

MÉMOIRES

DE

L'ACADÉMIE DES SCIENCES

ARTS ET BELLES-LETTRES DE CAEN



PRINCIPALES DATES
DE L'HISTOIRE DE L'ACADÉMIE DE CAEN

Reconnue d'utilité publique par décret impérial du 10 août 1853

- 1652 Des réunions littéraires, constituant le noyau de la future Académie, se tiennent autour de Moisant de Brieux à l'Hôtel du Grand-Cheval, place Saint-Pierre, à Caen.
Ces réunions durent jusque vers 1701, sous la direction de Regnault de Segrain.
- 1662 Premières réunions de l'Académie de Physique de Caen autour d'André Graindorge et Daniel Huet.
- 1667 Des statuts sont accordés à l'Académie de Physique, qui se disperse pourtant vers 1676.
- 1705 Lettres patentes de Louis XIV établissant officiellement l'Académie Royale des Belles-Lettres de Caen, qui retombe en sommeil vers 1714.
- 1731 M^{gr} de Luynes devient Protecteur de l'Académie.
- 1753 L'Académie reprend ses séances à l'Hôtel du Grand-Cheval (Hôtel d'Escoville), sous la direction de l'intendant Fontette.
- 1754 Publication du tome I^{er} des *Mémoires* de l'Académie.
- 1793 La Convention supprime les Académies.
- 1800 Le Préfet Dugua ressuscite l'Académie sous le nom de Lycée.
- 1802 L'Académie prend son titre définitif d'Académie des Sciences, Arts et Belles-Lettres de Caen. Elle n'a pas interrompu son activité depuis lors.
- 1811 Reprise de la publication des *Mémoires*.
- 1853 L'Académie est reconnue d'utilité publique par le décret impérial du 10 août.
- 1955 L'Académie tient une séance publique solennelle à l'occasion de la célébration du quatrième centenaire de Malherbe.
- 1970 L'Académie fixe de nouveau son siège à l'Hôtel d'Escoville, acquis par son fondateur Moisant de Brieux.
- 2002 En juin, l'Académie célèbre ses 350 ans d'existence en présence de M. Pierre Messmer, chancelier de l'Institut de France, et de très nombreux invités.

MÉMOIRES

DE

L'ACADÉMIE DES SCIENCES,
ARTS ET BELLES-LETTRES DE CAEN

Fondée en 1652. Lettres patentes de 1705.
Reconnue d'utilité publique (décret du 10 août 1853).

CONFÉRENCE NATIONALE
DES ACADÉMIES
CAEN 2022

Avec le concours de la Ville de Caen, du Conseil départemental du
Calvados et de la Région Normandie.

TOME LVII

2022

Siège : Hôtel d'Escoville, 12 place Saint-Pierre, 14000 CAEN

Tél./Fax. : 02 31 86 14 16

Courriel : accaen.scabl@orange.fr

Site Internet : <http://www.academiecaen-scabl.com>

Les auteurs des conférences ici publiées sont seuls responsables des opinions qu'ils expriment.

Reproductions : tous droits réservés

Hommage à Henry Delisle 28 avril 2022

Henry Delisle, décédé le 18 avril 2022 présidait l'Académie des Sciences Arts et Belles-Lettres de Caen depuis le 1^{er} janvier 2021, il y avait été élu membre associé le 14 janvier 2017, puis en était devenu vice-président en janvier 2019.

Inspecteur général de l'Agriculture, ancien député de Lisieux, ancien conseiller général du Calvados, il a toujours été passionné de culture. Il est l'auteur, avec Marc Gauchée de *Culture rurale, culture urbaine ?* paru aux éditions Le cherche midi en 2006 où il multiplie les propositions en vue de favoriser le développement des pratiques culturelles dans les espaces ruraux et s'interroge sur les chances d'une politique culturelle en milieu rural. « Nous vivons dans une époque qui a vu triompher une idéologie capitaliste et libérale qui s'interdit de concevoir un projet collectif autrement que comme la somme des projets individuels ». Comment construire du commun ? Cette interrogation ne cessera de le hanter.

Il avait également remis au ministre de l'Agriculture un rapport remarqué d'évaluation de la convention entre le ministère de l'Agriculture et celui de la Culture qui organisait les relations entre les deux ministères au sein des établissements d'enseignement agricole. Il a aussi présidé pendant longtemps le jury de recrutement des professeurs d'éducation culturelle de ces lycées.

Auteur d'un ouvrage au titre provocateur *Si vous aimez le parti socialiste, quittez-le*, Henry a aussi présidé l'association des Amis d'André Lemaître peintre du pays d'Auge.

C'est sa passion pour la culture qui a guidé son activité au sein de l'Académie. Citant Paul Valéry, il aimait répéter que si « la singularité de l'Académie est d'être indéfinissable » elle n'en doit pas moins se comporter tel « un phare de la culture ».

Sa préoccupation première fut de donner toute sa place à l'Académie dans la ville, pour preuve l'installation du secrétariat au rez-de-chaussée de l'Hôtel d'Escoville, avec une vue directe sur la place et l'église Saint-Pierre au cœur de la cité historique. Ce fut la première étape de la stratégie mise en place par Henry.

Henry Delisle avait aussi et surtout pour ambition de marquer la place de l'Académie dans la vie culturelle de la cité, l'homme politique n'était jamais loin, son activisme a parfois bouleversé le chemin sagement balisé de notre vénérable institution.

Au sein de l'Académie, Henry, sous des dehors parfois sourcilleux et bourrus, veillait à l'expression de la collégialité dans le fonctionnement des instances, dans le respect de l'activité de chacun.

Dès son élection en 2017, Henry s'est totalement impliqué dans la réalisation de l'ouvrage *Quelque chose en nous de Normandie*, ouvrage inspiré par la Normandie, son histoire, sa culture, ses populations, ses ressources, son positionnement en France et dans le monde. Il dirigea très vite le comité de rédaction avec enthousiasme et imagination, en bouscula allègrement le programme, les après-midis de travail en groupe furent organisés à son domicile et se sont souvent achevés telles les aventures d'Astérix – certes non pas par un banquet, mais par un dîner partagé dans la bonne humeur. À lire dans cet ouvrage son très beau poème « Le port ».

Son souci de la visibilité de l'Académie dans l'espace public l'engagea à redéfinir et actualiser le site internet qu'il imaginait devenir un incontournable outil de communication de notre Compagnie. C'est avec la même ambition qu'il transforma la maquette notre revue interne semestrielle *La Gazette* pour lui donner une forme attractive et familière.

Ses éclats de rire, ses explosions de colère sincères ou simulées dont les murs de la salle Moisant de Brieux conservent le souvenir, une attention pleine d'empathie pour ses nombreux amis, l'engagement sans faille pour les causes auxquelles il se donnait, une joie de vivre toute rabelaisienne exprimaient la complexité d'un homme pétri de contradictions assumées.

Toujours habité par la recherche de la grandeur de l'Académie, Henry Delisle a initié « les mardis de l'Académie », rencontres avec des personnalités reconnues, échanges et réflexions sur des sujets en débat dans la société. Grandement perturbés par la crise sanitaire, ces mardis de l'Académie ont permis d'entendre Jean-Pierre Jouyet et Dominique Raimbourg.

Mais la grande affaire qui occupait et préoccupait Henry Delisle reste l'organisation par l'Académie de Caen, en octobre 2022 de la Conférence nationale des académies. Cette manifestation exceptionnelle qui réunit tous les deux ans, dans l'une des trente-deux Académies de région les délégués de l'ensemble, il l'avait sur les pas de ses prédécesseurs à la tête de l'Académie, obtenue de haute lutte. Le thème de réflexion qu'il a voulu consacrer à un sujet porteur de débat dans la société sera le suivant : « Intelligence artificielle, biotechnologies, neurosciences : vers quel humanisme ? ». Il avait l'intention d'y aborder un aspect politique : les relations entre IA et Démocratie.

Il avait aussi imaginé réserver un accueil exceptionnel à nos confrères des autres académies. Henry n'était jamais à court d'idées et d'ambition. Il avait envisagé de créer un spectacle (opéra, pièce de théâtre, lectures...). La réussite de la comédie musicale *Les couleurs du temps* imaginée, conçue et créée par lui et donnée au conservatoire de Caen en juin 2017 l'y avait encouragé. La crise sanitaire et la maladie l'en auront empêché.

Son investissement au sein de l'Académie et auprès des élus municipaux aura favorisé la tenue de la Conférence nationale des académies dans les locaux prestigieux de l'Abbaye-aux-Hommes.

Pour l'ensemble de son action, pour les résultats obtenus et aussi pour les projets engagés, l'Académie des Sciences, Arts et Belles-Lettres de Caen veut rendre un hommage solennel et présenter ses remerciements chaleureux à son président Henry Delisle.

Didier LAFORGE
Vice-président de l'Académie des Sciences, Arts
et Belles-Lettres de Caen

CAEN VILLE D'INVENTEURS ET DE SAVANTS

par M. Joël BRUNEAU

Maire de Caen

Membre d'honneur de l'Académie des Sciences Arts
et Belles-Lettres de Caen

Incontestablement, la vie des hommes a été davantage bouleversée lors des deux derniers siècles que lors des deux millénaires qui ont précédé. Cette évolution notable et rapide est en très grande partie due à l'avènement des énergies fossiles. Rétrécissement continu des temps de parcours, accroissement de la production industrielle et mécanisation du travail agricole ont donné lieu à une société nouvelle. Une société au sein de laquelle la force physique des hommes a pu être décaplée par l'énergie et les machines.

Cette révolution technique, technologique et industrielle est aujourd'hui à l'origine du monde que nous connaissons. Un monde toujours en mouvement, qui se dirige vers une nouvelle grande révolution technologique. Si la révolution industrielle a permis de remplacer la force des ouvriers par celle, bien supérieure, des machines, cette nouvelle révolution semble, elle, ouvrir la voie à l'augmentation de nos capacités cognitives grâce à l'intelligence artificielle (IA), au *deep learning* (DL) ou encore aux biotechnologies.

Ces innovations, dont certaines sont déjà présentes dans notre quotidien, suscitent des sentiments contraires. Si certains louent l'espoir qu'elles portent, notamment dans les domaines de la santé, de la recherche ou même de l'administration publique, d'autres se méfient d'un monde désincarné, déshumanisé, dont le fonctionnement serait entièrement dicté par des ordinateurs.

Quoi qu'il arrive nous pouvons être certains que ces innovations seront à l'origine, c'est d'ailleurs le propre des

innovations, de changements majeurs dans notre manière de penser et d'organiser notre société. Derrière ces grandes innovations qui ont marqué et marqueront notre histoire, nous trouvons toujours de grands défricheurs. Des savants, des scientifiques, des investisseurs, des créateurs qui donnent un cap au progrès technique et qui se livrent une compétition de tous les instants pour faire progresser nos sociétés.

Être capable de réunir ces savants a toujours été pour les états, mais aussi les régions et les villes, un moyen de puissance, une manière de montrer son ambition. Dans l'histoire, de nombreuses villes ou cités-états ont su attirer ou faire naître les grands génies et les grandes entreprises innovantes. De Florence à San-Francisco, elles ont toutes réussi à cumuler des atouts indispensables à la formation d'un terreau fertile à l'innovation, à la création. Ce sont très souvent de grandes villes universitaires, attractives, à la position géographique idéale et qui disposent d'une vie culturelle et intellectuelle bouillonnante.

Parmi ces villes emblématiques auxquelles nous pouvons systématiquement raccrocher au moins un grand scientifique, un inventeur ou une entreprise, je crois que Caen a toute sa place. Elle possède un terreau fertile pour l'innovation, un terreau très ancien et qui n'a jamais cessé d'être conforté. Un terreau qui a donné naissance hier à de grands noms scientifiques et qui continue aujourd'hui d'être pourvoyeur d'innovations et d'entreprises créatrices. Notre ville a trop longtemps été considérée comme une belle endormie ou comme une ville dont l'histoire ne se résumait qu'à l'épopée de Guillaume le Conquérant, au Débarquement de Normandie et à la présence du Mémorial. Je crois que nous devons voir à quel point son essence profonde est celle d'une ville de savants, d'inventeurs, de défricheurs.

Caen est avant toute chose une ville millénaire, dont la première mention dans un document officiel date de 1025. Alors qu'elle n'était composée que de plusieurs petits bourgs,

elle prend son essor grâce à Guillaume de Normandie, lui-même innovateur dans le domaine politico-administratif. Organisée autour du château et de ses deux abbayes, l'Abbaye-aux-Hommes et l'Abbaye-aux-Dames, elle devient rapidement la deuxième ville de Normandie et concentre lettrés, juristes et marchands. Ces abbayes, plus que des lieux de savoirs deviennent aussi des modèles de l'architecture romane normande qui va irriguer l'Europe pendant plus de deux siècles. Les cathédrales de Canterbury et de Rochester en Angleterre, l'abbaye de Dunfermline en Écosse, le Palais des Normands et l'église San-Cataldo en Sicile et les nombreux châteaux normands du sud de l'Italie en sont quelques exemples particulièrement frappants. Cette première couche sédimentaire profonde marque déjà la Ville et lui donne cette vocation intellectuelle qui se prolonge aujourd'hui avec sa vie culturelle, son pôle de recherche et son université.

En 1432, une deuxième couche importante s'ajoute à la ville. Avec pour but originel de former leurs élites administratives, les Anglais, qui occupaient alors Caen, fondent une Université, largement plébiscitée par les élites locales. Elle est alors la troisième Université anglaise après les très prestigieuses Universités d'Oxford et de Cambridge. Au fil des décennies, l'Université de Caen s'affirme comme l'une des universités françaises les plus puissantes et formant de nombreux intellectuels.

Concomitamment à cet essor, Jacques Moisant de Brioux fonde en 1652 l'Académie des Sciences, Arts et Belles-Lettres de Caen, première académie fondée en France après celle de Paris. Dix ans plus tard, l'un des plus grands savants de son temps, Pierre-Daniel Huet et son ami André Graindorge fondent l'Académie de Physique de Caen, première de province à être reconnue par le roi. La fondation de ces deux académies et le développement de l'Université coïncident alors avec l'émergence de grands intellectuels, dans des domaines aussi divers que les lettres, l'histoire, les mathématiques, la physique ou encore la médecine. En 1689,

Mme de Sévigné parlera même de Caen comme « la source de nos plus beaux esprits ».

De beaux esprits qui, comme elle, participeront à la création de salons littéraires renommés au XVIII^e siècle. Inspirés par les idées des Lumières, ils attirent durant la période révolutionnaire des personnalités influentes, à l'image de la jeune Charlotte Corday ou de nombreux députés Girondins ayant fui Paris pendant la Convention. Dans la continuité de ces idées des Lumières et du mouvement des encyclopédistes se forment également à Caen, dès le XVIII^e siècle, mais surtout au XIX^e siècle, de nombreuses sociétés savantes. D'horizons multiples, elles témoignent de la grande diversité des groupes intellectuels caennais. Des sociétés sont par exemple fondées autour des questions d'agriculture, de commerce, de sciences naturelles, d'histoire ou encore de beaux-arts. Comme d'autres grandes villes françaises et européennes, Caen prend elle aussi part à la grande poussée scientifique du XIX^e siècle.

Ce sont toutes ces couches sédimentaires qui ont contribué à faire de Caen une ville de savants, un terrain particulièrement fertile à la création, l'innovation et à la réflexion.

Au cours du XX^e siècle, les guerres, et plus spécifiquement la Seconde Guerre mondiale ont marqué un tournant dans la dynamique de la ville. Durant l'été 1944, Caen est bombardée et l'Université est détruite. En quelques semaines, elle perd plus de la moitié de ses habitants, qui fuient dans les campagnes alentours. Ce moment aurait pu marquer un coup d'arrêt à l'histoire intellectuelle et universitaire caennaise et aurait pu reléguer la ville au second rang, derrière Rouen avec la disparition totale de l'Université.

Mais, forte de son histoire millénaire et de ses racines profondes, Caen s'est relevée. Dès l'automne 1944, les étudiants sont installés dans les locaux de l'École normale d'instituteurs en attendant la reconstruction de l'Université. Après près de 10 années de travaux, la nouvelle Université de

Caen est inaugurée en 1957 derrière le château, au cœur de la cité. Une décennie plus tard, le nombre d'étudiants augmentant, le ministère de l'Éducation nationale décide, en 1968, de construire un nouveau campus. Ce deuxième campus, dont la construction au nord de la ville s'est étalée sur deux décennies a été la première pierre de la fondation d'un grand pôle d'enseignement supérieur et de recherche sur le plateau nord de Caen. Entre 1973 et 1985 sont regroupés sur ce lieu tous les grands outils de recherche en science que sont le Centre de lutte anti-cancer François Baclesse, le CHU, le Grand accélérateur national d'ions lourds (GANIL), l'un des plus grands laboratoires de recherche en physique atomique et nucléaire au monde et enfin le Centre d'imagerie et de recherche en neurosciences (CYCERON). Une concentration d'outils de recherche exceptionnelle pour une ville de 100 000 habitants qui témoigne de l'ancrage puissant du monde de la recherche et de l'innovation à Caen.

Dans *Le chiffre et le songe*, Jacques Blamont, l'un des pères du programme spatial français, au travers de l'exemple du palais d'Uraniborg, plus grand observatoire d'Europe au XVI^e siècle, nous montre combien la concentration d'infrastructures de grande qualité était un accélérateur de découvertes scientifiques. Demain, notre grand défi sera de renforcer le pôle d'innovation caennais en réunissant les acteurs de la recherche, les étudiants et les industries. Pour cela, nous avons fondé en 2018 l'association EPOPEA, pour concentrer tous les acteurs du plateau nord autour d'un projet commun. Très prochainement, ce grand campus de l'innovation accueillera une Maison des chercheurs, à vocation internationale, ainsi qu'une Maison de l'innovation. L'enjeu de ces projets est double. Ils doivent à la fois nous permettre de renforcer nos outils d'innovation en mettant de nouveaux outils performants à la disposition des chercheurs de nouveaux outils performants. Ils doivent aussi nous permettre de valoriser davantage ces formidables atouts qui, bien souvent, ne sont pas spontanément associés à Caen dans les esprits.

Comme nous avons pu le voir, Caen s'est construite depuis 1 000 ans comme une ville de savoir, de culture, de recherche, d'inventions, de savants. Depuis 1 000 ans, Caen a vu émerger de nombreuses grandes figures dans des domaines très divers. Nous pouvons par exemple citer François de Malherbe pour les lettres, Arcisse de Caumont et Michel de Boüard pour l'histoire et l'archéologie, le docteur Thibout de la Fresnaye, inventeur de génie en médecine, Dumont d'Urville, grand explorateur, Augustin Fresnel en physique et en optique, Pierre-Simon de Laplace, l'un des plus grands mathématiciens français ou encore Cécile Dewitt-Morette, brillante chercheuse à la frontière entre les mathématiques et la physique quantique.

Toutes ces figures, souvent trop peu connues du grand public, ont bénéficié du fertile terreau caennais et contribuent aujourd'hui à la renommée de la ville. Tous avaient pour objectif de faire progresser le monde en lui apportant de grandes découvertes. Ils représentent l'esprit de cette ville, une ville humaniste, résolument tournée vers la science et le savoir.

Aujourd'hui encore, cet esprit rayonne et permet à Caen de se placer parmi les villes les plus attractives de France, à la fois pour sa qualité de vie reconnue, mais aussi pour sa dynamique en matière d'emploi et d'innovation.

Aux côtés des grandes entreprises historiques de la Communauté urbaine, nous avons la chance de voir émerger une constellation de jeunes entreprises innovantes qui s'appuient sur cette histoire et toutes ces grandes structures de recherche que nous avons évoquées. Dans la grande bataille internationale pour l'innovation, nous avons la chance de pouvoir compter sur cet écosystème riche de sa diversité, dans des domaines aussi variés que le médical, le biomédical, les neurosciences, l'intelligence artificielle, le *big data* ou encore la robotisation avec des applications aussi diverses que la détection précoce des cancers et l'amélioration de leur traitement, la facilitation des dialyses, la lutte contre les maladies neurodégénératives, la meilleure compréhension du

fonctionnement de notre cerveau, l'amélioration de la navigation en mer par le *big data*, la robotisation des entreprises et usines ou encore la cybersécurité. Tous ces champs d'étude et ces domaines d'application auront un impact considérable sur le monde de demain.

Symboliquement, notre cœur de ville est aujourd'hui encadré par les deux abbayes, à l'origine de son développement, par le château, qui témoigne de son caractère de place juridique et de lieu de pouvoir, par l'Université, créée il y a près de 600 ans et bien sûr par l'Orne, vecteur de commerce et reliant Caen à la mer et au reste du monde. Cet encadrement symbolique du centre de la cité témoigne de la riche histoire de Caen, capitale régionale, ville de savoirs et d'échanges.

Tout au long de son histoire, notre ville a vu plusieurs couches de sédiments s'accumuler pour former cette identité caractéristique. Une identité profondément humaniste, marquée par ses grandes figures mais aussi par de profonds traumatismes qui lui ont donné une vocation universelle.

À l'heure où nous devons faire face à de nombreux défis collectifs et planétaires et à l'heure où nous vivons une nouvelle révolution technologique majeure, je crois que Caen aura toute sa place dans les réflexions qui animeront notre société, comme elles animeront très certainement le colloque de la Conférence nationale des académies. Forte de son histoire millénaire mais surtout, forte de ses Caennais, chercheurs et entrepreneurs, Caen saura apporter sa pierre aux futures grandes innovations et réflexions génératrices de progrès.

C'est ce visage que Caen entend montrer en accueillant cette année la Conférence nationale des académies et qu'elle offrira à tous avec force en 2025 en fêtant son millénaire.

Bienvenue à Caen et à bientôt.

INTELLIGENCE ARTIFICIELLE BIOTECHNOLOGIES, NEUROSCIENCES : VERS QUEL HUMANISME ?

LE CHOIX D'UN THÈME

Lorsque, en janvier 2020, la CNA a confié à l'Académie des Sciences, Arts et Belles-Lettres de Caen l'organisation de la rencontre de 2022, une interrogation a immédiatement circulé chez les académiciens : quel thème proposer ? Il ne conviendrait pas de vanter les charmes et les richesses de la Normandie. Caen partage avec Cherbourg et Rouen le privilège d'être une des trente-deux académies réunies dans la Conférence nationale. Henry Delisle, alors vice-président, proposa de réfléchir au sujet, non sous la forme d'un colloque scientifique réservé aux spécialistes, mais en abordant les multiples enjeux portés par les promesses comme par les dangers contenus dans l'expression la plus manifeste de la révolution technologique à l'œuvre, les algorithmes. De longs débats s'ensuivirent avant de parvenir à la formulation de l'appel à communication qui a été largement diffusé au sein des académies : « Intelligence artificielle, biotechnologies, neurosciences : vers quel Humanisme ? » Notre président Henry Delisle est décédé le 18 avril dernier : cette rencontre lui est donc dédiée, et un hommage particulier lui est rendu.

L'intelligence artificielle est partout dans notre quotidien. Elle informe et multiplie les résultats des biotechnologies comme des neurosciences. Elle se loge jusque dans l'art, et la littérature en fait abondamment son miel. Pourquoi, alors, proposer un thème aussi banal à la réflexion des participants à la CNA ? Justement parce que sa présence produit des effets porteurs de lourds enjeux économiques, sociaux, politiques, éthiques, métaphysiques (« Comment prendre soin des "enfants Crispr" ? », titrait *Le Monde* du 20 avril 2022). Les risques de voir se fabriquer une société éclatée, déshumanisée

sont réels ; l'automatisation et la robotisation conduisent à la suppression de nombreux emplois et transforment le statut social et moral du travail ; les manipulations génomiques vont-elles déboucher sur des individus « augmentés » ? Mais, en même temps, les réussites technologiques, les innovations industrielles, les développements de la robotique sont éclatants, entraînant des progrès décisifs en médecine, dans la recherche scientifique, dans la vie quotidienne. Se trouve alors posée la question des interactions entre l'intelligence artificielle – les intelligences artificielles ? – et l'intelligence humaine – les intelligences humaines ? –. Et celle de la régulation, juridique ou politique de leur usage – de leurs usages. Il apparut alors qu'un programme riche et porteur de débats pouvait être construit.

L'APPEL À COMMUNICATION

Le texte d'orientation ci-dessous, accompagné d'un appel à communication a donc été lancé en mai 2021 en direction de l'ensemble des académies réunies dans la Conférence nationale.

INTELLIGENCE ARTIFICIELLE, BIOTECHNOLOGIES, NEUROSCIENCES : VERS QUEL HUMANISME ?

Parmi les questions vives et les querelles qui agitent l'espace public, les enjeux, les promesses, les dangers de ce que communément on désigne sous les termes « intelligence artificielle » occupent une place de choix, tant dans les cercles académiques cultivés que dans les conversations privées. L'alliance impossible de deux termes appartenant à des champs sémantiques étrangers l'un à l'autre, l'ambiguïté de la formule, les approximations qu'elle autorise, les conséquences potentiellement dangereuses des applications qu'elles contiennent sont contredites par les résultats de la révolution numérique, de la robotique, des technologies médicales de pointe. Les neurosciences constituent de ce point de vue un champ de réalisations et de promesses porteur d'améliorations dans la vie de la cité et dans la vie des individus et sont grosses des mêmes inquiétudes. Alors, intelligence artificielle, neurosciences, technologies de pointe préparent-elles un univers déshumanisé dominé par de glaçantes technologies ? un monde orwellien de surveillance permanente et de contrôle social généralisé ? un nouvel âge d'or pour l'humanité ? le dépassement de l'homme dans un transhumanisme « augmenté » ou son accomplissement ? La réflexion sur les sciences et les technologies contemporaines, loin des certitudes du café du commerce, réveille et renouvelle la vieille inquiétude pascalienne : « Car, enfin, qu'est-ce que l'homme dans la nature ?¹ », à laquelle Jacques Ellul², Gilbert Simondon³ ou Bernard Stiegler dans une optique différente, ajouteraient « et dans la technique ».

1 PASCAL, *Pensées*, Paris, Robert Laffont, 1960, p. 180 (Guersant, 1954, 415).

2 Jacques ELLUL, *Le bluff technologique*, Paris, coll. Pluriel, 2012.

3 Gilbert SIMONDON, *Du mode d'existence des objets techniques*, Paris, Aubier, 1989 (première édition 1958).

Bien que datant de 1994, la très longue notice que le *Trésor de la Langue française* consacre au mot « intelligence » fournit une définition éclairante de l'expression « intelligence artificielle » : « Recherche de moyens susceptibles de doter les systèmes informatiques de capacités intellectuelles comparables à celles des êtres humains ». Le *Dictionnaire de l'Académie française*, dans sa neuvième édition, à la rubrique « intelligence » renvoie « intelligence artificielle » à l'article « artificiel : Intelligence artificielle, ensemble de propriétés rapprochant du cerveau humain certains systèmes informatiques très évolués ». Le *Dictionnaire de français Larousse* en ligne donne : « Intelligence artificielle (IA) : ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine ». Même Wikipédia y va de sa définition : « un ensemble de concepts et de technologies plus qu'une discipline autonome constituée ».

Déjà, au siècle dernier, le succès de la littérature d'anticipation et de science-fiction traduisait, au-delà des frissons procurés par l'imagination, les inquiétudes et interrogations philosophiques sinon métaphysiques des lecteurs inquiets de la vision du monde à venir promise ou dénoncée par des auteurs à l'imagination fertile. Aujourd'hui, la réalité rejoint voire dépasse la fiction : systèmes informatiques, machines capables de simuler l'intelligence humaine dont la mise en œuvre produit miracles ou catastrophes, neurosciences cliniques, cognitives ou sociales et leurs applications recouvrent de multiples secteurs de la société, vie quotidienne, santé, éducation, économie, justice, politique, mais aussi arts, culture. La Chine installe un dispositif de reconnaissance faciale qui fiche et contrôle les citoyens, les réseaux sociaux et la publicité devancent et modèlent nos désirs, la gestation pour autrui laisse entrevoir de potentielles manipulations eugéniques, les tentatives de régulation par les États laissent les GAFAs indifférents.

Mais, en même temps, la Commission européenne publie un *Livre blanc sur la souveraineté numérique*⁴ et fait adopter un règlement général sur la protection des données⁵ très protecteur, le Parlement français légifère « Pour une République numérique »⁶, la Commission nationale informatique et libertés (CNIL) organise un débat public sur les enjeux éthiques que soulèvent les algorithmes et l'intelligence artificielle⁷, les comités d'éthique multiplient mises en garde et recommandations⁸, les penseurs de la technocritique⁹, insistant sur les usages, réintègrent la technique, constitutive de l'homme même, dans le processus d'humanisation. Les questions abondent, qui exigent des

4 *Livre blanc sur la souveraineté numérique, une approche européenne axée sur l'excellence et la confiance*, COMMISSION EUROPÉENNE, Com (2020) 65 final, 19/02/2020.

5 Rectificatif au Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données et abrogeant la directive 95/46/CE (règlement général sur la protection des données), *Journal officiel de l'Union européenne*, L 127 2 du 23/ 05/ 2018.

6 Loi n°2016-1321 du 7 octobre 2016 pour une République numérique, *Journal officiel de la république française*, n° 235 du 8 octobre 2016.

7 CNIL, *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi Pour une République numérique, décembre 2017, 80 p.

8 Par exemple, COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE, Position du CCNE sur le décret 2020-384 du 1^{er} avril 2020 (17 avril 2020) ; Contribution du Comité consultatif national d'éthique à la révision de la loi de bioéthique 2018-2019 (25/09/2018) ; Données massives et santé : état des lieux, prospective et nouvelles questions éthiques (29/05/2019). COMITÉ NATIONAL PILOTE D'ÉTHIQUE DU NUMÉRIQUE, Enjeux d'éthique liés aux outils numériques en télémédecine et télésoin dans le contexte de la Covid 19, *Bulletin de veille* n° 3, 21/07/20.

9 Par exemple, STIEGLER, Bernard, *La Technique et le Temps*, 3 volumes, 1. *La faute d'Épiméthée* ; 2. *La Désorientation* ; 3. *Le Temps du cinéma et la Question du mal être* ; suivi de *Le nouveau conflit des facultés et des fonctions dans l'Anthropocène*, Paris, Fayard, 2018, 970 p.

réponses approfondies. Que nous apprennent l'histoire des techniques, la littérature ? Miniaturisation, contrôle et stockage des données, manipulations génétiques, quelle responsabilité du savant, de l'ingénieur, du chirurgien, du généticien, du juriste, de l'homme politique ? À nouvelles technologies, nouveaux usages : comment se les approprier ? Que deviennent les valeurs de liberté, de démocratie, d'état de droit face à Google ou à Tencent ? Quel contrôle démocratique exercer sur les nouveaux maîtres des technologies : comment conserver une souveraineté nationale ? Le progrès, illusion dangereuse pour l'humanité ou promesse d'un avenir meilleur ?

Pour répondre à ces questions, il faut d'une part comprendre le système cérébral humain et d'autre part le comparer aux modèles de calcul introduits jusqu'à présent. En 1936, Allan Turing définit un système de calcul communément appelé machine de Turing. Les systèmes informatiques et par conséquent l'intelligence artificielle sont basés sur cet archétype calculatoire. Les tentatives pour dépasser un tel modèle ont donné naissance à des paradigmes qui, soit sont, au mieux, aussi performants, soit se sont heurtés aux limites de la physique théorique. S'appuyant sur le modèle de Turing, la thèse d'Alonzo Church – « tout ce qui est calculable est calculable par la machine de Turing » – conduit à une question : les neurosciences ont-elles un modèle cérébral permettant de simuler une machine de Turing ? Nous savons qu'il existe des problèmes non calculables comment les définir sous un angle neuroscientifique ? Ces questions nous ramènent au problème de la conscience.

LES CONTRIBUTIONS

L'appel à communication a permis de recueillir treize propositions de communication dont la variété des aspects abordés et la richesse des contenus a conduit à demander à leurs auteurs de rédiger un texte complet et de préparer une

communication orale. Les origines, l'histoire, les réussites, les inquiétudes par rapport aux mésusages toujours possibles ; la fonction de l'art et de la culture ; les relations entre l'homme et la machine, l'intelligence artificielle et l'intelligence humaine ; et aussi les interrogations éthiques, philosophiques, voire métaphysiques : autant d'aspects abordés dans les communications. Un dernier texte est venu se joindre aux treize précédents, sans que l'auteur puisse se déplacer pour prononcer sa communication, mais le sujet traité comble une lacune en présentant la question de la justice pénale confrontée à l'intelligence artificielle. Un dialogue entre deux anciens directeurs éclaire l'histoire et le rôle fondamental de CYCERON dans le développement à Caen d'un pôle de recherche internationalement reconnu dans le domaine des neurosciences. Enfin, l'Académie a ouvert le débat en donnant la parole à deux contributeurs particuliers, le jeune créateur caennais d'une entreprise qui fabrique des robots, enthousiaste et engagé ; et le commissaire européen Thierry Breton porteur de deux directives récentes destinées à assurer à l'Union européenne la maîtrise de sa cybersécurité. Notre confrère Jean Guglielmi, décédé le 2 janvier dernier, avait entrepris la rédaction d'une communication qu'il avait intitulée « Apprendre et penser resteront-ils l'apanage du vivant ? » qu'il n'a pu achever : l'ébauche figure également dans les textes reproduits, avec l'hommage qui lui a été rendu.

Sans doute l'appel n'a-t-il pas comblé toutes les attentes de la commission de préparation de la rencontre. Ainsi aucune proposition ne s'est manifestée pour traiter la question de la littérature ou du cinéma d'anticipation. Les conséquences sociales de l'évolution du travail, l'utilisation ou le détournement de l'intelligence artificielle à des fins de manipulation politique ne sont abordés directement par aucune contribution. Mais la diversité et la richesse des questions traitées sont en elles-mêmes amplement suffisantes pour ne nourrir aucun regret.

UN ITINÉRAIRE DE LECTURE

Le thème de travail proposé par l'Académie des Sciences, Arts et Belles-Lettres de Caen est en adéquation avec la Ville, son passé et aussi son avenir : « ville d'inventeurs et de savants », comme le souligne Joël Bruneau son maire, Caen abrite aussi la plus ancienne académie créée après l'Académie française, en 1652. C'est donc un Caennais qui ouvre la rencontre : Alain Bretto s'interroge sur les relations entre le calcul mathématique et le raisonnement humain et, remontant aux bâtons d'Ishango ou aux *Éléments* d'Euclide, brosse une vivifiante histoire du calcul qui le conduit jusqu'à Hilbert et Gödel, puis à Turing et à ses machines d'où sortiront les algorithmes et l'intelligence artificielle, « outil puissant qui deviendra prégnant dans nos sociétés ». Ce lumineux **printemps** de la science semble promis à un bel avenir, « et les fruits passeront les promesses des fleurs » chantait déjà Malherbe. Rebondissant sur Gödel et Turing, Jean Richert poursuit l'exploration du cheminement conduisant à la « simulation cognitive », à la cybernétique puis à l'intelligence artificielle, cette « science cognitive » qui rejoint la philosophie, la psychologie, les neurosciences, la linguistique et l'anthropologie. Il souligne ce qui reste la différence ontologique entre l'animal, la machine la plus perfectionnée, le robot et l'homme : la conscience, la « conscience de soi ». Mais les progrès fascinants des robots utilisant des algorithmes de plus en plus performants ne conduiront-ils pas à une « intelligence artificielle surhumaine » jusqu'au moment où « les potentialités de la machine deviennent supra-humaines » ? Faudra-t-il alors envisager un « futur transhumaniste », ce rêve d'un homme nouveau « métissé, hybridé » par les nanotechnologies ou les biotechnologies ? La question conduit Michel Chein à se demander si *Homo sapiens* ne va pas céder la place à un étrange et inquiétant *Homo numericus*, produit des réseaux de neurones artificiels, soumis à son environnement numérique. La question centrale de la

réflexion est ainsi posée : le calcul, les algorithmes ont entraîné de considérables progrès, des améliorations inappréciables dans la vie des sociétés et des hommes, mais la machine « apprenante » ne risque-telle pas de créer une catégorie d'individus post-humanistes sinon transhumanistes imperméables aux préoccupations éthiques ?

Les cinq communications suivantes soulignent les aspects favorables des progrès scientifiques liés à l'intelligence artificielle, aux neurosciences, à la robotique. Cette phase d'**épanouissement** semble se dérouler sans véritables limites. Jean-Claude Marchal articule la réflexion sur l'humanisme, la philosophie des Lumières et les perspectives transhumanistes. Le transhumanisme, nous dit-il « n'est pas une métaphysique, c'est une philosophie de la technique et de l'action ». Des avancées scientifiques, de « la convergence NBIC », faut-il conclure que « l'euthanasie de la mort va devenir l'ultime frontière pour l'Humanité » ? Et les espérances transhumanistes d'un homme augmenté de ses pouvoirs sur lui-même jusqu'à une illusoire immortalité ne sont-elles pas les héritières des Lumières, de leur foi dans la science et dans la liberté humaine ? Aujourd'hui, dit Patrick Penel « le numérique est devenu l'affaire de toute la société » et le mythe de la « machine pensante » va-t-il accomplir la promesse prométhéenne de transformation de la nature et de la nature humaine ? Mais une IA conquérante n'a pas encore effacé le rêve non accompli d'une mythique « post-humanité ». La création du laboratoire CYCERON relatée par Jean-Michel Derlon et Francis Eustache, et celle de l'entreprise Conscience Robotics par Iliès Zaoui illustrent ces avancées technologiques qui concourent au mieux-vivre des humains. Et l'exemple rapporté par Frédéric Thibault-Starzyk de l'utilisation de l'intelligence artificielle dans l'intégration des disciplines universitaires pour surmonter les incertitudes des savoirs compartimentés montre que savoirs et techniques désormais réunis sont porteurs de développements sans fin dont il faudra nécessairement encadrer légalement les usages. Car toutes les

techniques aujourd'hui réunies sont également susceptibles de conduire à une société de surveillance généralisée.

Des vagues d'**ombres et de lumières** recouvrent donc les terrains conquis par les applications de l'intelligence artificielle comme celles des neurosciences. Mareike Wolf-Fédida, à partir de l'analyse d'une expérimentation dans le secteur de la psychothérapie rogerienne simulant une conversation entre deux machines, souligne l'altérité radicale de la communication humaine et les limites de l'intelligence artificielle avec ses impacts négatifs sur la culture. Et l'esquisse de Jean Guglielmi s'interroge et s'inquiète : « apprendre et penser resteront-ils l'apanage du vivant ? ». La brillante et roborative analyse, conduite par Jacques Bouineau à partir de sa position d'historien du droit, des conséquences de la numérisation de la société souligne les dangers que la réification globale – « Les hommes étant désormais traités comme des choses qui possèdent un logiciel [...] avec comme objectif d'atteindre des performances » – fait courir à la démocratie et aux libertés. Faut-il se résoudre à n'être plus qu'un *homo connecticus* aliéné ? « Et si le sursaut venait de l'homme lui-même ? ». Mais les « hommes sont viscéralement violents » et la société – et non la foule – repose sur d'indispensables « principes régulateurs ». À l'heure où l'Union européenne impose aux plateformes numériques, aux réseaux sociaux et aux places de marché de vente en ligne des obligations, des devoirs et des responsabilités qui constituent une indéniable victoire politique, c'est un de ces aspects majeurs que le Commissaire Thierry Breton aborde en présentant la politique de la Commission et de l'Union européenne en matière d'intelligence artificielle et de souveraineté numérique : réguler, non pas brider, mais fixer des « règles proportionnées et flexibles [...] pour répondre aux risques spécifiques posés par les systèmes d'IA ».

Des **questionnements** demeurent, l'imagination s'empare des technologies numériques : l'irrationnel, l'inconscient, le beau sont percutés par l'intelligence artificielle et les

neurosciences. Jacques Boucharlat explore la question de la folie, de l'angoisse, des névroses et examine les modalités d'intervention de l'intelligence artificielle, des algorithmes dans les thérapies comportementales, avec tous les risques de dérive liés à « l'autonomisation de robots » utilisés comme aides-soignants dans certains services. Pour Richard Forestier, l'art est au croisement des technologies et de l'humanité. Or les nouveautés technologiques actuelles pénètrent profondément l'art contemporain, déstabilisent « l'Art conventionnel » et transforment l'activité artistique même, sans que changent ses finalités : « l'œuvre d'Art s'impose comme une manifestation privilégiée de l'Humanisme dans les activités humaines ». Le sonnet *Voyelles* d'Arthur Rimbaud fournit à Bernard Devauchelle un point d'appui pour analyser cette « confusion des sens » que traduit la synesthésie, à travers l'expérience du chirurgien, transformée radicalement par la cybernétique et ses applications technologiques, imagerie médicale, microscope, robotique.

Vers quel humanisme nous dirigeons-nous ? Une nouvelle **sagesse** est-elle en germe dans ces réflexions autour de la science et de l'éthique ? Face à l'irruption du numérique dans le monde du quotidien, Claude Kirchner nous invite à concevoir, tout en s'appuyant sur l'éducation, une éthique de la conversion numérique qui susciterait le respect de valeurs humanistes telles que « le respect des données personnelles, la transparence des procédures utilisées, l'auditabilité des programmes mis en œuvre ». Éric Suraud refuse de se résoudre au « scénario catastrophe de la disparition de l'Homme submergé par le numérique ». Il pointe quelques limites aux intelligences artificielles, la perception de l'ambiguïté, l'intuition de l'inconcevable, l'accès aux sentiments. Bien plus que sur une disparition, c'est sur une « renaissance de l'Homme » qu'il parie, sur un « nouvel Homme » qui aura retrouvé « la place dominante dans le libre usage de ses outils techniques [...] Continuons donc d'agir au mieux ». Et pour que notre action au service du bien s'exerce « dans une oasis

de fraternité », une place centrale doit être réservée à la justice et à son exercice. Jean Touzet, après un long cheminement professionnel de magistrat spécialisé dans la justice pénale, s'interroge sur l'articulation entre l'intelligence artificielle et ses applications, les perspectives transhumanistes et l'exercice de la justice pénale. Quelle aide les outils de l'intelligence artificielle peuvent-ils apporter à un juge dans la recherche de la vérité des faits, dans la détermination de la sanction. Et que pourrait apporter le transhumanisme à la détermination de la sanction pénale. Reste que la question de la responsabilité – celle de l'accusé, comme celle du juge – ne saurait être déléguée à une machine qui aurait reçu la personnalité juridique sans que de graves dérives ne soient contenues en germe dans une telle décision. Utilisation raisonnée comme aide à la décision, oui. Personnalité juridique conférée à une machine : danger. Pour conclure, Francis Eustache développe les apports des neurosciences à la connaissance du fonctionnement du cerveau et à la mise au point de démarches thérapeutiques nouvelles en matière d'altérations cérébrales. Il souligne la multiplicité des approches pour étudier les liens entre cerveau et cognition et montre l'articulation entre les données individuelles et les dimensions sociales de la mémoire collective. Progrès scientifiques, progrès technologiques et progrès thérapeutiques s'imbriquent et se renforcent au service du patient et plus largement au service de l'homme et de l'humanité.

Edgar LEBLANC
Secrétaire-adjoint de l'Académie des
Sciences, Arts et Belles-Lettres de Caen

LE CALCUL LE RAISONNEMENT HUMAIN ET L'INTELLIGENCE ARTIFICIELLE

par M. Alain BRETTO

de l'Académie des Sciences, Arts et Belles-Lettres de Caen

Le raisonnement, le calcul, la décision, forment une grande partie des occupations humaines. Ces activités s'appuient sur la logique, c'est-à-dire une suite de propositions « emboîtées » de manière cohérente par argumentation afin d'arriver à une conclusion. Cet enchaînement ordonne les idées afin de donner du sens à celles-ci.

La plupart des humains raisonnent de manière constructiviste, c'est-à-dire fabriquent effectivement l'objet étudié. Par exemple faire un gâteau, bâtir un immeuble, additionner l'entier a avec l'entier b , calculer la gestion des stocks, élaborer une fiche de paye. Un raisonnement constructiviste doit convaincre qu'une propriété est vraie, et doit également expliquer pourquoi elle est vraie. Par ailleurs, il doit montrer que l'essence – ce qui fait qu'une chose est ce qu'elle est – prouve l'existence. Une preuve d'existence est suffisante dans le cas de la non-effectivité. Ainsi ces deux types de dialectique ne sont pas, à mon sens en opposition. Le constructivisme peut être considéré comme une tentative d'appréhender les choses dans leur globalité, contrairement à l'opinion de Merleau-Ponty qui avance, dans *L'œil et l'esprit* que « la science manipule les choses et renonce à les habiter ».

Si nous voulons créer les objets étudiés il nous faut un mécanisme de fabrication : l'algorithmique fournit un tel dispositif. Un algorithme est une suite finie d'instructions finies ; par exemple : aller à la ligne k du texte ; pour tout i de 1 à n , ajouter au nombre x_i le nombre 2 ; une recette de cuisine est un algorithme. De là, un raisonnement constructiviste peut être vu comme un algorithme.

L'informatique est l'ensemble des activités humaines qui abordent l'information en utilisant des algorithmes, notamment grâce aux ordinateurs. La nature des problèmes traités par celle-ci est principalement de deux types. Calculatoire : faire une tarte, additionner deux nombres ; décisionnel : problème auquel on peut répondre soit par oui, soit par non, par exemple, est-ce qu'un voyageur de commerce peut visiter $k > 0$ villes en moins de p kilomètres ?

Donc l'informatique traite une information que l'on peut « mécaniser », ou alors des questions dont la réponse est soit oui, soit non.

Un algorithme est codé sur une machine par une suite de 0 et de 1 car l'ordinateur ne traite que des éléments binaires appelés bits (un bit classique est soit 0 soit 1). Un algorithme programmé, c'est-à-dire écrit dans un langage compréhensible par la machine est composé, de manière unique, d'une suite de 0 et de 1. De manière réciproque, toute suite de 0 et de 1 correspond exactement à un programme.

Cette suite est simplement un entier naturel écrit en base 2. De là, un algorithme est équivalent à un nombre.

Ainsi, raisonner c'est calculer ou calculer c'est raisonner !

Cela peut en partie expliquer la fascination des nombres. Ceux-ci véhiculent le raisonnement, mais ils peuvent aussi être vus comme une forme symbolique et non comme un symbole, c'est-à-dire une manière de rendre le monde intelligible et non un élément concret que l'on met en relation avec un élément abstrait, car le nombre est déjà une abstraction. Néanmoins je n'entrerai pas dans le détail des formes symboliques liées aux nombres. Ce qui nous intéresse ici c'est le calcul et il ne faut pas confondre celui-ci avec la notion de nombre.

Ainsi nous sommes amenés à nous interroger sur le sens profond d'un calcul. Pour cela il faut comprendre comment le calcul est apparu dans l'histoire humaine, donnons une histoire succincte de la notion de calcul.

LE CALCUL, UNE BRÈVE HISTOIRE.

Le calcul est né bien avant l'écriture qui, elle, est inventée en Mésopotamie en 3500 avant J.-C.

Les bâtons (os) d'Ishango, sont considérés comme le plus ancien outil calculatoire jamais mis au jour. Découverts en Afrique (Congo) ils sont datés d'environ 20 000 ans avant J.-C.

À Babylone, au 3^e millénaire avant J.-C. on retrouve les premières traces d'algorithmes, ils permettent la résolution d'équations algébriques. L'algorithme de la tablette YBC 7289 connu sous le nom « d'algorithme de Babylone » permet d'approximer racine de 2 et d'ailleurs la plupart des racines carrées, cela 2 500 ans avant J.-C. Ce morceau d'argile est une première forme connue du théorème de Pythagore.

Le livre VII des *Éléments* d'Euclide (III^e siècle avant J.-C.) fournit un algorithme qui a donné du fil à retordre à pas mal d'entre nous. Il permet de calculer le PGCD¹ de deux nombres. Non seulement sa convergence est prouvée mais il est basé sur un concept fondamental en informatique : l'itération. Les treize livres qui constituent ces *Éléments* sont fondamentaux dans la pensée moderne. Ceux-ci sont composés d'axiomes, de définitions, de théorèmes qui sont démontrés, ils traitent de la géométrie et des nombres. Ces ouvrages vont avoir une influence considérable sur l'évolution de la logique, concept introduit par Aristote. Ils auront un impact primordial sur le raisonnement humain dans la plupart des civilisations, ainsi que sur les sciences fondamentales développées par celles-ci. On retrouve ces recueils à Constantinople. La prise de la ville par les Turcs en 1453 provoqua l'exil des savants qui emportèrent leurs précieux manuscrits. Ceux-ci furent ensuite traduits en arabe et en latin, et seront étudiés durant des siècles dans toutes les grandes universités : Bagdad, Cordoue, Paris,

1 Plus grand commun diviseur.

Oxford, etc. Il y a de fortes chances que leurs contenus aient influencé la logique indienne. Ce succès est dû à la méthode employée, basée sur trois piliers : définitions, axiomatiques, assertions, qui sont prouvées en partant des deux premiers concepts.

Plus proche de nous, environ 80 ans avant J.-C., peut-être même avant, la machine d'Anticythère est un calculateur à engrenages qui détermine sans doute des positions astronomiques. Si c'est le cas, ce mécanisme est le premier connu qui traite de l'arithmétique modulaire. L'instrument décrit précédemment nous mène à un autre foyer intellectuel important. Le mathématicien et astronome chinois Sun Zi (III^e siècle après J.-C.) décrit un algorithme permettant de résoudre des systèmes d'équations particuliers (modulaires). En réalité le problème posé par Sun Zi peut s'écrire de la manière suivante :

« Quand le général Han Xing range ses soldats par trois, il reste deux soldats, quand il les range par cinq, il en reste trois et quand il les range par sept, il en reste deux. Combien l'armée de Han Xing comporte-t-elle de soldats ? » (*Tangente* 177 – Juillet 2017).

Cette question s'appelle « problème de Maître Sun » et donnera naissance au Théorème des restes chinois qui traite de la résolution de systèmes de congruences.

À titre indicatif donnons l'écriture moderne de cette énigme (le lecteur peu avisé des symboles mathématiques peut sauter les équations, car une interprétation est donnée ci-dessous) :

$$\begin{cases} x \equiv 2 \pmod{3} \\ x \equiv 3 \pmod{5} \\ x \equiv 2 \pmod{7} \end{cases}$$

Où \equiv est le symbole de congruence et « mod » est le symbole de modulo.

Bien qu'imparfaite la solution donnée par Sun est originale et imaginative, elle permet de résoudre des problèmes astronomiques du type :

« Existe-t-il un nombre x de jours pour que la pleine lune tombe au solstice d'hiver sachant qu'avant le solstice d'hiver il reste 11 jours et 5 jours avant la pleine lune ? » La question se traduit par le système d'équation suivant :

$$\begin{cases} x \equiv 11 \pmod{365} \\ x \equiv 5 \pmod{29} \end{cases}$$

Ce système peut se traduire par : existe-t-il un entier x qui, divisé par 365, donne un reste égal à 11 et qui, divisé par 29 donne un reste égal à 5 ?

Sun Zi est considéré en Chine comme étant le pendant de Pythagore en Occident.

Naturellement nous pourrions développer de manière détaillée l'histoire du calcul, jusqu'au XIX^e siècle mais cela nous mènerait hors de notre champ d'étude. Rapidement nous pouvons citer le mathématicien persan Al-Khwârizmi (VIII^e siècle) dont le nom donne le mot algorithme. Nous pourrions discourir sur Napier, Pascal, Leibniz (XVII^e siècle), précurseurs du calcul automatique, sur Charles Babbage (XIX^e siècle), qui construit un dispositif analytique qui est un précurseur de l'ordinateur moderne. Je ne résiste pas au plaisir de parler d'Augusta Ada King comtesse de Lovelace plus connue sous le nom d'Ada Lovelace, fille de Lord Byron « sainte protectrice de tous les programmeurs ». Elle décrit de manière précise la machine analytique de Charles Babbage et donna un programme destiné à être exécuté par celle-ci. De ce fait elle est considérée comme la première programmeuse du monde. Le langage de programmation orienté objet ADA a été nommé de la sorte en son honneur.

Pour comprendre le sens profond d'un calcul, on a besoin de saisir son déroulement. À la fin du XIX^e siècle tout s'accélère. George Boole, Auguste De Morgan, et d'autres jettent les bases d'une « algèbre logique ». Giuseppe Peano axiomatise l'arithmétique. La logique héritée de la Grèce antique n'est plus suffisante, elle ne permet pas de définir de manière rigoureuse les nombres. Nous sommes à l'orée du XX^e siècle, une crise des fondements ébranle l'édifice des mathématiques.

LES FONDEMENTS DES MATHÉMATIQUES REVISITÉS.

Le logicien Gottlob Frege caractérise une théorie mathématique comme devant être :

- cohérente : on ne peut pas démontrer une proposition et son contraire ;
- complète : tout énoncé peut être soit démontré, soit réfuté ;
- et décidable : tous les énoncés de la théorie peuvent être testés grâce à une procédure.

Environ à la même période Georg Cantor met au point la théorie des ensembles. Ces théories portent en elles des paradoxes qui sont mis en évidence grâce à des assertions relativement simples. Par exemple le paradoxe du barbier de Bertrand Russell : le barbier du village qui est un homme du village rase tous les hommes du village qui ne se rasent pas eux-mêmes. Question : le barbier se rase-t-il lui-même ou pas ? En fait aucune proposition ne peut exprimer quelque chose au sujet d'elle-même. D'autres paradoxes existent également. Les mathématiques sont un géant sur des pieds d'argile, il faut vaincre ces contradictions.

En ce début du XX^e siècle de grands scientifiques vont tenter de répondre au défi de construire les mathématiques sur des fondations solides. Luitzen Egbertus, Jan Brouwer et son élève

Arend Heyting créent la logique intuitionniste, qui n'utilise pas des axiomes qui vont à l'encontre de l'intuition, qui ne contient pas le tiers exclu, (une assertion est soit vraie, soit fausse) et par conséquent qui n'admet pas le raisonnement par l'absurde et qui n'accepte pas l'existence d'objets construits. Celle-ci donne naissance aux mathématiques constructivistes.

Ernst Zermelo et Abraham Fraenkel donnent une axiomatisation de la théorie des ensembles permettant de se soustraire aux paradoxes du type Russell.

David Hilbert propose au 2^e Congrès international des mathématiciens qui se tient à Paris en 1900 une liste de 23 problèmes pour le XX^e siècle. Deux nous intéressent particulièrement.

Le deuxième problème de cette liste est : peut-on prouver la cohérence (au sens défini ci-dessus) de l'arithmétique ? Le dixième : trouver un algorithme déterminant si certains types d'équations (équations diophantiennes : équations polynomiales à plusieurs variables dont les coefficients sont des entiers relatifs et les solutions, si elles existent, sont également entières aux signes près) ont des solutions.

Pour parfaire ce genre de question et asseoir les fondements des mathématiques, Hilbert élabore son fameux programme en 1920. Le but de celui-ci est de fonder les mathématiques sur une armature à toute épreuve : les mathématiques se déduisent d'un ensemble fini d'axiomes qui est cohérent (défini au sens ci-dessus). Puisque les structures finies sont « bien comprises » et ne présentent aucun paradoxe, alors montrons la non-contradiction de la théorie des ensembles, de l'arithmétique, grâce à des structures finies. Au début tout se passe bien. Kurt Gödel démontre en 1929 son théorème de complétude sur le calcul logique. Donc tout va bien pour le programme de Hilbert, c'est en 1931 que cela se gâte : Kurt Gödel publie ses théorèmes d'incomplétudes.

Le premier dit : « Toute théorie mathématique qui contient l'arithmétique est soit incohérente, soit incomplète. C'est-à-dire l'ensemble des assertions vraies est strictement plus grand

que l'ensemble des assertions démontrables. ». Le second affirme : « la cohérence d'une théorie mathématique qui contient l'arithmétique est indécidable à l'intérieur de cette théorie. »

Une théorie ne peut pas se prouver elle-même.

De manière générale nous pouvons dire que tout système axiomatisé, non-contradictoire, suffisamment puissant pour inclure l'arithmétique de Peano ne peut démontrer sa cohérence en s'appuyant sur ses axiomes. Par conséquent on pourrait en conclure que le rêve de Hilbert est voué à l'échec.

En fait les résultats de Gödel qui font appel à des notions informelles de calcul mécanique affirment qu'il y a des énoncés que l'on ne saura ni démontrer ni réfuter dans une théorie donnée, pas en mathématiques en général. Quand une proposition est indécidable dans une théorie, on peut toujours l'ajouter (ou sa négation) comme axiome. Ainsi la réponse au programme de Hilbert n'est pas aussi simple, et probablement pas aussi définitive, sans doute plus compliquée que celle qu'aurait fournie la véracité de ce programme. Faut-il pour autant abandonner le fait qu'une affirmation mathématique est une valeur de vérité universelle ? J'ai l'habitude de donner à mes étudiants une petite analogie : dans la géométrie euclidienne le segment de droite est le plus court chemin entre deux points, ce qui n'est plus le cas à l'échelle de notre bonne vieille terre. De là, la géométrie de la terre n'est pas euclidienne, mais cela n'enlève pas la valeur de vérité du segment de droite.

Revenons aux problèmes traités par des processus automatiques. Nous ne savons toujours pas ce qui est « calculable ». Puisque d'après les théorèmes d'incomplétude « tout ne peut pas être fait », peut-on se limiter à des assertions qui pourraient être « mécaniquement » prouvées. Gödel s'intéresse à cette question, il élabore les fonctions récursives (fonctions qui peuvent être calculées de manière mécanique). Il faut attendre 1936 pour que se dégage la structure

fondamentale des concepts calculatoires. Alan Turing, qui travaillait sur le problème de décision de Hilbert « peut-on décider de manière mécanique si un énoncé est prouvable ou non », définit un modèle abstrait de calcul : les machines de Turing-MT – pour nous cette terminologie sera équivalente à algorithme. Celles-ci décrivent le fonctionnement d'appareils qui allaient devenir les ordinateurs, elles définissent de manière précise la notion d'algorithme. Environ à la même époque Alonzo Church met au point le *lambda calcul* (base des langages de programmation) qui permet de réfuter le problème de décision de Hilbert. Voilà nous y sommes, les acteurs de la pièce sont brossés, l'informatique (théorique) est née. Deux nouvelles branches très actives aujourd'hui en découlent : la théorie de la complexité algorithmique qui formalise le temps de calcul d'un algorithme et la théorie de la calculabilité. Mais nous ne savons toujours pas ce qui est calculable.

On s'aperçoit très vite que tous ces formalismes sont équivalents. Reste maintenant à dégager la notion de « calculable », c'est le rôle de la thèse de Church.

Celle-ci, exprimée par Stephen Kleene, atteste :

« Tout ce qui est calculable est calculable par une machine de Turing (par un algorithme) ».

Le premier mot « calculable » est une notion intuitive. Par exemple : bouger un bras, prévoir un planning, faire une addition, gérer des stocks, sont des activités calculables. Par conséquent il n'existe pas de définition rigoureuse de ce terme générique, alors que la deuxième occurrence est définie de manière précise (au sens de Turing). Cela explique le fait que cette assertion soit une thèse (une opinion générale qu'on s'attache à défendre) et non une conjecture ou une proposition (une assertion qu'on peut démontrer). Nous pouvons reformuler cette phrase : tout ce qui est calculable est

calculable par l'homme et de là, calculable par machine de Turing ou par algorithme. Une question vient à l'esprit : existe-t-il des machines capables de calculer ce que l'homme ne peut pas concevoir. Nous savons qu'il existe en mathématique des objets qui sont inaccessibles à l'entendement humain. À ce niveau on entre dans la philosophie voire la théologie.

Terminons ce paragraphe par quelques éléments de décidabilité algorithmique. Un problème de décision est décidable s'il existe un algorithme qui prend en entrée le problème, et qui donne en résultat soit oui, soit non ; si aucun algorithme n'existe, il est dit indécidable. Tester si un nombre entier positif est pair est décidable. Le fameux problème de la halte est indécidable :

Existe-t-il un algorithme qui prenne n'importe quel algorithme en entrée et qui, en temps fini, renvoie soit oui si cet algorithme s'arrête soit non s'il ne s'arrête pas.

Il est semi-décidable s'il existe un algorithme qui répond oui (mais pas nécessairement non, il peut faire un calcul infini) ; le 10^e problème de Hilbert est semi-décidable.

Les années 1940 donnent naissance aux premiers ordinateurs (au sens d'aujourd'hui), ceux-ci vont, au fil des ans évoluer, devenir plus petits, plus puissants, la mini-informatique apparaît, puis la micro-informatique ; il faut faire communiquer entre eux ces nouveaux appareils : internet se développe, les téléphones portables, les tablettes, font leur apparition, pour arriver jusqu'à nous, aujourd'hui, dans un monde interconnecté. Pour ce faire il a fallu passer de la logique aristotélicienne – développée dans les *Premiers Analytiques* dont le socle est le syllogisme (espèce de syntaxique) et l'analytique (espèce de sémantique), ainsi que la logique stoïcienne qui introduit la notion de proposition, – à une logique plus élaborée et rigoureuse, développée à la base par G. Frege et B. Russell. Pourquoi cette évolution ? Pour au moins trois raisons :

□ L'idée d'asseoir sur un socle logique les mathématiques en fondant celles-ci, comme la géométrie d'Euclide, sur une axiomatique pertinente.

□ En réponse à la philosophie kantienne des mathématiques qui considère que cette science a quelque chose de plus que la logique, l'intuition, c'est-à-dire une notion extra-logique.

□ Le besoin de comprendre les méthodes de raisonnement de l'être humain, précisément en mécanisant les processus de pensée.

Comme nous l'avons vu, la notion de calcul est centrale dans cette progression. Le calcul utilise les nombres dont une définition précise se dégage à partir de la fin du XIX^e siècle. Qu'en est-il du futur ? Un nouveau traitement de l'information est sans doute en train d'émerger : l'informatique quantique. Le travail théorique est fondé : logique, modèle de calcul, algorithmique, etc. Reste à mettre au point le premier ordinateur quantique efficace. Mais surtout, comme nous allons le voir dans le paragraphe suivant, l'intelligence artificielle.

INTELLIGENCE ARTIFICIELLE (IA)

Sans sortir des généralités, nous pouvons dire que l'IA est un ensemble d'algorithmes dont la mise en œuvre permet à des ordinateurs de simuler l'intelligence humaine. Par conséquent, l'outillage algorithmique utilisé est d'une autre nature que celle qui est sous-jacente à l'informatique de type classique. Il permet, par exemple, l'apprentissage, l'auto-modification et l'évolution de certains paramètres.

Il existe deux types d'intelligence artificielle.

L'IA faible : son objectif est de plagier le plus fidèlement possible des comportements humains, comme jouer aux échecs, démontrer automatiquement un théorème, élaborer un emploi du temps, soutenir une conversation. En ayant l'apparence de l'intelligence, mais en étant dénuée des activités intellectuelles supérieures telles que l'émotion, la conscience, la sensibilité, la « machine » semble réagir comme si elle était douée de ces capacités mais cela n'est qu'une simulation.

L'IA forte a tous les attributs de la précédente, mais elle est munie d'une « vraie conscience » et pourrait ressentir de vrais sentiments tels que l'empathie, la joie, la tristesse, ceux-ci n'étant pas un simulacre, mais une émotion réelle.

Nous avons vu précédemment que l'informatique est basée sur les machines de Turing (MT), l'IA a pour fondement l'informatique, ainsi les MT sont en quelque sorte l'archétype de l'intelligence artificielle. Celles-ci ne considèrent que le constructivisme. De là, le raisonnement par l'absurde, c'est-à-dire une preuve où la non-existence entraîne une contradiction ne conduisant pas en soi à une construction de l'objet qui n'est pas pris en compte, de manière générale, par un algorithme. Par conséquent il semble difficile d'élaborer une IA forte. Pour ce faire, il faudrait développer un modèle de calcul qui soit plus puissant – au sens de la possibilité de calcul – que les machines de Turing. Toutes les tentatives qui ont été menées depuis la création des MT en 1936 se sont avérées infructueuses. Ainsi on peut conclure que ce modèle calculatoire est le plus performant pour une informatique basée sur le silicium et les algèbres de Von Neuman.

La loi de Moore indique que « quelque chose » double tous les dix-huit mois, cette chose étant « la puissance », « la

capacité », « la vitesse » de calcul. Bien que cette assertion commence à s'épuiser, elle est toujours d'actualité. L'intelligence artificielle est une grosse consommatrice de ces éléments. Cela permet d'utiliser des outils très sophistiqués tels que les réseaux de neurones artificiels multicouches (RNAM).

Rappelons qu'un neurone artificiel est un système informatique dont le fonctionnement est une simulation rudimentaire du fonctionnement d'un neurone biologique. Celui-ci est schématiquement une cellule nerveuse donnant une réponse après un stimulus. Ainsi, un réseau de neurones artificiels multicouches essaye de copier un cerveau humain. Notons que nous sommes très loin de la complexité cérébrale, néanmoins ce sont des outils qui sont largement utilisés en intelligence artificielle. En raison de l'amélioration des capacités de calcul, ils permettent de traiter les problèmes que l'IA faible est capable d'aborder avec de plus en plus d'efficacité.

INTELLIGENCE ARTIFICIELLE ET SOCIÉTÉ

Au vu de l'augmentation des capacités de calcul, de l'amélioration de la rapidité calculatoire, l'IA va s'imposer dans la société et dans tous les domaines socio-économiques. Quelles sont les conséquences ?

Il est clair que beaucoup de domaines vont être impactés : le e-commerce, la médecine, l'agriculture, la défense, l'éducation, le journalisme, et bien d'autres disciplines seront bouleversées. La plupart des professions seront entièrement redéfinies, cela implique une révolution copernicienne de la société.

Les répercussions vont être très importantes sur les activités humaines, qui jusqu'à présent étaient relativement protégées. Les deux siècles précédents ont vu, sous l'impulsion du « machinisme » les métiers manuels bouleversés, certains ont

disparu. L'intelligence artificielle se confronte, non seulement à des occupations qui ont peu de valeur ajoutée – les nouvelles caisses de supermarché sont capables de scanner, de manière instantanée un panier de produits – mais également à des fonctions plus cérébrales. Une partie de l'intellect utilisée dans un métier considéré comme intellectuel, enseignant, ingénieur, journaliste, médecin, etc., utilise ce que nous pouvons appeler l'« intelligence répétitive ».

Une tâche utilise ce type d'intelligence si elle n'a pas besoin d'imagination ou si elle peut se dérouler sans aucune originalité dans son déroulement, le processus est en quelque sorte prédictible et fixé.

Diagnostiquer un rhume, faire une tarte aux pommes, noter une copie, font partie de ce type de fonctionnement. Ainsi, une grande partie des professions intellectuelles seront modifiées, le robot-journalisme est en pleine expansion ; les cabines de diagnostics médicaux commencent à être utilisées ; la chirurgie robotisée se développe rapidement, (bien que dans ces deux derniers exemples la présence d'un médecin reste indispensable). Bien d'autres spécialités faisant appel à de hautes capacités cognitives vont subir des changements majeurs.

Il y a un grand risque que le travail tout du moins sous la forme que nous connaissons « disparaisse ». Sous l'impulsion de l'automatisation d'une partie de la plupart des occupations professionnelles considérées comme intellectuelles, la société, dans son ensemble, va muter. Donnons quelques exemples : le monde occidental connaît des destructions massives d'emplois dans différents secteurs, notamment dans l'industrie. Celles-ci sont dues aux délocalisations et à l'automatisation grandissante de la plupart des tâches liées à ces activités. Jusqu'à présent, beaucoup d'économistes considéraient que cela n'était pas un réel problème, vu que l'on pouvait former les chômeurs à des métiers plus spécialisés faisant appel à des capacités intellectuelles plus importantes. Si l'IA peut remplacer une partie des emplois à forte valeur cérébrale, il est

illusoire de vouloir préparer des inactifs à des professions qui de toute façon seront rattrapées par les algorithmes.

John Locke base la propriété sur le droit naturel qui est l'ensemble des droits que l'homme possède à sa naissance du fait qu'il appartient à l'humanité. La propriété privée est un droit naturel obtenu au prix du travail et des efforts que l'on dépense pour mettre en valeur celle-ci. Si le travail « disparaît » comment justifier celle-ci, notamment de quelle manière légitimer l'appropriation des moyens de production ?

Les dangers d'un développement anarchique de l'IA sont réels. La surveillance généralisée des citoyens est déjà une réalité. Tout gouvernement, même démocratique sera tenté d'utiliser ces outils. Les banques, les assurances pourraient prendre des décisions basées sur des algorithmes autonomes. Allez expliquer à une personne que son prêt n'est pas accordé, ou qu'il ne peut pas s'assurer parce qu'une machine a dit non.

Nous avons introduit ci-dessus les réseaux de neurones artificiels comme étant des instruments fondamentaux en intelligence artificielle. Un problème important se pose avec ces appareillages, c'est l'effet « boîte noire ». Si on stoppe un tel réseau avant la fin du calcul, on est incapable de remonter dans celui-ci et on ne peut expliquer toutes les étapes calculatoires. En fait, on connaît les entrées, on connaît les sorties, mais entre ces deux phases aucune opération ne peut être explicitée. Cela pose de graves questions d'éthique. Comment faire confiance à une machinerie qui prend une décision alors qu'on est incapable de comprendre et d'en connaître le pourquoi.

Cette incompréhension est due au fonctionnement des RNAM. Les données sont analysées couche par couche, sachant que les données d'entrée du niveau $i + 1$ sont en quelque sorte les sorties du niveau i . Entre les deux strates, nous ne sommes pas en mesure d'analyser les résultats : il y a

une incompréhension totale. Cela n'est pas le cas dans un programme informatique classique où chaque phase est entièrement déterminée. Expliciter chaque pas calculatoire d'un réseau de neurones artificiels multicouches est un enjeu majeur, notamment dans les utilisations sensibles telles que la sécurité ou la médecine.

Par ailleurs, malgré le peu d'affichage des entreprises de l'IA, des décisions erronées sont courantes en intelligence artificielle. Le réseau peut avoir un fonctionnement tout à fait normal et arriver à des décisions mensongères. Ces erreurs résultent d'une part du calcul intensif et de la complexité de ce celui-ci, mais elles émanent également du manque de conscience de la machine, elle apprend sans comprendre. S. M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi et P. Frossard ont montré, dans un article publié en 2017 qu'on peut toujours trouver une perturbation infime, de telle sorte que la probabilité qu'un réseau neuronal profond se trompe dans sa prédiction soit importante. Cela pourrait entraîner des dysfonctionnements notoires dans divers secteurs socio-économiques.

L'étude de la « bêtise artificielle » est en pleine expansion. D'autres périls sont également à craindre : les armes autonomes, la manipulation de l'information déformant la réalité, l'abolition de la vie privée, l'uniformisation des solutions proposées aux problèmes sociaux, etc. Un autre danger important est l'abêtissement de la population. En moins de quinze ans le quotient intellectuel de la population française – et dans une plus grande généralité en Occident – a diminué de 10 points. Cette baisse est liée à l'utilisation intensive de l'informatique dès le plus jeune âge, cela empêche le développement de l'esprit critique et restreint l'apprentissage du raisonnement.

Tous ces risques peuvent être circonscrits au travers d'une éthique de l'intelligence artificielle admise par l'ensemble des

sociétés humaines. Cela pourrait prendre la forme d'une charte régissant les développements et les utilisations de l'IA.

L'intelligence artificielle est un outil puissant et qui deviendra prégnant dans nos sociétés ; le problème est la pertinence des solutions proposées. Une IA apprend mais n'a pas la conscience de cet apprentissage, donc elle n'a pas la faculté de discernement quant à la justesse et au bien-fondé de la réponse proposée. Ce qui doit nous mettre en garde contre une confiance aveugle et naïve en l'IA.

BIBLIOGRAPHIE :

ARENDRT, Hannah, *Condition de l'homme moderne*, Paris, Éd. Calmann-Lévy, Pocket, 1994.

BELNA, Jean-Pierre, *Histoire de la théorie des ensembles*, Paris, Ellipses, 2009.

BERGSON, Henri, *Les deux Sources de la Morale et de la Religion*, Paris, Félix Alcan, 1932.

BRETTO, Alain, *Intelligence artificielle : la réalité & le mythe*, Gap, Éditions DésIris, 2020.

CORI, René et LASCAR, Daniel, *Logique mathématique*, tome 1, Paris, Dunod, 2003.

DURHAM, T., « On Artificial Stupidity », *Computing*, 1985, p. 4-5.

FLORIDI, Luciano, *The Philosophy of Information*, Oxford, Oxford University Press, 2011.

FREGE, Gottlob, *Les Fondements de l'arithmétique*, Paris, Éditions du Seuil, 1970.

VAN HEIJENOORT, Jean, *From Frege to Gödel : A Source Book in Mathematical Logic*, Harvard University Press, 1990.

JULIA, Luc, *L'intelligence artificielle n'existe pas*, Paris, First Edition, 2019.

LOCKE, John, *Traité du gouvernement civil*, London, 1690, édition numérique produite par Jean-Marie Tremblay, Cégep de Chicoutimi.

MOOSAVI-DEZFOOLI, Seyed-Mohsen, FAWZI, Alhussein, FAWZI, Omar et FROSSARD, Pascal, « Universal adversarial perturbations », *Computer Vision and Pattern Recognition (CVPR)* actes de la conférence IEEE, 2017, p. 86-94.

NORTH WHITEHEAD, Alfred, et RUSSELL, Bertrand, *Principia Mathematica*, 3 vol., Cambridge, Cambridge University Press, 1910-1913.

O'NEIL, Cathy, *Weapon of Math Destruction*, New-York, Crown, 2016.

Ouvrage collectif, *Intelligence Artificielle*, Paris, Flammarion Champ-actuel, 2018.

VILLANI, Cédric, « Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne », Mission parlementaire du 8 septembre 2017 au 8 mars 2018 rapport officiel, 2018.

UNE BRÈVE HISTOIRE D'INTELLIGENCE DE CONSCIENCE ET D'ÉTHIQUE

par M. Jean RICHERT

de l'Académie d'Alsace des Sciences, Lettres et Arts

« Il y a donc lieu d'espérer que l'Amérique, d'ici à quelques générations, en produisant autant d'hommes occupés d'ajouter à la masse des connaissances que l'Europe entière, en doublera au moins les progrès, les rendra au moins deux fois plus rapides¹. »

« Je suis plus vieux que le temps parce que je suis conscient². »

« Si l'histoire héroïque et malheureuse de la cybernétique doit nous apprendre quelque chose, c'est probablement qu'à côté de l'esprit de pionnier, la modestie, le doute raisonné et l'attention à la tradition nourrie d'esprit critique sont des vertus indispensables à l'aventure de la connaissance³ »

INTRODUCTION

L'Intelligence Artificielle (IA) est née dans la seconde moitié du siècle dernier. Fille de l'informatique, elle-même descendante de la logique et des mathématiques, elle s'est développée et son application a investi tous les domaines et secteurs des sciences et de la technique jusqu'au temps présent grâce aux progrès fulgurants de la puissance des ordinateurs actuels.

Après avoir évoqué quelques éléments historiques concernant ses débuts et ses développements, le but de la présente dissertation est d'évoquer les principes de base sur

1 CONDORCET, 1847.

2 PESSOA, 2004.

3 DUPUY, 2005.

lesquels est né et repose ce nouveau domaine. Nous aborderons ensuite son contenu et ses potentialités, les développements envisagés dans le futur, ses limitations connues et probables et nous analyserons et discuterons ses effets éthiques sur le comportement des individus et des sociétés, les utilisations potentiellement bienveillantes et malveillantes par les personnes qui la développent et la maîtrisent, les ambitions portées par certains pratiquants de la discipline.

LES PRÉMICES DE L'INTELLIGENCE ARTIFICIELLE : KURT GÖDEL ET ALAN TURING

La première moitié du XX^e siècle a connu des développements fondamentaux considérables dans l'ensemble des sciences. Les sciences physiques ont déclenché le mouvement avec l'élaboration de la théorie quantique (1900-1930) qui décrit la nature et les phénomènes qui régissent le comportement des objets microscopiques constituant l'ensemble de notre univers.

Une seconde révolution s'est annoncée avec les travaux de Kurt Gödel (1906-1978)⁴, un mathématicien féru de logique né à Brünn (aujourd'hui Brno en Tchéquie) et étudiant à Vienne qui était alors un foyer de l'*intelligentsia* européenne. Il émigra aux États-Unis en 1933. Génie dans son domaine, il démontra deux assertions fondamentales qui « secouèrent » l'ensemble de la communauté scientifique, à savoir :

a) toute théorie mathématique contient des propositions dont on ne peut vérifier la validité. La véracité de certaines d'entre elles, que l'on peut croire universelles échappe à une preuve rigoureuse (1930). Il y a une incomplétude non surmontable, à la base des mathématiques qui constituent par conséquent un système qui n'est pas fermé sur lui-même.

4 NAGEL, NEWMAN, GÖDEL, GIRARD, 1989.

b) les mathématiques sont construites sur les lois de la logique mais il peut être impossible de prouver la consistance de certaines assertions c'est-à-dire de fixer l'absence de contradiction entre certaines propositions.

Cette seconde partie a été démontrée à la même époque par John von Neumann, un mathématicien d'origine hongroise installé, lui aussi, aux États-Unis.

Pour arriver à ces résultats Gödel utilisa une méthode d'approche révolutionnaire qui consistait à transposer les propositions de la logique en expressions arithmétiques sur lesquelles on peut appliquer des transformations de type calculatoire appelées « algorithmes », des procédures manipulatoires qui permettent de résoudre des problèmes. Pour exemple, l'algorithme le plus simple est celui de l'addition d'entiers naturels.

Ces résultats s'avèrent très décevants pour la communauté des mathématiciens mais la méthode introduite et utilisée par Gödel fut une source d'inspiration pour un jeune anglais, Alan Turing (1912-1954). En effet, l'utilisation d'une approche calculatoire dans le cadre de la logique lui suggéra l'idée de concevoir une machine virtuelle (non matérielle) qui « calcule » sur une bande de papier des propositions logiques formalisées en termes de nombres binaires (des 0 et des 1) en lieu et place du système décimal ordinaire. Le principe en est le suivant : la bande avance dans la machine et son action est gouvernée par un algorithme qui génère le calcul. L'algorithme fait avancer la bande et son action permet de distinguer une proposition vraie d'une proposition fausse. Si cette dernière est vraie la machine s'arrête au bout d'un certain temps, si elle est fausse la machine poursuit sa marche à l'infini dans le temps. La validité de la proposition est décidable dans le premier cas et non décidable dans le second⁵.

5 BODEN, 2021 ; GUERRAOUI, NGUYÊN HOANG, 2020 ; ANDRÉ, 2012.

La structure fondamentale de l'ordinateur moderne était née : le ruban correspond à la partie matérielle (partie *hardware*) et l'algorithme créé par un opérateur à la partie opératoire (partie *software*).

Cette procédure géniale permit à la fois de « vérifier » le contenu des théorèmes de Gödel et de von Neumann et ouvrit la voie au calcul informatique développé ultérieurement. Il restait cependant à savoir s'il pouvait exister une « hyper-machine » qui transcenderait les capacités de la machine décrite et qui serait capable de surmonter la question de l'indécidabilité : cette tentative fut l'objet d'une thèse préparée par Turing en 1936 à Princeton sous la direction d'une autre sommité de l'époque, Alonzo Church. Il n'en fut rien, la réponse fut négative et la situation en resta là.

CYBERNÉTIQUE ET SCIENCES COGNITIVES

Les résultats évoqués suscitèrent de nombreuses réflexions dans le monde du savoir de l'époque. Une activité intense se développa aux États-Unis pendant la période de guerre en Europe. C'est ainsi qu'à partir de 1942 se mirent en place à New-York les conférences Macy (1942-1953) conçues comme un lieu d'échanges d'idées en vue de définir un rapprochement entre les différents domaines du savoir, les sciences cognitives destinées à faire progresser la connaissance de l'homme et de ses potentialités. Elles rassemblèrent des sommités mondiales actives dans ces domaines dans le but de progresser et d'arriver à une définition générale et cohérente d'une science du fonctionnement de l'esprit. Parmi les membres de ce groupe figuraient des mathématiciens et des physiciens célèbres, la plupart d'entre eux étaient des européens réfugiés aux États-Unis, John von Neumann déjà cité plus haut, Norbert Wiener et Claude Shannon, personnalités connues pour leurs travaux et intéressées par la réalisation de ce qui allait devenir plus tard l'ordinateur digital opérationnel. Dans le contexte des activités de cette conférence on peut citer plus particulièrement celle

organisée par von Neumann en 1948 dont le sujet était : « *The general and logical theory of automata* »⁶ (La théorie et logique générale des automates) qui laissa des traces de première importance dans le milieu des sciences de l'informatique naissante⁷. C'est ainsi que naquit la « cybernétique », dénomination d'origine grecque trouvée chez Platon, signifiant « pilotage d'un navire » et utilisée pour la première fois par Ampère (1834) pour désigner le gouvernement des hommes.

Cependant, au début des années 1960, les personnalités participant aux conférences Macy durent se rendre à l'évidence que leurs centres d'intérêt et leurs méthodes d'approche ne se recouvraient pas assez pour assurer un dialogue productif. Ils se séparèrent en groupes de recherche distincts, ceux qui s'interrogèrent sur les mécanismes de la vie, le cerveau, et ceux qu'intéressait « l'esprit ». La cybernétique se divisa alors en trois branches distinctes identifiées par Marvin Minsky, informaticien et roboticien américain considéré comme un précurseur d'une science nouvelle, l'Intelligence Artificielle (IA). Celui-ci construisit en 1951 la première machine dite « de neurones en interaction aléatoire » au MIT (*Massachusetts Institute of Technology*), l'ancêtre des dispositifs électriques complexes développés et progressivement perfectionnés dans le cadre de l'IA. Suivant ses suggestions on identifia alors les domaines suivants :

La « simulation cognitive » : il s'agissait de reproduire pour le maîtriser, le plus souvent sur l'ordinateur, le fonctionnement d'un système, un résultat obtenu par ailleurs, par exemple l'énoncé d'un théorème mathématique connu⁸. Cette approche était essentiellement un programme informatique permettant de construire une théorie plus précise

6 VON NEUMANN, 1996, VON NEUMANN, 1998.

7 DUPUY, 2005.

8 *Id.*

à partir de modèles intuitifs plus ou moins clairs, de tester la consistance logique, la complétude, la « clarté » de ces idées intuitives ou de déchiffrer le contenu de données empiriques dans le cadre de prises de décision.

La « seconde cybernétique », une théorie des « systèmes auto-organiseurs »⁹ devant permettre de voir comment des systèmes évoluent et créent de nouvelles structures complexes à partir d'un ordre initial, la « morphogénèse ». Elle visait à l'élaboration d'une méthode de description universelle commune aux différents domaines de la science.

L'IA, objet central de la présentation, décrétée science cognitive, se rajouta aux sciences cognitives connues au XX^e siècle (la philosophie, la psychologie, les neurosciences, la linguistique, l'anthropologie) : une approche mécaniste, digitale, calculable au sens de Turing pour répondre à des questions concernant la connaissance.

Tous ces développements menèrent à la fin du siècle passé à la création d'organismes de recherche interdisciplinaires dédiés à ce type de recherche. Pour exemple, en France le CNRS créa l'Association pour la recherche cognitive et l'Institut des sciences cognitives du CNRS (1992) comportant un secteur de science fondamentale de la cognition et un secteur applicatif de l'ingénierie appelé « cognitique ».

L'INTELLIGENCE ARTIFICIELLE

C'est John Mc Carthy, un chercheur en informatique au MIT et à l'Université de Stanford aux USA, qui a forgé en 1956 le concept et le nom d'« Intelligence Artificielle » (AI en anglais) lors d'une réunion d'informaticiens à *Dartmouth College* en Angleterre pour distinguer cette approche de celle d'une autre figure de la cybernétique, Norbert Wiener,

9 ATLAN, 1979.

mathématicien célèbre et esprit universel intéressé par les problèmes liés à la communication, un des animateurs des conférences Macy après la Seconde Guerre mondiale.

Le mot « intelligence » est polysémique, il peut caractériser à peu près toutes les activités du vivant, des hommes, des animaux, des plantes. L'intelligence présente différents degrés, c'est chez l'homme qu'elle est la plus développée. Parmi les formes les plus élémentaires on peut citer la distinction des formes et des objets, des sons et de manière générale, de tous les phénomènes perçus dans l'espace et dans le temps, leur mémorisation et leur reproduction.

Les potentialités de l'IA se situent à ce niveau. La qualification « artificielle » correspond au fait qu'elle se développe en dehors de l'homme bien qu'elle soit issue du cerveau de l'homme, car elle est un « enfant » de la machine de Turing, une procédure de traitement de données dans le but de réaliser une tâche donnée. Comme on l'a vu, elle résulte de la combinaison d'un support matériel, l'ordinateur moderne et d'un ou plusieurs algorithmes de calcul utilisés en vue d'accéder au but cherché.

La miniaturisation vertigineuse de l'électronique à partir des années 1960 a permis d'accéder à l'utilisation effective de l'IA grâce à la réalisation d'ordinateurs de plus en plus rapides et pourvus de capacités de plus en plus grandes, performance liée aux progrès fulgurants de la miniaturisation. En 1951 un transistor avait une largeur de 10 mm, aujourd'hui 100 000 transistors tiendraient dans le diamètre d'un cheveu.

Le principe d'exécution d'une tâche donnée repose sur un concept virtuel élémentaire appelé « perceptron » imaginé en 1957 par Frank Rosenblatt, un psychologue du Laboratoire d'aéronautique de l'Université Cornell (USA). Ce type de concept qui est central est une imitation virtuelle très primitive du cerveau par sa structure et son fonctionnement. Il est constitué d'un ensemble de perceptrons constitués de

« neurones » (ici représentables sous forme de « fils » de transmission) qui relie des « synapses » (des « centres de connexions) » dans lesquels évoluent des « courants » (des transferts d'informations) gouvernés par un programme informatique. Le fonctionnement est piloté par ce programme implanté dans un ordinateur qui impose les opérations à effectuer. Le programme est codé sous forme d'instructions logiques en termes de langages de programmation. La structure dynamique ainsi conçue, un ordinateur et un programme qui pilote des informations suivant le principe décrit, est l'outil utilisé en pratique depuis le milieu du XX^e siècle jusqu'à nos jours. Dans sa structure, ce type d'architecture peut être vu comme une représentation très simpliste de la structure du cerveau humain.

L'exécution d'une tâche avec ce dispositif commence par une phase dite d'« apprentissage » lors de laquelle le programme d'exécution est conditionné de telle sorte que les données de départ entrant dans le circuit produisent un résultat final à la sortie qui soit aussi proche que possible des données d'entrée : c'est la phase qui doit fixer les quantités qui entrent dans le calcul menant de l'entrée à la sortie. Le travail d'ajustement nécessite généralement un très grand nombre de répétitions.

Cette phase étant effectuée (le *learning*, l'apprentissage), le programme de reproduction ainsi généré pourra servir à effectuer des applications ultérieures. En utilisant cette façon de procéder on peut en effet imaginer de très nombreuses applications. Pour en citer quelques-unes : la reconnaissance d'empreintes digitales, de filigranes de passeports, la lecture et les corrections orthographiques et grammaticales, la reproduction de sons, la robotique industrielle et domestique, la gestion, la médecine (entre autres l'intervention chirurgicale à distance, l'orthopédie, l'endoscopie), la logistique du transport et de la circulation, la banque et les finances, l'activité juridique, le renseignement civil et militaire, les arts

et la culture, plus récemment la recherche fondamentale dans le domaine des sciences de la nature, la biologie, les sciences de l'atmosphère et de l'univers.

Dans l'état du savoir-faire actuel l'essentiel du fonctionnement bien maîtrisé se réduit à deux domaines : le *Machine Learning* (ML, apprentissage machine) et le *Deep Learning* (DL, apprentissage profond). La « machine » est l'ordinateur et les instruments à l'entrée et à la sortie qui y sont raccordés. DL est une extension de ML.

La maîtrise de ces procédures a conduit récemment à des résultats remarquables :

– Des travaux très récents en mathématiques fondamentales concernant le domaine de la topologie (l'étude des formes d'objets) et de la théorie de la représentation (l'étude des symétries) se sont appuyés sur l'IA pour chercher et découvrir des liens jusqu'alors inconnus entre différentes propriétés des objets et concepts étudiés. De façon remarquable ces recherches ont mené à l'énoncé de nouveaux théorèmes, le « graal » en mathématiques¹⁰.

– Un autre exemple impressionnant très récent est l'application du DL à l'élaboration d'un programme permettant de comprendre en le reproduisant virtuellement, le mécanisme complexe de repliement des protéines¹¹, une performance de la recherche fondamentale en biologie reconnue comme exceptionnelle dans les revues *Science* et *Nature* pour l'année 2021.

– Fin 2021 les principes de l'IA ont été appliqués dans le domaine technique de la génération d'énergie nucléaire par la méthode dite de la fusion, la procédure qui se déroule dans le soleil, plus « propre » sur le plan de l'émission de radioactivité et plus puissante que celle de la fission utilisée actuellement¹².

10 DAVIES et coll., 2021.

11 BAECK, 2021.

12 HIRLIMANN, 2018.

Une expérience de déclenchement du processus et de sa maintenance pendant quelques secondes a été effectuée récemment dans le laboratoire international JET (*Joint International Torus*) situé en Angleterre. L'expérience a été pilotée par un programme d'IA, une extension du *Deep Learning*, qui permet de générer, tout au long du fonctionnement, une adaptation automatique optimisée des conditions d'évolution de l'expérience par le programme lui-même, le « RDL » (*Reinforcement Deep Learning*, en français : programme d'apprentissage profond renforcé)¹³.

INTELLIGENCE ET CONSCIENCE HUMAINE

L'histoire montre que les procédures décrites ci-dessus font explicitement référence au concept d'« intelligence » qui trouve toute sa complexité dans le fait que l'homme et les animaux d'ordre supérieur possèdent un cerveau dont la constitution et les capacités fonctionnelles sont fascinantes et largement incomprises. Théologiens et savants ont, au cours des siècles, essayé de localiser cette propriété et de définir ses fonctions. Jusqu'au début du XIX^e siècle le cerveau humain fut considéré comme étant le siège de l'âme¹⁴ et de l'esprit identifié à l'âme, de l'esprit « décomposé » par Descartes en trois parties (pensée, imagination et mémoire)¹⁵. Dans sa définition matérialiste moderne l'esprit est la manifestation de phénomènes physiologiques régis par les lois de la physique. Pour certains pratiquants des sciences cognitives contemporaines, il doit exister entre le cerveau et l'esprit le même rapport que celui qui existe entre la partie matérielle (*hardware*) et la partie logique (*software*) dans le monde de l'informatique.

13 DEGRAVE, FELICI, BUCHLI *et al.*, 2022.

14 BOUCHET, MASSON, 1980.

15 DESCARTES, 1637.

Mais l'homme possède une qualité supplémentaire par rapport à l'animal, à savoir celle qui fait qu'il est conscient de son existence et sait créer. Selon Étienne Balibar¹⁶ c'est le philosophe John Locke qui a introduit le concept moderne de conscience dans un écrit intitulé *Essay on Human Understanding* édité en 1690 et traduit en français par Pierre Coste¹⁷ en 1700 :

« Car, puisque la conscience accompagne toujours la pensée et que c'est là ce qui fait que chacun est ce qu'il nomme soi-même et par où il se distingue de toute autre chose pensante : c'est aussi en cela seul que consiste l'identité personnelle ou ce qui fait qu'un Être raisonnable est toujours le même¹⁸. »

La science moderne a développé quantité d'études concernant le concept de « conscience » en introduisant différentes définitions qui dépendent de l'usage qui en est fait dans différents domaines du savoir, essentiellement ceux de la neurologie, de la psychologie, de la philosophie, de la psychanalyse et de la neurobiologie. Les psychanalystes par exemple¹⁹ distinguent différents types de conscience, à savoir la conscience perceptive, celle de l'action, la conscience de sa propre vie et celle de l'éveil. Nous serons amenés à concentrer notre attention sur la conscience perceptive dans le cours du développement du sujet.

Le passage de l'IA à l'IAG : un « saut »

Partant de l'état des performances actuelles de l'IA, l'imagination, la curiosité et la démarche continue vers le « plus » a poussé la recherche à essayer de concevoir des outils qui ne sont pas uniquement dédiés à des tâches d'exécution mais aussi à s'attaquer à des tâches non répétitives

16 BALIBAR, 1998.

17 BOUILLET, CHASSANG (dir.), 1878.

18 *Essai philosophique concernant l'entendement humain*

19 LECHEVALIER, 2007, p. 437-454.

nécessitant un degré « d'intelligence » plus élevé. Cette étape est celle de l'IAG (Intelligence artificielle généralisée) ou IA forte.

Comment peut-on caractériser une IAG par rapport à l'IA ? Cette question de fond est abordée de façon très claire par Alain Bretto²⁰ qui énumère un ensemble de concepts et de propriétés qui définissent et permettent de distinguer l'IAG de l'IA : l'interprétation de l'information à disposition, c'est-à-dire son sens et sa signification, la pensée, l'intelligence, la raison, l'intuition, l'inférence inductive et la conscience définissent les propriétés liées à l'humain en tant qu'être pensant. Ces propriétés sont comparées à celles qui définissent une entité travaillant au niveau de l'IA (une machine, un ordinateur). La plupart d'entre elles recouvrent celles de l'humain. Ce qui différencie fondamentalement la machine de l'humain, la notion de conscience, ne peut être présent dans l'ordinateur car la machine ne peut jamais traiter un processus autoréférentiel, c'est-à-dire se reconnaître elle-même et dire : « je suis conscient d'être conscient », « je peux me définir moi-même ». Ceci montre qu'il doit exister au niveau du cerveau humain une structure et des mécanismes fort différents du type « perceptron » et de ses performances, le principe suivant lequel se déploie l'activité de l'ordinateur. À priori le « pas » vers une IAG n'est pas réalisable, en tout cas si l'on s'appuie sur l'argumentation présentée plus haut. Il n'est pas difficile de s'en douter aussi quand on se rapporte, dans une approche physiologique évoquée ci-dessous, aux connaissances acquises et non définitives sur la structure et le fonctionnement du cerveau humain.

La conscience de soi-même

Il est empiriquement possible de distinguer différents types de niveaux de conscience : ainsi la conscience des contrastes (clair/obscur) ne devrait pas poser de problème quant à son

20 BRETTO, 2021.

intégration dans le traitement IAG. Par contre il n'en est pas de même pour ce que les philosophes appellent la « conscience perceptive », la conscience de soi-même, de ses sentiments, de son histoire (ontogénie), de sa conviction d'exister face à l'autre qui pense que son interlocuteur pense lui-même²¹.

Sur le plan biologique, la localisation de la conscience est une question ouverte. La connaissance intime des pensées, la vie affective, les perceptions seraient « hypothétiquement localisées de façon diffuse dans le « cortex ». Le schéma le plus convaincant serait celui du neurologue américain Marsel Mesulam, neurologue américain de la *Northwestern Feinberg School of Medicine* de Chicago²². Il apparaît que le cortex frontal qui est aussi le siège des fonctions exécutives joue effectivement un rôle essentiel, il peut s'opposer aux comportements automatiques et permettre l'expression de comportements contradictoires.

David Chalmers, philosophe et spécialiste reconnu en sciences neuronales, admet que ce problème est sérieux et incontournable, c'est l'objet de grandes controverses philosophiques qui sortent du cadre de la réflexion scientifique : peut-on parler d'états de conscience de la machine ? Aucune théorie n'est satisfaisante, la véritable intelligence implique la compréhension. Une quantité impressionnante d'idées et de propositions plus ou moins convaincantes, tant philosophiques que scientifiques ont été avancées mais il semble bien qu'aucune ne soit satisfaisante pour pouvoir envisager un moyen de surmonter le passage de l'IA à l'IAG²³.

Pour ce qui concerne le concept de vie²⁴ il n'existe pas de définition universellement acceptée mais neuf caractéristiques peuvent être mentionnées : l'auto-organisation, l'autonomie,

21 LECHEVALIER, 2007, p. 437-454.

22 MESULAM, 2008.

23 BODEN, 2021.

24 *Ibid.*

l'émergence, le développement, l'adaptation, la réactivité, la reproduction, l'évolution et le métabolisme. Les huit premières pourraient être décrites (instanciées) en langage informatique, la dernière ne le peut pas. Aucun ordinateur, aucun robot ne peut métaboliser car le métabolisme est un processus complexe qui utilise des substances biochimiques et des échanges d'énergie pour assembler et entretenir un organisme. Une IAG liée à la vie paraît par conséquent impensable. Et pourtant certains chercheurs ont déjà poussé plus loin...

Quelques éléments physiologiques concernant le cerveau humain

Dans un travail datant du début du siècle (2004) le neurophysiologiste américain Prix Nobel de Médecine Gerald M. Edelman²⁵ a développé une théorie en soi très détaillée quoique toujours incomplète sur la question. Le point important est la distinction qui est faite entre une conscience primaire que l'homme partage avec les animaux supérieurs et une conscience d'ordre supérieur réservée à l'humain. La naissance de ces types de conscience est explicitement décrite en termes de connexions spécifiques entre différentes zones du cerveau reliées entre elles par des circuits de neurones. Il reste cependant dans la description détaillée d'Edelman des inconnues importantes reconnues par leur auteur, elles concernent le fonctionnement exact des mécanismes électro-physiologiques actifs.

Ce fait n'est pas étonnant. Le cerveau humain est et reste un vaste objet d'étude. Il s'agit de comprendre le fonctionnement d'un organisme contenant près de 86 milliards de neurones macroscopiquement assez bien connu (voir les travaux évoqués ci-dessus) mais, comme on peut s'en douter, un chantier particulièrement complexe car structure et

25 EDELMAN, 2008 ; EDELMAN, TONINI, 2000.

comportement fonctionnel sont intimement liés. Pour aborder cette question il est indispensable d'approfondir la connaissance de cette structure. Les études actuellement envisagées et en cours utilisent des ordinateurs particulièrement puissants qui ont pour but de permettre le passage d'une investigation de la connectivité dans la masse cérébrale à l'échelle micrométrique (1/10 000^e cm) à une investigation à l'échelle nanométrique (1/10 000 000^e cm). Cette étude a été faite très récemment sur un être microscopique (le *Caenorhabditis elegans*) qui possède 302 neurones et un modèle liant structure et fonctionnement a été développé. La recherche est encore loin de l'étude du cerveau humain²⁶!

Quelques réflexions philosophiques contemporaines concernant l'intelligence supérieure

La question de la conscience a une histoire très riche dans le domaine de la philosophie, histoire qui n'épuise pas, et de loin, le sujet²⁷. En effet, si Descartes était sûr de son fait en affirmant l'expression universellement connue « Je pense, donc je suis », certains philosophes postérieurs se sont attachés à contester la solidité de cette affirmation et ceci jusqu'à l'époque actuelle. Ainsi Hume a-t-il soutenu la thèse selon laquelle notre conscience de nous-mêmes ne serait qu'une fiction.

La question a été reprise dans un travail récent²⁸ dans lequel l'auteur confronte la position de Descartes et celle de Wittgenstein pour montrer ce qui est appelé le « caractère insaisissable » du sujet. Le philosophe Paul Ricoeur dit que ce qu'il appelle le « Soi » contient une part de contingence, qu'il est « le produit de l'histoire que nous racontons à nous-mêmes

26 AMUNTS, LIPPERT, 2021, p. 1054-1055.

27 BESNIER, 2020.

28 GILLOT, 2020.

sur nous-mêmes et puise à la mémoire de ce que nous avons été »²⁹.

Ces points de vue du bien-fondé de la notion de subjectivité, propriété de l'homme pensant ont été explicitement ou implicitement repris par certains théoriciens modernes des sciences cognitives. En effet, pour certains philosophes et pour certains praticiens de l'IA il est difficile d'accepter que l'homme puisse être bloqué par le problème insurmontable de la conscience de soi dans la recherche d'une issue à la réalisation d'une IAG.

La réflexion moderne concernant la conscience a été menée dans de nombreuses discussions et confrontations entre diverses approches de pensée menées par des chercheurs reconnus dans différents domaines, en particulier Daniel Dennett, David Chalmers et John Searle, Allen Newell, Herbert Simon, Aaron Sloman et Hillary Putnam, des personnalités du monde de la philosophie, des sciences cognitives, de la psychologie, de la linguistique, des sciences sociales, des mathématiques et de l'économie.

Pour Dennett le sentiment d'identité personnelle, un « moi biologique » se comprend par le fait que les vivants possèdent l'instinct élémentaire de la distinction entre le « moi » et « l'autre ».

Concernant l'esprit, Putnam soutient que celui-ci est ce que fait le cerveau et Sloman pense que l'esprit peut être conçu comme une « machine virtuelle », non matérielle, un système de traitement de l'information que l'informaticien a en tête lorsqu'il crée un programme et que l'utilisateur s'approprie lorsqu'il l'utilise³⁰.

Pour Newell et Simon l'esprit-cerveau est un « système de symboles physiques » (SSP) disposant de moyens pour effectuer une action intelligente. Un SSP qui effectuerait de

29 BESNIER, 2020.

30 BODEN, 2021.

bons calculs serait effectivement capable d'être intelligent, une IA « forte ».

Par contre, pour Searle, ce système serait dans l'erreur : le calcul formel tel qu'il est conçu ici ne peut pas fournir « l'intentionnalité », c'est-à-dire la signification ou la compréhension. Les programmes conçus par des humains sont de type syntaxique et sémantiquement vides. L'intentionnalité humaine pourrait être due à une propriété spécifique du cerveau, à savoir le fait que ce dernier fonctionne sur un substrat physique, les neuro-protéines constituant un support qui permet l'exécution de calculs que la matière inerte (métal et silicium) ne peut réaliser. La spécificité du support matériel complexe serait à lui tout seul le substrat permettant à l'humain de réaliser son existence et ses performances intellectuelles. Où est l'explication définitive ?

DES PROJECTIONS AMBITIEUSES MAUVAIS RÊVES OU RÉALITÉS LIÉS À LA FOI DANS UNE IAG D'ORDRE SUPÉRIEUR

Conceptions et réalisations actuelles, projections vers le futur

Faisant fi de la question de l'obstacle lié à la nature de la conscience spécifique à l'homme, croyant fermement à l'évolution exponentielle des idées et des techniques liées à l'informatique, la course consciente (ou inconsciente) pour atteindre les performances du cerveau humain ont mené un certain nombre de spécialistes et de savants plus ou moins sérieux à l'idée que les développements en cours pourraient mener à un basculement, un moment précis dans l'histoire de la cybernétique où le potentiel de l'IA atteindrait un niveau de performance égal à celui du cerveau humain. Ce phénomène hypothétique, déjà évoqué par Condorcet (!), correspondrait à un moment où la croissance technologique aurait mené à une situation incontrôlable et irréversible. L'homme aurait en face

de lui un agent intelligent capable d'améliorer ses propres performances et d'acquérir ainsi une « super-intelligence ».

Parmi les premières personnes qui ont songé à cette issue on trouve des personnalités aussi importantes que John von Neumann et, étonnamment, Stephen Hawking qui s'en est cependant dédit ultérieurement. En 2012-2013 certains spécialistes estimaient que la probabilité pour que l'IAG (Intelligence Artificielle Généralisée) soit au point dans la décade 2040-2050 était de l'ordre de 50 %.

Cette situation singulière a cependant soulevé un certain scepticisme, il paraissait difficile d'imaginer concrètement ce que serait le monde « post-critique » évoqué ci-dessus. Il pourrait *a priori* être concevable que la génération d'une IA en concurrence avec l'humain au niveau de la rapidité d'exécution soit envisageable. Cette rapidité pourrait être générée à la fois par l'usage d'un support matériel conséquent pour les machines et un perfectionnement des algorithmes qui pilotent le développement des opérations. Elle reste cependant plus qu'hypothétique. En effet en 2012 la simulation d'une seconde de 1 % de l'activité cérébrale réelle impliquant 1,73 milliard de neurones et 1,4 million de milliards de synapses nécessitait quarante minutes de temps d'exécution sur un ordinateur de la dernière génération³¹! Toutes ces réflexions et projections ont généré commentaires et critiques de la part de philosophes, de psychologues et finalement aussi de scientifiques sceptiques.

L'Intelligence Artificielle Surhumaine (IAS)

Un autre aspect fascinant dans l'évolution des potentialités humaines est celui de l'apparition de la conscience perceptive dans un contexte extra-humain. Le génie inventif de l'homme moderne ne s'arrête pas à la mise au point d'une IAG : l'IAS doit nous porter au-delà des potentialités humaines. Cette dernière étape dans l'évolution des machines menant

31 *Ibid.*

au « surhumain » est supposée naître de la « singularité » citée plus haut. Cette singularité correspondrait à un changement plus ou moins radical et brusque comme c'est le cas d'un système physique quand il passe d'un état de départ dans un état aux propriétés radicalement différentes (un exemple simple est celui du passage de l'état solide à l'état liquide sous l'action d'un paramètre physique extérieur). Ici la singularité doit correspondre au moment où les potentialités de la machine deviennent « supra-humaines ».

Cette vision a ses initiateurs et ses croyants, en particulier Ray Kurzweil, un inventeur futuriste américain travaillant dans divers domaines d'application des technologies modernes et lauréat de prix célèbres aux États-Unis. Pour lui, la mise en évidence de cette singularité correspondra au moment où la pensée humaine et l'existence biologique « fusionneront » avec la technologie pour faire un monde où il n'y aura plus de distinction entre l'homme et la machine ou entre les réalités physique et virtuelle. Le monde de l'intelligence artificielle des « S-croyants » se divise en pessimistes (qui y voient un danger mortel) et les optimistes comme Kurzweil. C'est dans ce contexte que se sont développées les thèses transhumanistes et la possible réalisation de l'état d'immortalité.

Comprendre le comportement du cerveau

À défaut de réalisations concrètes allant dans le sens du dépassement des potentialités de l'IA par une possible ou impossible IAG, la recherche poursuit l'étude approfondie du cerveau, son fonctionnement et ses pathologies. Dans le cadre de la WBE (*Whole Brain Emulation*)³² la recherche essaie de simuler le fonctionnement de ses neurones, leurs connexions et leurs capacités de traitement de l'information. Ces projets sont souvent reliés à la recherche concernant les maladies mentales. Tel est le cas de la maladie d'Alzheimer pour citer

32 SANDBERG, BOSTROM, 2008.

la plus connue. Un certain nombre d'autres projets allant dans ce sens ont été développés à travers le monde. On peut citer le projet BRAIN lancé par le Président Obama et des soutiens financiers privés dont le but est d'établir une carte de la connectivité du cerveau de la souris qui doit être suivie par l'établissement de celle de l'homme. Cette initiative a connu des tentatives antérieures en Allemagne et au Japon avec le projet NEST (Technologie de Simulation Neuronale) pour développer un ordinateur spécifique appelé « K ». Simuler le fonctionnement du cerveau en vraie grandeur avec un ordinateur est une gageure dont on peut estimer l'ampleur comme cela a été évoqué ci-dessus. Plus modestement, mais efficacement, se développe l'adaptation de prothèses artificielles aux parties défaillantes du corps humain.

L'IMPACT ÉCONOMIQUE SOCIÉTAL ET MORAL DE L'IA ET DE SON DÉVELOPPEMENT

Tout objet, toute réalisation concrète ou virtuelle, tout développement technique que l'esprit humain peut concevoir, réaliser, mettre en application pratique a, comme une pièce de monnaie, un endroit et un envers. L'homme n'échappe pas aux effets positifs et négatifs de ce qu'il peut imaginer et mettre en pratique pour lui-même et les autres. L'humanité a vécu, vit et continuera à vivre cet état de fait dans le futur. Si ce fait n'est pas une loi de la nature, il est bien inscrit dans l'histoire.

En conséquence, il est indispensable de se poser des questions quant aux conséquences générées par les développements des technosciences actuelles et en gestion qui régissent à peu près tous les domaines de la vie actuelle. Le concept de progrès n'est pas forcément lié à celui de bonheur. Nos ancêtres lointains, les Sapiens, ont expérimenté cette réalité quand ils sont passés de l'état de chasseurs-cueilleurs nomades à celui d'agriculteurs sédentaires³³. Une plus grande

33. HARARI, 2012 ; HARARI, 2017.

assurance de survie due à la culture des céréales engendrant plus d'efforts physiques et moins de liberté a suivi une période de sécurité de survie alimentaire plus aléatoire mais une liberté de déplacement et de fatigue physique moins grande. Où en sommes-nous aujourd'hui et où risquons-nous d'aller ? Si l'IAG et l'IAS paraissent loin ou même hors de portée pour créer des problèmes, l'existence de l'IA amène des progrès constants pour ce qui concerne le bien-être de la personne mais crée aussi des problèmes et de nouveaux soucis sérieux aujourd'hui et pour le futur. Le catalogue des questions ouvertes est vaste et couvre à peu près tous les domaines de l'activité humaine.

Les effets économiques à long terme et l'impact de l'utilisation des procédures d'IA sur l'emploi ne sont pas clairs. Au vu de l'état actuel de la situation, on peut dire que les progrès techniques à venir déboucheront certainement sur l'introduction intensive de robots programmés ; ils risqueront de nuire à l'emploi de « cols bleus » mais augmenteront la taille du monde des « cols blancs ». En fait l'utilisation de robots n'aura d'issue positive que si les gains de productivité attendus sont redistribués sur l'ensemble de la population. Il n'est pas clair que la théorie de Schumpeter³⁴ (l'innovation par la destruction, la création de postes de travail après destruction d'autres postes) s'appliquera telle qu'elle est pratiquée actuellement car la requalification du personnel pourrait être plus difficile, le niveau exigé par le travail pouvant être très élevé. Néanmoins les prévisions faites dans le cas d'un développement lent des possibilités de l'IA permettent de penser, sauf contretemps tel que celui créé par la Covid, que le nombre d'emplois pourrait encore augmenter dans le futur. Si l'on est optimiste on peut espérer que les risques qui grandissent avec ces nouvelles technologies pourront se stabiliser et même être minimisés.

34 VOLTZ, 2019.

Cependant, en envisageant un développement plus rapide de la technologie et de la programmation menant à une IAG, si tant est qu'elle soit possible, la situation concernant les risques s'avérerait beaucoup moins claire sinon imprévisible. Il est difficile d'imaginer ce qui pourrait se passer au cas où les machines et leurs performances atteindraient un niveau particulièrement élevé. Dans le cas d'une telle situation, les grands « managers » modernes (Bezos, Thiel et Musk) ont suggéré la création d'une société à but non lucratif qui garantirait un développement responsable de l'IA... La surveillance du développement pourrait se faire par la création d'un secteur public indépendant dédié à une régulation dans la réalisation des procédures. Une entente sur ce projet devrait nécessairement se faire sur le plan planétaire. Il est difficile de croire qu'un tel développement soit envisageable si tant est que nous en arrivions là.

Les développements techniques liés à l'IA pourront certainement contribuer positivement à l'amélioration des performances dans le secteur de la santé, c'est-à-dire permettre des diagnostics plus rapides et plus sûrs, des actes plus pertinents, la mise au point de prothèses de plus en plus adaptées dans de nombreux domaines de la médecine, servir en agriculture, mettre à disposition des moyens de surveillance plus performants pour éviter les catastrophes naturelles, plus généralement pour prévenir les désastres déjà existants et attendus dans le domaine du dérèglement climatique.

Les questions sociétales et comportementales soulevées par l'IA et les technosciences sur l'évolution d'un monde ouvert à la communication en ce début du XXI^e siècle laissent entrevoir un lent mais irrémédiable « glissement » du pouvoir de la gouvernance du politique aux monopoles planétaires et ses conséquences³⁵. Après la Seconde Guerre mondiale la poussée d'abord imperceptible mais toujours croissante du

35 DISCEPOLO, IZOARD, 2021.

pouvoir de « managers » entreprenants a mené à la création de grands groupes économiques. Une catégorie d'individus usant de ses talents est en train de créer une oligarchie monopolistique de patrons de « super-business » ayant en face d'elle une société de consommation qui a d'abord été essentiellement alimentaire mais qui se transforme rapidement, grâce au développement et à l'usage de l'outil informatique, en une société de consommation de tout ce qui est accessible à l'achat, c'est-à-dire tout bien matériel disponible.

Cette évolution a été encouragée par le développement de l'informatique, en particulier dans la Silicon Valley dont les activités alternatives aux programmes de développement des États ont mené au monde des « *Big Brothers* » actuels, les Gates, Page et Brin, Musk, Bezos et autres Zuckerberg.

C'est ainsi que s'est créée une société de communication, celle de la commande par ordinateur, faisant souvent illusion en générant une véritable euphorie de fausse liberté et de fausse puissance chez beaucoup de pratiquants du « téléphone intelligent », en particulier chez les jeunes.

En effet, une conséquence évidente de cet état de fait est l'addiction des individus aux machines qui éloignent les usagers les uns des autres, se prêtant ainsi à une manipulation qui pousse à la consommation, à l'intrusion rampante du monde économique dans la vie privée, une véritable menace sur le sens du mot « liberté » tant revendiqué en ces temps de pandémie. Dans cette évolution, l'électronique rapide qui était considérée comme la « servante de la science » dans les années 1960 est devenue, en tant qu'ingrédient essentiel de l'IA et de ses applications, une « reine » dominatrice, pièce centrale de l'asservissement du monde humain à l'utilisation du « tout électronique » dans tous les domaines de la vie pour s'informer, commander et acheter tous les biens matériels imaginables et unimaginable – et souvent pour éviter de devoir réfléchir.

La malveillance et les questions éthiques : quelques exemples³⁶

L'informatique « maillée » est vulnérable, elle mène à des risques majeurs si elle correspond à un développement non contrôlé et (ou) non contrôlable : la criminalité digitale est apparue il y a peu de temps et s'est particulièrement manifestée lors d'élections récentes dans les pays démocratiques.

Pour le moment le fait que des machines puissent générer des actions criminelles de leur propre volonté ou initiative n'est heureusement pas envisageable. La programmation humaine peut par contre faire de grands dégâts sur l'ensemble de la planète. On peut citer la probable intervention extérieure de la Russie dans les élections américaines de 2016, une situation peut-être semblable lors des élections françaises de 2017. Ces faits montrent les faiblesses des systèmes de protection devant les attaques criminelles qui mènent à la désinformation et aux informations mensongères (*fake news*). Plus récemment (juillet 2021) il a fallu se rendre compte des dégâts causés par l'affaire « Pegasus », la mise à disposition à tout bon payeur d'un programme permettant d'entrer dans les réseaux les plus « sensibles » pouvant mettre en péril la sécurité civile et militaire des États à travers le monde. Les mauvaises intentions d'individus et de véritables groupements de *hackers* sans scrupules seront un danger perpétuel pour le monde actif des entreprises modernes et des États tant que les réseaux existeront. Ce sont là aussi des dangers mortels pour les démocraties, pour le bon fonctionnement des sociétés, tant sur le plan des individus que celui des collectivités, pour leur survie sociale et économique.

36 GUERRAOUI, NGUYÊN HOANG, 2020.

SI UN FUTUR « TRANSHUMANISTE » POUVAIT S'INSTALLER : QUELS REMÈDES POUR Y ÉCHAPPER ? ³⁷

La convergence technologique pour créer du « radicalement nouveau » peut transgresser les limites de l'éthique, c'est-à-dire qu'elle peut violer l'intégrité physique et morale de l'homme.

Elle se décline essentiellement dans quatre domaines : les nanotechnologies, les biotechnologies, les techniques de l'information et les sciences cognitives. Si demain le danger post-humaniste ne peut pas venir de la création de machines faites de métal et de silicium la redoutable possibilité d'envisager le cerveau humain adjoint ou même implanté d'un potentiel artificiel pourrait rester envisageable. Comme évoqué plus haut, la médecine traumatique est capable de compenser un handicap physique tel que l'absence d'un membre en implantant une prothèse que le cerveau peut « piloter ». Son extrapolation, grâce aux développements des biotechnologies à venir, pourra se faire dans le futur pour le bienfait de l'humanité, au fur et à mesure de l'accroissement de la connaissance du cerveau. Mais on imagine aussi les problèmes que de telles opérations pourraient soulever quand les zones concernées commandent des fonctions comme la volonté, la prise de décision : quid du concept d'éthique ? L'entrée possible dans un monde métissé ou hybridé et hostile ? En créant un climat d'influence sur le mental et en essayant d'affaiblir ou même de faire disparaître la capacité de distinguer le vrai du faux – ce qu'elles font d'ailleurs déjà aujourd'hui – elles pourraient mener à la déstabilisation des sociétés humaines avec des conséquences imprévisibles, imaginables et inimaginables. Il paraît difficile de voir de quelle manière une telle situation pourrait être évitée.

37 GIORGINI, 2017 ; FROMENT, MAIGNANT, 2021.

EN CONCLUSION

L'Intelligence Artificielle est une réalisation remarquable créée par le génie de quelques-uns qui peut trouver des applications pour tout ce qui concerne l'homme et ses besoins. Comme tous les instruments créés au cours de millénaires depuis son apparition sur la terre, cet outil peut contribuer à son confort et son bonheur. L'extension de ses réalisations est assez vaste pour couvrir tous les aspects de la vie, autant matériels que spirituels, sa conception constitue un outil dont la mise en œuvre apporte chaque jour de nouvelles applications.

Mais les performances de l'IA sont loin d'atteindre celles de l'humain, il est difficile de croire qu'un jour quelque « robot » puisse être capable de faire la loi sur terre. Il a fallu des millions d'années pour que se développe cette formidable unité de commande, le cerveau, avec ses multiples propriétés, structurelles et fonctionnelles plus ou moins bien cernées, capables d'atteindre des objectifs qui semblent *a priori* hors de la portée de l'homme.

Nous savons cependant que tout nouveau développement et toute découverte peuvent avoir deux faces. Il ne faudrait pas que certains progrès, l'usage d'informations non contrôlées, utilisées sans esprit critique ou même de façon perverse, mènent la société humaine vers un état de conscience morale individuelle dégradée et au rejet d'une éthique menant à l'apaisement de tensions sociales exacerbées. L'homme se doit de respecter l'homme et la nature dans laquelle et de laquelle il vit. Les connaissances et leurs applications acquises grâce aux progrès techniques devraient y contribuer.

Je voudrais remercier ma consœur Christiane Roederer et mon confrère René Voltz pour leurs conseils éclairés ainsi que Madame Violaine Acker pour son aide précieuse sur le plan linguistique et technique.

RÉFÉRENCES

- AMUNTS Katrin, LIPPERT Thomas, « Brain research challenges supercomputing », *Science*, n° 6571, vol.374, 2021, p. 1054-1055
- ANDRÉ Michel, *La cathédrale de Turing, de George Dyson*, Pantheon Books, 2012.
- ATLAN Henri, *Entre le cristal et la fumée*, Paris, Seuil, 1979.
- BAECK Minkyung et coll., « An accurate prediction of protein structures and interactions using a three-track neural network », *Science*, n° 373, 2021, p. 871.
- BALIBAR Étienne, « L'Invention européenne de la conscience », *La conscience et ses troubles*, Paris – Bruxelles, De Bœck, 1998.
- BESNIER Jean-Michel, *Demain les Posthumains, Le futur a-t-il encore besoin de nous ?*, Fayard/Pluriel, 2020.
- BODEN Margaret A., *L'Intelligence Artificielle*, EDP Sciences, 2021.
- BOUCHET Alain, MASSON Jean-Louis, « Les localisations anatomiques de l'âme au cours des siècles », *Histoire des Sciences médicales*, n° 14 (1), 1980, p. 95-106.
- BOUILLET Marie-Nicolas, CHASSANG Alexis (dir.), « Pierre Coste », *Dictionnaire universel d'Histoire et de Géographie*, Paris, Hachette, 1878.
- BRETTO Alain, *Intelligence artificielle : la réalité & le mythe Ébauche d'une théorie générale de l'intelligence artificielle*, Gap, éditions DésIris, 2021.
- CARITAT Antoine, Marquis de Condorcet, *De l'influence de l'Amérique sur l'Europe*, Paris, Firmin Didot frères, impr. de l'Institut, 1847.
- DAVIES Alex et coll., « Advancing mathematics by guiding human intuition with AI », *Nature*, Vol. 600 p. 70.
- DEGRAVE J., FELICI F., BUCHLI J. et al., « Magnetic control of tokamak plasmas through deep reinforcement learning. », *Nature*, vol. 602, p. 414–419, 2022.
- DESCARTES René, *Discours de la Méthode*, 1637.

- DISCEPOLO Thierry, IZOARD Celia, préface et postface de *1984* George Orwell, Agone, 2021.
- DUPUY Jean-Pierre, *Aux origines des sciences cognitives*, Paris, La Découverte, 2005.
- EDELMAN Gerald M., *Biologie de la conscience*, Paris, Editions Odile Jacob, 2008.
- EDELMAN Gerald M., TONINI G., *Comment la matière devient conscience*, Paris, Editions Odile Jacob, 2000.
- FROMENT A, MAIGNANT G. (dir.), *Philosophie et Éthique de la Smart city : Homo futurus, La Société d'Ecologie humaine, Quelles évolutions biologiques pour l'espèce humaine ?*, Nice, Les Editions Ovadia, 2021.
- GILLOT Pascale, *La question du sujet – Descartes et Wittgenstein*, Paris, CNRS Éditions, 2020.
- GIORGINI Pierre, *La tentation d'Eugénie : l'humanité face à son destin*, Montrouge, Bayard, 2017.
- GUERRAOUI Rachid, NGUYÊN HOANG LÊ, *Turing à la plage*, Paris, Dunod, 2020.
- HARARI Yuval Noah, *Sapiens – Une Brève Histoire de l'Humanité*, Paris, Albin Michel, 2012.
- HARARI Yuval Noah, *Homo deus – Une Brève Histoire du futur*, Paris, Albin Michel, 2017.
- HIRLIMANN Charles, *La révolution CO₂*, The BookEdition.com, 2018.
- LECHEVALIER Bernard, LECHEVALIER Bianca, « Aborder la question de la conscience », *Revue française de psychanalyse*, 2007/2 (Vol. 71), p. 437-454.
- MESULAM Marsel, « Chapitre 11. Les centres nerveux : lieux topographiques et lieux fonctionnels » dans *Traité de neuropsychologie clinique*. Francis Eustache éd., Paris, De Boeck Supérieur, 2008, pp. 125-140.
- NAGEL Ernest, NEWMAN James R., GÖDEL Kurt, GIRARD Jean-Yves, *Le théorème de Gödel*, Paris, Sciences, Seuil, 1989.
- NEUMANN, John von, *Théorie Générale et Logique des automates*, Ceyzérieu ChampVallon, 1998.

NEUMANN, John von, *Le Cerveau et l'Ordinateur*, Paris, Flammarion, collection « Champs », 1996.

PESSOA Fernando, *Le livre de l'intranquillité*, Paris, Christian Bourgois Editeur, 2004

SANDBERG Andus, BOSTROM Nick, *Whole Brain Emulation : A Roadmap*, Technical Report #2008-3, Future of Humanity Institute, Oxford University.

VOLTZ René, « L'Innovation clé de l'Évolution Économique » dans *Akademos*, Institut de France, 2019.

Pour les personnes intéressées par la composante informatique :

NIELSEN Michael : *Neural Networks and Deep Learning – Cours d'initiation* <http://neuralnetworksanddeeplearning.com>
+ références dans les textes

L'INTELLIGENCE ARTIFICIELLE VA-T-ELLE CONTRIBUER À TRANSFORMER *HOMO SAPIENS* EN *HOMO NUMERICUS* ?

par M. Michel CHEIN

de l'Académie des Sciences et Lettres de Montpellier

Le titre de cette conférence : *Intelligence artificielle, neurosciences, biotechnologies : vers quel Humanisme ?* pose lui-même de nombreuses questions. Des technologies – Ô combien importantes puisqu'elles concernent la vie, mais quid des sciences de la vie ? –, des sciences – Ô combien importantes puisqu'elles concernent, entre autres, le cerveau, mais quid des technologies neuronales ? –, un domaine – l'Intelligence artificielle (IA dans le texte) objet de nombreuses craintes et de tout autant d'espairs, dont on parle tous les jours avec des acceptions différentes : difficile d'identifier science, technique, sorcellerie ? – et l'informatique et le numérique ? Et pourquoi ces domaines conduiraient-ils à un seul Humanisme, ne faudrait aussi s'interroger sur les humanismes possibles et quels sont les choix, en particulier politiques, permettant, si ce n'est d'aller vers un « avenir radieux » du moins d'éviter le pire... Dans l'introduction j'explique les limitations drastiques – ne considérer que le domaine de l'aide à la décision qui fait l'objet des deux parties centrales de ce texte – qui m'ont semblé pertinentes pour aborder ce sujet immense (toute la société est concernée, de l'individu à l'avenir de la planète), passionnant (quelles limites, jusqu'où pourrait-on aller ?), inquiétant (du contrôle des applications à celui de la population), complexe (qui a une vision du Web 3 ?) et urgent (la technologie court plus vite que les débats) de cette conférence. En conclusion, je propose quelques recommandations pour éviter le pire.

INTRODUCTION

Inutile d'insister sur les aspects positifs de nombreuses applications de l'IA, en santé – par exemple dans l'analyse d'images, l'analyse des EEG, le monitoring, la robotique médicale –, en recherche d'informations et recommandation, pour traquer le harcèlement en ligne, calculer des itinéraires en utilisant l'état du trafic, dans la conduite d'avion, les anti-spams, ces applications sont très présentes dans les médias.

Inutile également de dénoncer certaines pratiques bien connues comme l'hameçonnage personnalisé (c'est-à-dire la détermination de cibles intéressantes et la génération de messages crédibles en fonction des profils de ces cibles) ; les atteintes aux droits de l'homme, en particulier à la vie privée ; tout ce qui conduit à une véritable perversion du débat démocratique, les manipulations politiques par des campagnes de propagande automatisées et personnalisées, en utilisant des données personnelles vendues ou volées (il est bien connu maintenant que dans la campagne de Donald Trump, la société Cambridge Analytica détermina des États pouvant basculer, puis dans ceux-ci des électeurs incertains, et enfin envoya des messages ciblés ; le *Brexit* aurait été entaché lui aussi par ce type de manipulation et les élections politiques doivent maintenant être protégées contre ces attaques) ; la fabrication de *fake news* en utilisant de fausses photos ou la manipulation de vidéos avec des systèmes permettant la synthèse de la parole. Tous ces dangers sont renforcés parce que les techniques de l'IA sont un monopole des géants de l'informatique qu'ils soient américains (GAFAM) ou chinois (Alibaba, Baidu, Tencent, Xiaomi, etc.) et lorsqu'elles sont aux mains d'un régime autoritaire ou totalitaire comme en Chine, elles pourraient conduire rapidement à un contrôle total

de la population et à la construction d'un « homme nouveau » (ce que décrit avec angoisse Kai Strittmatter ¹).

Cependant, on peut se demander si, dans un pays démocratique, il est nécessaire de s'inquiéter, c'est-à-dire se demander si les techniques d'IA sont porteuses, en elles-mêmes, d'un tel danger, si l'utilisation massive de systèmes intégrant de l'IA pourrait contribuer à transformer l'Homme en un *Homo Numericus* premier pas vers un *brave sujet décervelé* ? Les techniques d'IA sont-elles si puissantes que cela ?

Schémas de Winograd

Auguste, le clown, essaie depuis plusieurs minutes de mettre une immense statue dans une minuscule valise. Monsieur Loyal, qui observe ses nombreuses tentatives suscitant les rires du public, lui dit au bout d'un moment : « Mais arrêtez donc, vous voyez bien que la statue n'entre pas dans la valise, elle est trop grande ! » Auguste s'arrête, regarde attentivement, l'air étonné, la statue et la valise, sort un long mètre jaune pliant en bois de sa chaussure, mesure la statue et la valise, recommence, lève les yeux au ciel et finalement, sûr de lui, lui répond « mais non, elle est trop petite ! ».

Depuis Turing et son jeu de l'imitation², l'écriture d'un programme de *questions/réponses* en langue naturelle est un problème scientifique qui a suscité de nombreux travaux mais qui n'a toujours pas de solution générale satisfaisante. On s'est aperçu que *le test de Turing* (le jeu de l'imitation dans sa version simple) conduisait à écrire des programmes consistant à tromper une personne plus qu'à apporter des réponses directes et claires aux questions posées, ceci en utilisant jeux de mots, plaisanteries, citations, débordements émotifs,

1 Dictature 2.0 *Quand la Chine surveille son peuple (et demain le monde)*. Tallandier, 2020.

2 Par exemple voir Test de Turing dans Wikipedia.

apartés, etc. Pour expliquer des réponses étranges faites par un programme celui-ci se faisait passer pour un jeune Ukrainien de 13 ans maîtrisant mal l'anglais ! Très critiquée la *Loebner Competition* entre programmes prétendant passer le test de Turing s'arrêta en 2016. Winograd suggéra un cadre limité et précis pour ce problème, que Levesque proposa d'appeler schéma de Winograd.

Par exemple, le schéma associé au dialogue des clowns est le suivant :

Énoncé : *La statue n'entre pas dans la valise car elle est trop [grande/petite].*

Question : *qu'est-ce qui est trop [grande/petite] ?*

Réponse : *[la statue/la valise].*

Plus généralement, un tel schéma consiste en un énoncé avec une alternative, une expression spéciale entre crochets (ici grande ou petite), et une question concernant cette expression spéciale avec les contraintes suivantes :

la question consiste à demander quel est le référent du pronom de l'énoncé (ici elle),

deux réponses sont possibles suivant l'expression spéciale tirée au hasard (grande ou petite),

répondre doit être très facile, quasi instantané, pour des humains,

la réponse doit faire appel à des connaissances sur le monde,

on doit utiliser des raisonnements de bon sens,

on ne doit pas pouvoir répondre en utilisant des méthodes statistiques ou des moteurs de recherche (on dit qu'elle doit être *google-proof*),

l'évaluation est faite automatiquement en ne comptant pas de la même façon une bonne et une mauvaise réponse, les mauvaises réponses sont pénalisées³.

3 Pour N questions sur un domaine la note est : $\max(0, N - k)$. Nombre Réponses Fausses/ N), où $k \geq 2$ est un coefficient pénalisant les mauvaises

Voici un autre exemple :

Énoncé : *Paul a essayé de joindre Georges sur son téléphone, mais il [n'a pas réussi / n'a pas répondu].*

Question : *Qui [n'a pas réussi / n'a pas répondu] ?*

Réponses : [*Paul / Georges*]

Une compétition existe sur ce sujet depuis 2011 (à laquelle ni IBM ni Google, entre autres, ne participent) dont les résultats montrent bien les difficultés posées par un raisonnement de sens commun simple pour un humain (cf.⁴ pour un ensemble de schémas de Winograd).

Les schémas de Winograd ne sont qu'un exemple de tâches intellectuelles simples pour un humain que la machine ne sait pas bien résoudre aujourd'hui – il en existe beaucoup d'autres, « un crocodile peut-il courir un 110 mètres haies ? » est un exemple célèbre de question à laquelle un programme ne sait pas répondre... tant qu'on ne lui a pas donné la réponse.

Le désastre des deux tentatives du robot de conversation (*chatbot*) Tay de Microsoft est un exemple bien connu dans le domaine du traitement automatique des langues de limite de systèmes utilisant des techniques d'apprentissage. Tay devait passer pour une adolescente, initialement décrite, en particulier, à partir de données publiques, puis évoluer en discutant sur Twitter. En quelques heures, les tweets de Tay sont devenus racistes, misogynes et antisémites ! Si le système de Microsoft avait sans doute bien choisi les données initiales il n'avait pas prévu les tweets des utilisateurs.

Même si des systèmes de traduction sont spectaculaires pour certains types de textes, l'IA a des progrès à faire dans le

réponses (par exemple avec $k=2$ si une réponse sur 2 est fausse la note est 0).

⁴ <https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WSCollection.html>

domaine du traitement automatique des langues ! Mais au fait, c'est quoi l'IA ?

Polysémie galopante

Le terme IA est surchargé de significations, ce qui fait que l'IA peut parfois apparaître comme une espèce d'auberge espagnole où chacun apporte ce qu'il souhaite. En se limitant à ses usages scientifiques, alors qu'il y a des usages commerciaux, politiques, voire idéologiques, on peut constater quatre phénomènes qui expliquent cette polysémie galopante.

Un phénomène de restriction : l'IA c'est l'apprentissage numérique (notons que c'est l'usage dominant aujourd'hui dans la presse, l'économie et la politique).

À ce phénomène de réduction est associé un phénomène concomitant d'expansion, l'IA c'est toute l'informatique plus la robotique, plus l'aide à la décision, plus, etc.

Il existe également un processus d'évaporation et ce depuis le début, des domaines s'autonomisent ou s'immergent dans l'informatique. Un exemple ancien est celui de la linguistique computationnelle, un exemple actuel est celui des sciences des données. Ce qui était jusqu'à ces dernières années une partie de l'IA devient une science autonome.

On assiste aussi à un phénomène de neutralisation de certains domaines de l'IA par l'informatique, qui en les absorbant, fait perdre à ces domaines leur étiquette IA. En effet, les chercheurs en IA sont des explorateurs de l'informatique. Ce sont des chercheurs en IA qui sont à l'origine de nombreuses notions devenues banales en informatique : le temps partagé, les interfaces homme-machine, les langages fonctionnels, les langages objets, le *backtracking*, l'*alpha-beta pruning*, le *garbage collector*, les systèmes multi-agents, la programmation logique, la programmation par contraintes, etc. Ce que Nick Bostrom (directeur du *Future Humanity Institute*) décrivait en 2006 de la manière suivante : « Beaucoup d'IA de pointe a filtré dans

des applications générales, sans y être officiellement rattachée car dès que quelque chose devient suffisamment utile et commun, on lui retire l'étiquette d'IA. Aujourd'hui ce serait plutôt l'inverse ! Quand l'IA a le vent en poupe, comme en ce moment, tout le monde fait de l'IA parce que c'est excitant, à la mode, et c'est aussi pour bénéficier des financements des plans nationaux ou pour conquérir des marchés. »

Depuis environ 70 ans l'IA, en tant que discipline scientifique avec ses nombreux journaux scientifiques et congrès internationaux, avec ses sociétés savantes (tous les pays économiquement développés ont une telle société savante), s'est développée suivant les grands axes : Apprendre, Acquérir et Représenter des connaissances, Reasonner, avec des ruptures, des échecs et des succès.

L'IA faible, qui consiste à construire des programmes simulant des tâches intellectuelles spécifiques – un tel programme est souvent appelée *une IA* – a obtenu de nombreux succès.

L'IA forte, qui a pour objectif de construire un programme simulant toutes les capacités cognitives humaines, voire plus, n'est toujours qu'un projet, ce n'est qu'une idée, une conjecture qui fait travailler certains chercheurs et suscite de nombreux débats et controverses, ce qui est sûr c'est que ce n'est pas pour demain. Il n'y a pas que la beauté et les arts, les sentiments et la conscience qui soient hors du champ actuel de l'IA, personne ne compte sur une machine pour inventer une nouvelle théorie dans n'importe quel domaine scientifique. Et rappelons que la complexité des machines est limitée et que l'être humain est d'une complexité biologique (matérielle) beaucoup plus grande que celle des ordinateurs physiques (même quantiques... lorsqu'on les maîtrisera).

Existe-t-il des intelligences artificielles intelligentes ?

« *Il* ne comprend même pas ce qu'il fait », « *il* ne comprend même pas pourquoi il le fait », « *il* ne sait même pas que ça ne

s'applique pas dans ce cas-là », autant d'expressions qui conduiront à penser qu'*il* n'est pas très intelligent. Difficile de dire qu'on est intelligent si on ne comprend pas ce qu'on fait, si on ne peut pas dire pourquoi on fait telle chose, pourquoi on utilise tel outil, quelles sont ses conditions d'usage. Fabien Gandon dans *Les IA comprennent-elles ce qu'elles font ?* explique que comprendre « c'est saisir le sens, les finalités, les causes et conséquences, les principes. Comprendre quelque chose, c'est recevoir ou élaborer une représentation de cette chose, c'est s'approprier une conceptualisation reçue ou construite, qui permettra notamment de produire un comportement intelligent. »

Dans le cas du *chatbot* Tay, ce que Microsoft n'avait pas envisagé, alors que des spécialistes des réseaux sociaux auraient pu le prédire, était qu'un groupe d'utilisateurs de Twitter commencerait immédiatement à tweeter des commentaires racistes et misogynes. Tay a rapidement appris et a incorporé ces commentaires racistes dans ses propres tweets parce qu'il ne comprenait pas le sens des mots.

Comprendre c'est être capable de répondre à une succession de *pourquoi ?* par des réponses de plus en plus précises correspondant à des niveaux de compréhension de plus en plus détaillés au fur et à mesure que des *pourquoi ?* s'enchaînent. Par exemple, différents niveaux de compréhension, d'explication, de la succession du jour et de la nuit, pourraient être, dans un premier temps : le soleil est une source fixe de lumière et la terre tourne sur elle-même, puis ajouter que la terre tourne autour du soleil, puis que l'axe de rotation de la Terre fait avec la perpendiculaire au plan de l'écliptique un angle constant de $23^{\circ}26'$, puis introduire les variations de luminosité, de température et de composition atmosphérique de la Terre, etc.

Dans le titre de ce paragraphe « intelligence artificielle » est une locution désignant un programme, ou un ensemble de programmes utilisant certaines techniques et certaines

données, alors que l'attribut « intelligent » est pris dans son sens humain habituel, si bien qu'on peut dire, aujourd'hui, qu'une intelligence artificielle n'est pas très intelligente voire que toutes les intelligences artificielles sont stupides puisqu'elles ne comprennent pas ce qu'elles font ni pourquoi elles le font !

C'est une personne qui fixe les objectifs du programme (jouer à un seul jeu ou à plusieurs, jouer le mieux possible ou avoir plusieurs niveaux de jeu, battre un champion du monde, avoir une interface graphique en 2 ou 3 dimensions, etc.) qui décide de sa conception et de sa construction (quelle combinaison de différentes techniques : apprentissage automatique, recherche arborescente, fonction d'évaluation, Monte-Carlo, règles logiques, accès à des bases de parties, etc.), de ses conditions d'usage (essayer d'utiliser Deep Blue pour jouer au Go !), de son évaluation (peut-on imaginer Deep Blue sans Kasparov ou Alpha-Go sans Lee Sedol ?).

Et lorsque l'on parle d'un artefact « intelligent » c'est dans le sens où sa conception et sa réalisation ont nécessité beaucoup d'intelligence humaine ou qu'il a un comportement efficace pour une tâche limitée ; à ce dernier sens de nombreuses IA sont souvent beaucoup moins « intelligentes » qu'un mécanisme d'horlogerie, et s'il existe, certes, des IA qui sont des programmes très complexes, c'est l'intelligence des hommes qui les ont réalisées.

Comprendre le sens d'un symbole, d'un mot en particulier, nécessite que celui-ci soit relié dans le cerveau à des zones cérébrales le concernant (cf. pour une présentation amusante le blog binaire⁵). Nils Nillson, chercheur éminent en Intelligence Artificielle récemment décédé, a écrit un article intitulé *Human-Level Artificial Intelligence ? Be Serious !*⁶

5 <https://www.lemonde.fr/blog/binaire/2021/12/14/que-se-passe-t-il-dans-les-cerveaux-des-cons/>

6 *AI Magazine*, 25th Anniversary Issue.

Gérard Berry, lui aussi chercheur éminent en informatique, disait en 2016 : « Je n'ai jamais été déçu par l'intelligence artificielle parce que je n'ai pas cru une seule seconde en l'intelligence artificielle. Jamais. ». Et, un peu plus tard, il précisait son point de vue en écrivant que l'IA est la « quête d'une intelligence non naturelle, implémentée sur un ordinateur par exemple. Certains disent que le sujet n'existera que quand l'ordinateur sera capable de pouffer de rire alors qu'on lui raconte une histoire drôle qu'il ne connaît pas. Mais ce sont des mauvaises langues. J'ai appris que Google et Facebook s'étaient cotisés pour construire le premier ordinateur réellement intelligent. Je connais la première question qu'ils vont lui poser : « Dieu existe-t-il ? » Et je devine la réponse : « Maintenant, oui. » Mais trêve de balivernes : pour les gens sérieux du domaine, c'est aussi une grande quête sur les possibilités réelles des machines comparées à nos propres limites, qui peut aussi nous éclairer sur ce que nous appelons sans trop de modestie l'« intelligence humaine »⁷.

Mais alors... pourquoi s'inquiéter ?

Reconnaissant que les IAs sont plutôt stupides que géniales, même si elles peuvent être des merveilles techniques. Alors, pourquoi s'inquiéter ?

Il suffit de quelques titres de livres récents écrits par des Français (la littérature anglaise est beaucoup plus importante) – *L'Intelligence artificielle n'existe pas*, Luc Julia (informaticien, co-inventeur de Siri), First éd. 2019 ; *L'Intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical*. Éric Sadin (philosophe), L'Échappée, 2018 ; *La fabrique du crétin digital*. Michel Desmurget (neuroscientifique), Le seuil, 2019 ; *L'homme nu. La dictature*

⁷ Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris, Odile Jacob, 2017.

invisible du numérique. Marc Dugain (romancier, journaliste), Christophe Labbé (journaliste), Plon, 2016 – pour se dire que malgré les nombreux succès de l’IA, dont je ne parlerai pas car ils sont quotidiennement exposés dans les médias, les choses ne sont pas si simples.

Étant informaticien, spécialiste d’IA mais absolument pas philosophe, je limiterai drastiquement mon propos pour essayer de ne pas dire trop de bêtises. J’aborderai les liens entre IA et Humanisme en m’intéressant principalement à la question suivante : à quels effets anthropologiques pourrait mener, ou mène déjà parfois, l’utilisation de certains logiciels d’aide à la décision utilisant des techniques d’IA et dont les objectifs concernent des humains – ou plus généralement le vivant – si leurs usages n’étaient pas rigoureusement contrôlés ?

Donc il ne sera pas tellement question de science, un peu de technologie mais surtout d’applications dans un domaine particulier. Cette sélection, en partie arbitraire comme toute classification, ferait sans doute réagir Philippe Descola, puisque d’une certaine manière elle calque la séparation entre les choses et l’humain, comme si les conditions de création, de distinction, d’usage des choses n’avaient rien à voir avec l’humain, d’autant plus que parmi ces choses j’inclus des artefacts ! Cependant, je pense que cette séparation permet de mettre en évidence, dans les applications de l’IA, ce qui tend à déshumaniser l’humain, à le transformer en un homme informatique ou numérique, oxymore s’il en est, du moins de mon point de vue.

AIDE À LA DÉCISION

La définition classique de l’aide à la décision, telle que celle proposée par Bernard Roy : « L’aide à la décision est l’activité de celui qui, en prenant appui sur des modèles, aide à obtenir des éléments de réponse aux questions que se pose un

intervenant dans un processus de décision, éléments concourant à éclairer la décision et à recommander un comportement de nature à accroître la cohérence entre l'évolution du processus et les objectifs de cet intervenant », peut être généralisée à l'acception actuelle en considérant non plus seulement des humains mais aussi en utilisant des agents artificiels, des IAs (si on généralise aussi l'intervenant à des agents artificiels on obtient des systèmes automatiques ou robotiques dont je ne parlerai pas ici).

Par exemple, IBM décrit ainsi son système *Watson for Cybersecurity* utilisé pour la détection et l'investigation d'attaques : « la détection d'attaques se fait en corrélant les incidents, et l'analyse de sécurité peut solliciter l'aide de Watson pour comprendre l'origine du problème et les signaux remontés » et, complète Hugo Madeux, « la logique s'inscrit toujours dans le cheminement, l'enrichissement, l'analyse, le raisonnement ». Ainsi, la décision finale est prise par un humain et l'IA ne fournit que des outils d'aide permettant d'utiliser au mieux l'intelligence humaine. Mais les outils informatiques transforment, comme n'importe quel outil, ceux qui les utilisent. Les outils de l'IA étant des outils « intellectuels » ils transforment la manière de penser de ceux qui les utilisent et peuvent même parfois prendre leur place...

Donnons quelques exemples pour préciser les systèmes qui nous intéressent plus particulièrement ici.

Themis est un système comprenant un micro et un haut-parleur commandés par un logiciel de reconnaissance du langage qui, expliquerait un enfant, « crie » lorsqu'on prononce un « gros mot ». Naturellement, même si un tel système risque d'avoir des difficultés à reconnaître certaines figures de style il pourrait être utilisé pour apprendre à parler « correctement » une langue, correctement au sens du concepteur du logiciel !

À l’opposé, il existe des systèmes de sécurité qui « crient » lorsqu’ils anticipent une menace, par exemple lorsque la cuve d’un réacteur nucléaire risque de se fendiller.

Nous ne nous intéressons qu’à des systèmes dont les décisions concernent des humains et pas des objets. Ces systèmes peuvent être plus ou moins incitatifs (e.g. systèmes de recommandation dans le marketing), prescriptifs (e.g. utilisés par exemple dans les banques ou les assurances) ou coercitifs (e.g. refus d’embauche, surveillance par la police ou condamnation par la justice)⁸.

La plupart des différents systèmes prédictifs dont on parle aujourd’hui utilisent les mêmes techniques d’apprentissage, particulièrement des réseaux de neurones profonds (*deep learning*), sur des données massives (*big data*) annotées. Avant d’aller plus avant il est nécessaire d’expliquer très rapidement ce qu’est le *deep learning*.

Apprentissage automatique par réseaux de neurones

Nous présentons le principe de ces méthodes (supervisées) à partir du problème de la reconnaissance de chiffres manuscrits. Ce problème, facile pour un humain est resté longtemps difficile pour un programme (les premiers *captcha* étaient composés de lettres ou des chiffres). Il suffit d’essayer d’écrire un programme reconnaissant des chiffres manuscrits pour s’en convaincre. Par exemple, pour décrire un « un » manuscrit on pourrait dire qu’il est *souvent* composé de deux segments, le plus long, *proche* de la verticale, a son extrémité supérieure *voisine* de celle du segment le plus court. Ces deux segments, dans un rapport *approximatif* de 1 à 3, font un angle *d’environ* 45 degrés ; ils peuvent être complétés par une base horizontale, de longueur *proche* de celle du segment le plus court, accueillant en un point *proche* de son milieu l’extrémité inférieure du segment le plus long. Certains « un » manuscrits

⁸ Distinction proposée par Éric Sadin dans l’opus cité.

sont aussi composés d'un unique segment, etc. Bref, les nombreux cas, les exceptions, les approximations font que ce qui est simple pour un humain est difficile à expliciter et à formaliser donc difficile pour un programme, sauf s'il est capable d'apprendre.

L'apprentissage supervisé consiste à considérer de nombreux exemples de chiffres manuscrits, et à améliorer les paramètres du programme jusqu'à ce qu'il les reconnaisse suffisamment souvent. Les récents succès pour résoudre ce type de problèmes ont été obtenus avec des réseaux de neurones artificiels. Un neurone artificiel simple fonctionne comme un mécanisme d'admission à un examen. On a une note par matière, chaque matière est affectée d'un coefficient, et un candidat est reçu à l'examen si la somme de ses notes dépasse un seuil sinon il est collé.

Un réseau de neurones comme celui de la figure ci-dessous est composé de tels neurones, les sorties d'un neurone pouvant être reliées aux entrées d'autres neurones. Dans l'exemple de la reconnaissance des chiffres, on donne en entrées les valeurs des pixels d'une image d'un chiffre et on modifie les poids et les biais jusqu'à ce que pour un chiffre, par exemple « un », à force de fournir de nouveaux exemples de « un », le réseau ait une sortie toujours supérieure à 0.5 sur le neurone de sortie correspondant au chiffre « un », et pour les neuf autres neurones de sortie toujours une sortie inférieure à 0.5.

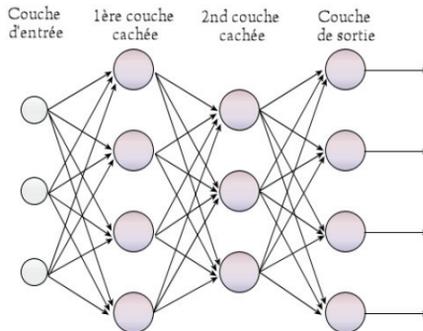


Figure : un réseau de neurones

La phase d'apprentissage est terminée lorsque le réseau donne les bons résultats sur les exemples. Le réseau est alors évalué sur un autre échantillon de données et si les résultats sont satisfaisants (s'il donne suffisamment souvent la bonne sortie) on considère que le réseau a appris, on peut alors l'utiliser pour reconnaître des données quelconques de chiffres.

Les banques, pour reconnaître les chèques, et les postes, pour reconnaître les adresses, ont été les premiers grands utilisateurs commerciaux de ces techniques qui ont depuis été utilisées avec succès dans de nombreux domaines (reconnaissance d'images, de sons, de mots, dans des jeux, traduction...). Trois raisons expliquent ces nouveaux succès : premièrement, l'informatisation de la société (internet, les réseaux sociaux, les *data centers*) permet d'avoir de très nombreux exemples ; deuxièmement, les machines de plus en plus rapides permettent de traiter ces grandes masses de données ; troisièmement, des algorithmes efficaces ont été inventés.

On vient d'expliquer très succinctement *comment ça marche* dans le cas d'un système apprenant à reconnaître un chiffre manuscrit, ou un animal, ou un bateau de pêche, ou une tumeur cancéreuse ou un blindé, etc. On peut remarquer que les humains n'apprennent pas comme ça, un enfant n'a pas besoin de millions d'exemples d'images de chats pour savoir reconnaître un chat !

Reprenons les étapes d'une application basée sur une méthode d'apprentissage :

1. définir l'objectif de l'application (reconnaître des nombres ou des champignons, faire acheter des livres, jouer – univers fermé – ou conduire une voiture – univers ouvert, non seulement l'environnement mais aussi les conducteurs, déterminer un risque d'attentat, etc.) ;

2. choisir une méthode (e.g. apprentissage supervisé par un réseau de neurones) ;
3. acquérir des données d'apprentissage et les analyser : les données sont-elles pertinentes, c'est-à-dire représentatives du but (reconnaissance de chèques ou risque d'attentat) ? Si le modèle d'apprentissage choisi a beaucoup de paramètres il faut beaucoup de données.
4. évaluer sur d'autres jeux de données que ceux utilisés dans la phase d'apprentissage et qui doivent être représentatifs des cas réels qu'il faudra traiter.

Qui décide ?

Nous évoquons ici un problème fondamental : le glissement d'un système d'aide à un décideur humain vers un système automatique. Un système d'*aide* à la décision sous-entend que c'est un intervenant humain qui prend la décision. Le décideur devrait donc être capable d'expliquer comment il a pris sa décision à partir des propositions faites par le système donc comprendre, et expliquer, comment le système fonctionne, pourquoi il a fait de telles propositions et pourquoi lui, le décideur, a fait son choix dans cet ensemble. Tout le monde a eu l'occasion d'obtenir comme réponse à une question concernant une décision « je n'y peux rien c'est l'informatique ! » demain nous entendrons « je n'y peux rien c'est l'IA ! ». Tous les systèmes d'aide à la décision sont susceptibles de ce glissement. C'est un problème crucial : de tels systèmes d'aide à la décision doivent rester « d'aide » à des décideurs humains et ne doivent pas prendre le contrôle en particulier quand leurs décisions concernent des humains qui ne sont pas réductibles à un modèle informatique.

Un exemple qui illustre dramatiquement cela est celui du passage de drones télécommandés à des armes autonomes puisqu'il s'agit dans ce cas de décider de qui vit et qui meurt

(pour ce sujet largement commenté on peut regarder la vidéo fictive mais réaliste *Slaughterbots – if human : kill()*⁹ !

Certaines méthodes ne donnent comme explication que la méthode elle-même (dans l'exemple précédent de reconnaissance de chiffres, pourquoi le programme a-t-il reconnu un 7 et pas un 1 ? Parce que j'ai fait de nombreux calculs aboutissant à cette conclusion, pourrait répondre le programme. Un système à base de règles pourrait expliquer qu'il avait conclu par un 7 et pas par un 1 parce qu'il avait détecté une ligne brisée constituée de deux traits, l'un oblique 2 fois plus long que l'autre horizontal, la jonction se faisant entre l'extrémité nord du premier et est du second, etc. Un directeur des ressources humaines utilisant un système d'aide au recrutement annonce à un candidat « nous sommes désolés : vous ne remplissez pas les conditions pour occuper ce poste. » Pourquoi, demande le candidat ? Le DRH devrait répondre : « Parce que les personnes que nous avons refusées (en fait, celles refusées dans les bases données utilisées pour l'apprentissage) avaient des paramètres semblables aux vôtres » ou bien « parce que les personnes que nous avons recrutées (ou recrutées ailleurs pour un tel poste) avaient des paramètres incompatibles avec les vôtres. » Dans les programmes basés sur de l'apprentissage automatique statistique, tout n'est affaire que de corrélations.

Neutralité et biais

Les moyens pour développer des applications étant nécessairement limités, et ceci quel que soit le domaine concerné (médical, politique, militaire, etc.) le but choisi pour le développement d'une application n'est pas neutre (dans toutes ces alternatives fictives, le *et* au lieu du *ou* exclusif serait souhaitable – maladie d'Alzheimer ou autisme ? manipulation de vote ou critiques au gouvernement ? menaces d'une grande

9 <https://www.youtube.com/watch?v=9rDo1QxI260>

puissance totalitaire ou terrorisme ?). Il n'y a pas non plus de neutralité dans le choix de la méthode, des données et de l'évaluation des résultats.

Parmi ces problèmes nous considérerons celui concernant les données. Avoir des données pertinentes, c'est-à-dire représentatives du domaine concerné, est un problème fondamental de statistiques c'est donc un problème fondamental des méthodes du type *deep learning*, et si les données concernent des humains leur analyse et leur évaluation sont du ressort des sciences humaines et sociales devant prendre en compte, entre autres, la manière dont elles ont été acquises (par achats, par butinage sur les réseaux, par *crowdsourcing*, extraites de bases de données expertes, etc.). Nous rappelons ci-dessous quelques exemples de biais dans les données utilisées par certains systèmes.

Médecine

Il ne s'agit pas ici de nier l'intérêt des systèmes d'analyse d'images (reconnaissance de certaines tumeurs par exemple), l'analyse des EEG, le monitoring, les systèmes prédictifs d'aide au diagnostic ou comme récemment celui concernant l'insuffisance cardiaque¹⁰ (de nature similaire à un système d'alarme sur une centrale nucléaire), mais de décrire les biais d'un système utilisé aux États-Unis identifiant des patients à haut risque nécessitant des soins spécifiques (de nature similaire à un système d'aide au recrutement ou d'attribution d'un crédit). L'un des paramètres utilisés par ce système étant le montant des dépenses de santé, il recommandait ces soins pour des patients blancs beaucoup plus souvent que pour des patients noirs¹¹.

10 <https://www.loria.fr/fr/evm-un-algorithme-predictif-de-linsuffisance-cardiaque-crea-a-nancy-avec-lia/>

11 <http://www.slate.fr/story/183384/systeme-sante-americain-algorithme-raciste>

À côté de nombreux succès, il faudrait pour le moins supprimer *personnalisée*, pour des systèmes basés sur l'apprentissage automatique, dans les objectifs de la médecine 4P – prédictive, préventive, personnalisée et participative –, car si les statistiques sont vraies en général (en santé publique, les méthodes bayésiennes fournissent des preuves statistiques de l'influence de certains facteurs sur certaines maladies) elles sont fausses en particulier (parce que dans ces méthodes chaque individu est un point isolé dans un espace de grande dimension), médecine personnalisée et médecine prédictive sont (presque) antinomiques¹² car une vérité statistique est vraie en général (en épidémiologie) mais peut être fausse en particulier (« mon voisin qui fume comme un pompier a 95 ans et se porte très bien ! »).

Recrutement

Face aux énormes quantités de CV reçus, de nombreux services de ressources humaines utilisent des systèmes plus ou moins automatiques d'aide au recrutement. Les systèmes fonctionnent généralement de la manière suivante : ils analysent les CV et fournissent une note, seuls les CV ayant passé un certain seuil sont lus, dans le sens lus par des humains. Le système développé puis abandonné par Amazon a été souvent commenté. En effet, ce système de recrutement avait été entraîné sur des données concernant majoritairement des hommes, le résultat fut de sous-noter les femmes et de sur-noter les hommes « leur logiciel de recrutement n'aime pas les femmes » a pu titrer Reuters¹³. « L'entreprise a tenté de le modifier pour le rendre neutre, mais a finalement estimé qu'elle était dans l'impossibilité de garantir que le logiciel

12 <https://theconversation.com/medecine-police-justice-lintelligence-artificielle-a-de-reelles-limites-170754>

13 <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

n'apprendrait pas d'autres façons discriminatoires de trier les candidats et a mis fin au projet¹⁴. »

Reconnaissance de genre

Le même type de biais a été mis en évidence par Joy Buolamwini¹⁵, une chercheuse du MIT, qui a démontré que les systèmes de reconnaissance du genre d'une personne à partir d'une photo faisaient beaucoup plus d'erreurs pour le classement des femmes à la peau sombre que pour celui des hommes blancs. Par exemple, le système de Microsoft classifiait incorrectement 1 % d'hommes blancs et 35 % de femmes à la peau sombre. Pourquoi ? Parce que leurs données d'apprentissage comprenaient beaucoup plus d'hommes blancs que de femmes noires.

Police

PredPol est un système de police prédictive basé sur une méthode pour prévoir les répliques des tremblements de terre. Son objectif était de déterminer les secteurs où la police devait être prioritaire. Les données utilisées étaient celles issues des plaintes des victimes, sauf pour les homicides, et pas celles des arrestations (sans doute pour ne pas être politiquement incorrect). Or, les enquêtes concernant les personnes s'affirmant victimes « montrent que la distribution des plaintes n'est pas homogène dans la population car certaines victimes pensent que la police ne peut rien faire pour régler leurs problèmes ou qu'il ne vaut pas la peine de déposer plainte. Le fait que les victimes ne recourent pas à la police s'explique par leur position sociale, leurs expériences passées avec la police, leur lieu de résidence et leur propension à agir dans l'intérêt de la vie du quartier. Les non-signalements sont des phénomènes sociaux en tant que tels, qui échappent complètement à

14 <https://www.lebigdata.fr/amazon-abandonne-ia-misogyne>

15 https://fr.wikipedia.org/wiki/Joy_Buolamwini

l'apprentissage statistique par les données enregistrées par la police. » Ainsi PredPol, qui conduit à ce comme l'écrit Cathy O'Neil, « Les innocents entourés de criminels se font malmener, tandis que les criminels entourés de bons citoyens respectueux des lois passent au travers ; et compte tenu de la forte corrélation entre les situations de pauvreté et le signalement de délits, ce sont les plus modestes qui continuent d'être pris au piège de ces filets numériques », a été abandonnée par la police de Los Angeles¹⁶.

Refus de demande de prêt

Le refus de demande de prêt, basé sur des systèmes intégrant des méthodes d'apprentissage, est un autre exemple de décision tellement fréquent que les médias s'en sont emparés¹⁷. Un chercheur en IA, mondialement connu, ayant vu une demande de prêt immobilier refusée alors qu'il avait des revenus suffisants et pérennes a réussi à obtenir le code du programme et il a découvert que c'était parce que le précédent propriétaire ainsi que de nombreux voisins de la maison qu'il voulait acheter avaient des dettes !

Justice

Les finalités du système Datajust, basé sur de l'apprentissage à partir des données du ministère de la Justice, sont décrites ainsi par ce ministère¹⁸ :

« 1° La réalisation d'évaluations rétrospectives et prospectives des politiques publiques en matière de responsabilité civile ou administrative ;

2° L'élaboration d'un référentiel indicatif d'indemnisation des préjudices corporels ;

16 Cathy O'NEIL *Algorithme. La bombe à retardement* (préface de Cédric Villani), Paris, Les Arènes, 2018.

17. <https://theconversation.com/discrimination-et-ia-comment-limiter-les-risques-en-matiere-de-credit-bancaire-167008>

18 <https://www.justice.fr/donnees-personnelles/datajust>

3° L'information des parties et l'aide à l'évaluation du montant de l'indemnisation à laquelle les victimes peuvent prétendre afin de favoriser un règlement amiable des litiges ;

4° L'information ou la documentation des juges appelés à statuer sur des demandes d'indemnisation des préjudices corporels. »

Malgré une requête déposée devant le Conseil d'État au nom de plusieurs associations, celui-ci a accepté qu'une période d'expérimentation de ce système se déroule jusqu'en mars 2022. Comme pour la médecine personnalisée, la justice ne peut être rendue automatiquement, elle est « personnalisée » dans son principe, même si les juges peuvent être aidés par des systèmes automatiques, dans la mesure où ils ne se fient pas aux propositions du système, que ce soit pour gagner du temps, par pression de leur hiérarchie ou tout simplement à cause d'une trop grande confiance dans les résultats fournis par ces systèmes.

Pour terminer cette partie mentionnons que ces problèmes se posent dans d'autres domaines. Par exemple, de nombreux systèmes de traitement automatique des langues ont été construits en utilisant la plateforme GPT3¹⁹ développée par OpenAI²⁰, qui utilise des milliards de variables, et permet, par exemple, après un mot de prévoir le mot suivant... mais avant de l'utiliser, comme dans tous les exemples précédents, il est nécessaire de s'interroger sur les qualités des données, en particulier les possibles biais dans les corpus utilisés pour l'apprentissage. S'il s'agit, par exemple, d'un système de traitement de la parole, les accents et les problèmes d'élocution des locuteurs sont à analyser.

19 <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000041763205/>

20 <https://openai.com/blog/gpt-3-apps/>

Limites scientifiques

Comme tout programme informatique un logiciel intégrant des techniques d'IA est soumis à des limites intrinsèques concernant, entre autres, son exactitude et sa complexité. Les problèmes abordés avec des techniques d'IA sont généralement très compliqués (voire indécidables) et les solutions obtenues sont ainsi nécessairement imparfaites : elles sont incomplètes (ne concernent pas tous les cas) et souvent seulement approchées (ne sont pas optimales). De plus, s'il est généralement difficile de prouver (mathématiquement) qu'un programme fait bien ce qu'on voudrait qu'il fasse, dans le cas de l'IA il est également difficile de déterminer précisément son domaine de validité et sa marge d'erreur. On peut se convaincre facilement que c'est le cas dans un système de conduite autonome mais aussi dans les exemples donnés ci-dessus.

D'où l'importance fondamentale d'une tâche particulièrement difficile : faire une « bonne » évaluation du système qui concerne, elle aussi, les statistiques et les sciences humaines et sociales.

ADAPTATION DE L'UTILISATEUR AU PROGRAMME : DE L'IDENTITÉ NUMÉRIQUE À L'HOMME NUMÉRIQUE

Dans la plupart des systèmes d'aide à la décision intégrant de l'IA, l'utilisateur, sauf s'il a de sérieuses connaissances scientifiques, un esprit critique développé et le sens de ses responsabilités personnelles pour pouvoir éventuellement contester ou s'opposer à un supérieur, risque de devenir un simple intermédiaire, de s'adapter au programme et pas l'inverse (beaucoup plus coûteux). Le décideur est amené à modifier son comportement, sans probablement s'en apercevoir, parce qu'il ne comprend pas le fonctionnement du programme : ne le comprenant pas, il se fie à ce qui ne devrait

être qu'une proposition pour en faire une décision, et le système d'aide risque de se transformer ainsi, imperceptiblement, insidieusement, en un système automatique, or un homme est plus complexe qu'une centrale nucléaire.

Dans le cas d'un système grand public c'est M. ou M^{me} Tout-le-monde qui risque d'être amené à se conformer au système pour l'utiliser au mieux. Ceci n'est pas spécifiquement lié à l'IA, c'est le cas de tout outil informatique, et même de tout outil, y compris une scie ou un marteau. Mais une scie n'a pas de modèle de l'utilisateur. Les systèmes intégrant de l'IA et concernant des personnes, contiennent des représentations informatiques, implicites ou explicites, de ces personnes qui seront amenées, pour utiliser efficacement ces applications, à ressembler à ces représentations numériques. Les systèmes de recrutement nous contraignent à nous conformer au « bon employé » du système ; les systèmes de traitement des langues vont nous apprendre à parler ou écrire d'une certaine manière ; la voiture autonome ne nous apprendra pas à conduire correctement mais à comprendre comment l'automate conduit et au pire à conduire comme un programme ; le portefeuille d'identités numériques veut nous faciliter les accès à certains services en nous remplaçant par des QR codes, etc.

Terminons par un dernier exemple, les systèmes de rencontres. Jessica Pidoux a étudié 22 tels systèmes dans sa thèse soutenue récemment à l'École polytechnique fédérale de Lausanne²¹. L'objectif d'un système de rencontres est d'apparier deux personnes au travers de l'interface dans laquelle elles essaient de se décrire et de décrire leurs désirs. Elles sont amenées à se conformer, pour augmenter leurs chances de recevoir des propositions de partenaires, à la façon dont elles imaginent que le logiciel fonctionne ce qui pourrait

21 <https://infoscience.epfl.ch/record/288400?ln=en>

tendre à leur imposer un comportement amoureux... De plus, certains systèmes apprenant à partir des actions des utilisateurs, ils peuvent, comme Tay, perpétuer ou amplifier des préjugés humains. « Tinder, par exemple, recommande des matchs basés sur un modèle patriarcal », explique Jessica Pidoux, car « Le système apprend que certains hommes plus âgés préfèrent les profils de femmes plus jeunes avec un niveau d'éducation inférieur, mais l'algorithme pourrait alors suggérer le même modèle à d'autres utilisatrices de l'application. »

L'IA n'est jamais seule. Même dans les programmes de jeux ce n'est qu'un outil parmi d'autres outils informatiques, de plus des IAs sont non seulement présentes dans les téléphones, les tablettes et les ordinateurs reliés via internet mais elles sont aussi très souvent présentes dans les objets connectés, les IAs font partie de l'*Internet of Things* (IoT). Un objet connecté est un objet possédant la capacité d'échanger des données avec d'autres entités physiques ou numériques via internet. L'IoT est une extension du réseau internet à des objets du monde physique via des capteurs et des effecteurs. C'est un réseau de réseaux permettant de relier des entités informatiques et des objets physiques pour échanger, stocker et traiter automatiquement des données. Ces dix dernières années l'IoT a connu une progression fulgurante, le nombre d'objets connectés hors téléphones et ordinateurs de tout type (y compris les tablettes), présents dans tous les domaines, dépasse la moitié du nombre total d'objets connectés évalué à une centaine de milliards, et son chiffre d'affaires se compte en centaines de milliards d'euros ! Même si les données personnelles proprement dites sont protégées, le croisement des métadonnées permet d'obtenir de nombreuses informations (y compris parfois nominatives, car elles peuvent, dans certains cas, permettre de lever l'anonymat de données). Que ce soit à partir des métadonnées des e-mail

expédiés ou reçus, (destinataire, date, objet, envoyé depuis un cybercafé ou de chez soi), des sms envoyés ou reçus (lieu, longueur), des photos (format, heure, lieu, appareil), de la navigation sur le web, des données fournies par des capteurs (consommation d'électricité ou d'eau), etc. les techniques permettent de connaître des quantités d'information sur votre vie privée (sur vos positions politiques, religieuses, culturelles, sur votre santé, vos loisirs, vos achats.

Cette évolution va démultiplier les applications de l'IA en amplifiant bénéfiques et risques. Les systèmes d'aide à la décision proliféreront et comme ils incitent les utilisateurs à se comporter comme un programme, il ne faudrait pas que cela transforme leur irréductible singularité en un ensemble fini de symboles régi par des procédures de calcul !

CONCLUSION

Après l'Homme à l'image de Dieu qui fut un progrès considérable pour l'humanité, que faire pour empêcher qu'il soit transformé en un Homme numérique et l'ordinateur, ou l'IA, en un nouveau veau d'or ?

Pour éviter une catastrophe, pour éviter que la fuite en avant impulsée par les GAFAM et autres entreprises de la Tech conduise à l'émergence dans notre société de fractures bien plus dangereuses qu'une « fracture numérique », pour éviter un effondrement de nos valeurs humanistes et en particulier la disparition du libre exercice de notre faculté de jugement et d'action, les quelques mesures suivantes me semblent nécessaires.

Formation

Dans le domaine de l'aide à la décision, mais plus généralement dans toute application informatique, en plus évidemment de connaissances en informatique, des connaissances en mathématiques ainsi que la maîtrise de la

méthode expérimentale (particulièrement importantes dans le cas de systèmes prédictifs) sont nécessaires.

Toute application informatique a des conséquences sur les individus et la société donc une formation en sciences sociales et humaines est nécessaire. Ces conséquences pouvant être dramatiques, l'éthique doit avoir une place importante dans les processus de décision et dans la formation (ce que nous développerons plus loin).

Naturellement, il ne s'agit pas que tous aient les mêmes connaissances et compétences, elles sont fonction de l'implication des personnes dans de tels projets. Cependant, tous, y compris les décideurs politiques (car il n'y a pas que le commun des mortels qui prend parfois l'IA pour une voyante ou une cartomancienne) devraient avoir les connaissances de base afin de pouvoir participer à un débat démocratique concernant les objectifs et les moyens de les atteindre, ainsi que l'évolution de ces domaines.

Débat démocratique

Les grandes entreprises présentent leurs projets comme un progrès et comme une évolution technologique inévitable. Elles sont puissantes et avancent rapidement. Pour ne pas être submergés par leurs objectifs, pour décider de la société dans laquelle nous voulons vivre, éducation et débat démocratique devraient être organisés par une structure neutre, donc par l'État (et pas par les réseaux sociaux ni par Marc Zuckerberg ou Elon Musk qui sont juges et parties).

Un tel débat démocratique devrait orienter l'élaboration de lois et de règlements régulant ces divers domaines, ainsi que des normes technologiques.

Pour ne donner qu'un exemple, il faudrait imposer la possibilité de se déconnecter facilement de l'IoT ! (Sa dénomination comme *Internet du Tout* fait frémir.) Chacun devrait pouvoir – si ce n'est être comme dans une chambre anéchoïde ce qui semble techniquement difficile – se

déconnecter de tout réseau quand il le souhaite et aussi de toute batterie de téléphone, d'ordinateur (souvenons-nous de Steve Jobs ou Bill Gates mettant un scotch sur leur caméra même lorsqu'elle n'était pas activée...) de même pour les puces RFID, y compris passives.

Éthique

Il y a tellement d'applications inquiétantes que de nombreux organismes, dans les pays démocratiques, se sont dotés de commissions d'éthique concernant l'IA, et 193 pays ont accepté récemment une résolution de l'UNESCO sur ce sujet²². En France l'alliance Allistene, qui regroupe CEA, CNRS, CPU, INRIA, Mines-Télécom, Écoles d'ingénieurs, a créé la CERNA²³ la Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique. Tout programme, tout système informatique, rappelle la CERNA dans un rapport sur l'*Éthique de la recherche en apprentissage machine* devrait respecter de nombreuses propriétés : *loyauté* (un système informatique est loyal s'il se comporte comme ses concepteurs le déclarent), *équité* (l'équité d'un système informatique consiste en un traitement juste et équitable des usagers), *transparence* (la transparence d'un système signifie qu'un utilisateur peut vérifier son comportement), *traçabilité* (la mise à disposition d'informations sur ses actions suffisamment détaillées pour qu'il soit possible après coup de suivre ses actions), *explicabilité* (le fonctionnement d'un système doit être compris par un utilisateur), *responsabilité* (le donneur d'ordre ou le concepteur étant responsable si le système est mal conçu, l'utilisateur étant responsable s'il a mal utilisé le système), *conformité* (un système numérique doit être conforme à son cahier des charges, ce qui signifie que le

22 <https://news.un.org/fr/story/2021/11/1109412>

23 <https://www.allistene.fr/cerna/>

système est conçu pour effectuer les tâches spécifiées en respectant les contraintes explicitées dans ce cahier).

Toutes ces propriétés devraient être vérifiées *avant* que le système soit utilisé, en analysant son code et ses données. Cependant, ces diverses propriétés peuvent être difficiles voire impossibles à réaliser dans des systèmes utilisant des techniques d'IA (en particulier l'explicabilité pour les systèmes d'apprentissage numérique). De plus, il y a une contradiction entre la rapidité des innovations techniques et la lenteur de leur évaluation qui nécessite la compréhension des conséquences, le contrôle non seulement de la conformité d'un programme aux lois mais aussi des propriétés que devrait avoir tout système informatique, ainsi que des procédures démocratiques, nécessairement lentes, afin que les citoyens puissent décider librement de leur destin.

L'urgence, en particulier de conquérir des marchés, dicte souvent sa loi, de plus, la restriction, voire la suppression, de crédits dans de nombreux secteurs, a conduit au remplacement de personnes par des programmes.

Certains ont poussé des cris d'orfraie contre les robots tueurs, ils avaient raison. Le contrôle par des personnes d'armes automatiques n'est pas toujours faisable, par exemple le chemin suivi par un drone dans un essaim de drones menant une attaque ne peut pas être contrôlé par des humains, et les conséquences peuvent être dramatiques, mais il n'y a pas que les robots tueurs... voici un extrait de ce même rapport de la CERNA :

« De façon caricaturale, un véhicule autonome qui se trouverait à choisir entre sacrifier son jeune passager, ou deux enfants imprudents, ou un vieux cycliste en règle, pourrait être programmé selon une éthique de la vertu d'Aristote – ici l'abnégation – s'il sacrifie le passager, selon une éthique déontique de respect du code de la route s'il sacrifie les enfants, et selon une éthique

conséquentialiste s'il sacrifie le cycliste – ici en minimisant le nombre d'années de vie perdues.

Le propos n'est pas ici de traiter de telles questions qui relèvent de la société toute entière... »

On peut se demander s'il est éthiquement responsable de continuer à essayer de construire des voitures autonomes, et plus généralement des véhicules autonomes ne circulant pas sur des voies spécialement aménagées. En effet, s'il y a un obstacle, Météor, Orlyval, un tramway ou un train n'ont pas à choisir, ils freinent pour éviter le choc ! Il semble inadmissible qu'on puisse demander à un algorithme de choisir les personnes à sacrifier, et donc demander à un informaticien de programmer de tels choix !

Le livre de Cathy O'Neil déjà cité – qui ne demande aucune compétence technique pour sa lecture – contient de nombreuses études bien documentées sur ce sujet et Cédric Villani termine ainsi la préface de ce livre : « Elle [Cathy O'Neil] est bien décidée à ne pas baisser les bras, et à faire porter sa voix autant qu'il le faut pour que nous puissions conserver notre humanité. Écoutez-la attentivement. »

Pour éviter les dangers inhérents à de nombreuses applications de l'IA, et plus généralement de l'informatique, une formation à l'éthique des étudiants, enseignants, ingénieurs, concepteurs et utilisateurs serait nécessaire.

Risques

Si de telles mesures n'étaient pas prises, la méfiance vis-à-vis de la science et de toute rationalité s'amplifierait avec toutes ses conséquences : impossibilité d'un débat public rationnel conduisant à une augmentation des fractures de notre société, sabotages (par exemple un sondage de septembre 2020²⁴ nous apprend que si moins de 20 % des personnes

24. <https://www.ifop.com/publication/les-francais-et-les-habitants-des-grandes-villes-soutiennent-ils-le-dploiement-de-la-5-g-en-france/>

interrogées se disaient favorables à la destruction des antennes-relais 5G, 48 % se disaient favorables à la suspension du déploiement de la 5G, un débat démocratique aurait sans doute limité les nombreuses destructions d'antennes 5G ou de matériels d'Enedis), refus des vaccins, etc.

Pour terminer, je voudrais revenir un instant à l'intelligence artificielle en tant que science en rappelant la première phrase de Turing dans son article fondateur de l'IA²⁵, dans lequel il propose de considérer la question « *Can machines think ?* » et dont l'une des dernières phrases est : « Nous pouvons espérer que les machines finiront par concurrencer les hommes dans tous les domaines purement intellectuels (*purely intellectual fields*). »

Donc il n'est pas question de l'homme dans sa totalité, ni même de ses *compétences* intellectuelles mais seulement de ses *performances* dans les « *purely intellectual fields* ». Pour simplifier, Turing proposait deux voies. La première était celle des jeux, et le succès d'AlphaGo le programme qui a battu Lee Sedol, 18 fois champion du monde, a perturbé le sens critique de certains, ce programme, répétons-le, est incapable d'expliquer pourquoi il a gagné, de commenter une partie, et est encore plus incapable d'apprendre à un humain à jouer au Go ! La deuxième voie était celle de la construction d'un système ayant les capacités humaines concernant le langage c'est-à-dire capable d'apprendre n'importe quelle langue et de communiquer et sur cette voie, on n'en est qu'au début !

L'intelligence artificielle en tant que domaine scientifique ne devrait pas inquiéter, au contraire elle permet de mieux comprendre certaines de nos compétences et la quête d'une IA générale ou forte non liée aux compétences humaines, est un domaine de recherche passionnant comme tous les domaines concernés par cette conférence et par tant d'autres. Mais il faudrait que la société se donne les moyens pour contrôler

25 Alan Turing, *Computing Machinery and Intelligence*, Mind, 1950.

démocratiquement et sérieusement ses applications dont les conséquences sur l'humanité sont capables du meilleur comme du pire.

TRANSHUMANISME, HUMANISME, PHILOSOPHIE DES LUMIÈRES DÉFINITION ET RELATIONS

*par M. Jean-Claude MARCHAL
de l'Académie de Stanislas*

Le transhumanisme, et ce néologisme plus récent le posthumanisme, mobilisent les biotechnologies, la robotique, les sciences cognitives et l'Intelligence Artificielle (IA) pour « créer » un homme nouveau voire une autre espèce humaine. R. Kurzweil¹ est l'un des prophètes de ce *credo* anthropotechnique. Cet enseignant au MIT s'est fait connaître par ses travaux et découvertes sur la reconnaissance optique des caractères, la reconnaissance vocale, les synthétiseurs électroniques. À partir de ses travaux sur l'IA il crée le concept de « singularité technologique »². Ce concept développe l'idée selon laquelle la croissance technologique de notre civilisation humaine, suivant les mêmes principes de la loi de Moore (voir *infra*) arrivera à une IA dépassant l'intelligence humaine, s'organisant en réseaux et pouvant à terme, l'homme s'hybridant avec la machine, lui donner le pouvoir de se reproduire d'une façon autonome... Ce faisant la singularité nous fait entrer de plain-pied dans la futurologie. Le terme de transhumanisme est ici inapproprié, raison pour laquelle nous parlerons dans cette perspective d'IA forte, de posthumanisme puisqu'il ne s'agit pas d'améliorer l'humanité mais d'en changer ! Ces anthropo-technologues se réclament d'un humanisme hérité du mouvement intellectuel de la Renaissance. Quels humanistes, quelles racines humanistes nourrissent ce transhumanisme ? Peut-on comparer la perte

1 KURZWEIL, 2013.

2 KURZWEIL, 2005.

d'influence de la scolastique médiévale, le nouvel essor intellectuel de la Renaissance avec les perspectives inquiétantes de la singularité technologique dont l'avènement serait à notre porte ?

DÉFINITIONS ET PERSPECTIVES

Le mot « humanisme » est polysémique. Le *Dictionnaire* de Furetière (1690) définit déjà le mot « humaniste » comme un homme qui « sait bien les lettres humaines, les humanités » et donne comme exemple « cet écolier est bon humaniste »³. Le dictionnaire *Petit Larousse* (2004) donne deux définitions du mot humanisme qui est pour la première une « position philosophique qui met l'homme et les valeurs humaines au-dessus des autres valeurs »⁴. Dans le prolongement de cette première définition : « L'humanisme s'est donné pour mission exclusive d'éclairer et de perpétuer la primauté de l'homme sur l'individu. L'humanisme a prêché l'homme »⁵.

La seconde définition du *Petit Larousse* bien que plus ancienne, est plus restrictive et circonscrit l'humanisme dans une philologie, un « mouvement intellectuel constitutif de la Renaissance, né en Italie au XIV^e siècle, qui gagna progressivement toute l'Europe pour s'épanouir au XVI^e siècle et qui fut marqué par le retour aux textes antiques dont il tira des modèles de vie, de pensée. (Principaux représentants : Pétrarque, Ficin, Pic de la Mirandole, Lefèvre, Reuchlin, Érasme) ».

Nous voyons déjà le nom de Jean Pic de la Mirandole. Nous verrons comment l'interprétation de son œuvre a pu abuser les philosophes humanistes. Aujourd'hui par extension, l'humanisme est « le résultat d'une formation qui embrasse la

3 FURETIÈRE, 1690.

4 PETIT LAROUSSE, 2003, p 518.

5 SAINT-EXUPÉRY, 1942, p 377.

culture littéraire, fondée essentiellement sur les cultures grecque et latine, et la culture scientifique »⁶. La meilleure approche sémantique du mot transhumanisme est certainement donnée par le philosophe G. Hottois :

« Le transhumanisme est un courant d'idées qui encourage le développement et l'usage des technologies matérielles afin d'améliorer, augmenter, étendre indéfiniment les capacités et performances (cognitives, physiques, émotionnelles) de l'individu qui doit rester libre d'y recourir « ou non »⁷. »

Le transhumanisme n'est pas une métaphysique, c'est une « philosophie de la technique et de l'action qui ne se préoccupe de ce qui est qu'afin de déterminer ce qui est possible et souhaitable »⁸. G. Hottois rapporte avoir : « utilisé les termes de "transhumain, posthumain, abhumain" en association avec "technoscience" dès les années 1970 »⁹.

Depuis cette période, ce courant de pensée, cet avenir qui apparaît à certains comme de la science-fiction est l'objet de nombreuses controverses de la part des philosophes, des scientifiques de toutes tendances, des politiques de tout bord, des hommes d'Églises, et... des auteurs de science-fiction. Pour mieux comprendre il faut expliciter le terme d'extropie et ses principes créés en 1995 par Max More. Pour la petite histoire Max T. O'Connor a pris le nom de Max More pour mieux proclamer sa volonté d'amélioration. Il a été pendant dix ans le président d'ALCOR¹⁰ dont les expériences de

6 www.cnrtl.fr/definition/humanisme

7 HOTTOIS, 2018, p. 75.

8 *Ibid.*, p. 79.

9 *Ibid.* p. 73.

10 *Alcor Life Extension Foundation* plus souvent appelée Alco est une société américaine située à Scottsdale, Arizona. Alcor est un centre de recherche – « *Cryonics* » – consacré à l'étude de la congélation du corps et du cerveau humains dans l'azote liquide après la mort légale dans l'espoir d'une résurrection en pleine santé dans l'éventualité de nouvelles technologies.

conservation par congélation des cerveaux après la mort légale (*CRYONICS*) sont regardées avec beaucoup de scepticisme par la communauté scientifique. La philosophie de l'extropie est la philosophie originale du transhumanisme. Max More la définit comme une volonté d'amélioration et de croissance d'un système vivant ou organisationnel. Le mot « extropie » se veut l'antinomie d'« entropie » en référence au deuxième principe de la thermodynamique¹¹ grossièrement assimilé au chaos¹². Il a établi en 1993 les *Principes extropiens*. L'extropie exprime une métaphore, plutôt qu'un terme technique, et n'est donc pas simplement le contraire de l'entropie.

Dans *Principes extropiens* revisités en 2003, il écrit :

« L'extropianisme est un transhumanisme. Les Principes extropiens définissent une version ou marque particulière de la pensée transhumaniste. Comme les humanistes, les transhumanistes privilégient la raison, le progrès et les valeurs centrées sur notre bien-être plutôt que sur une autorité religieuse externe. Les transhumanistes étendent l'humanisme en mettant en question les limites humaines par les moyens de la science et de la technologie, combinés avec la pensée critique et créative. Nous mettons en question le caractère inévitable du vieillissement et de la mort, nous cherchons à améliorer progressivement nos capacités intellectuelles et physiques, et à nous développer émotionnellement. Nous voyons l'humanité comme une phase de transition dans le développement évolutionnaire de l'intelligence. Nous défendons l'usage de la science pour accélérer notre passage d'une condition humaine à une condition transhumaine, ou posthumanisme¹³. »

11 Le terme d'entropie a été utilisé pour la première fois en 1865 par Rudolf Clausius en énonçant le deuxième principe de la thermodynamique.

12 En opposant d'une façon trop schématique ordre (extropie) et désordre (entropie) Max More scotomise la relation logarithmique entre entropie et probabilités établie par la fameuse et simple équation de L. Boltzmann.

13 MORE, 2003.

Il est l'un des premiers, avec sa femme Natasha Vita-More, à employer le terme de transhumanisme à cet usage spécifique d'amélioration par la science et le progrès. Il fait référence explicite à un humanisme qui aurait ébranlé l'autorité religieuse en faisant appel à la raison, au « progrès » et aux « valeurs centrées sur notre bien-être. » Cette filiation est-elle appuyée par la lecture des intellectuels auxquels il fait allusion ? Le concept d'une médecine « méliorative », la promesse d'un vieillissement prolongé dans un corps jeune dont la mort ne pourrait être qu'accidentelle est-elle soutenable par les possibilités de l'IA, des nanotechnologies et du génie biologique ? Ce débat d'idées pourrait n'apparaître que comme une séduisante spéculation intellectuelle s'il ne s'appuyait sur les travaux scientifiques les plus complexes et les plus sophistiqués autorisés par les possibilités exponentielles actuelles du calcul informatique. La loi à laquelle il a été fait allusion dans l'introduction a été énoncée par G. E. Moore en 1965¹⁴. Elle postule qu'à coût constant les semi-conducteurs doubleront tous les ans. Dix ans plus tard il revient sur cette première loi en recalculant le doublement des transistors sur une puce de microprocesseur tous les deux ans. Cette loi de Moore a été relayée au début du XXI^e siècle grâce aux possibilités offertes par les nanotechnologies et la puissance conjointe des algorithmes. En 2002 paraît le rapport de la *National Science Foundation* (NSF) et du *Department of Commerce* (DOC) : *Converging Technologies for Improving Human Performance : Nanotechnology, Biotechnology, Information Technology and Cognitive Science*¹⁵. Il sera édité en 2003. Le projet est clair : la convergence des technosciences a pour cible l'amélioration des performances humaines. Le rapport met en évidence tout ce qu'il est et sera possible de faire puisqu'au niveau nanométrique, il n'y a plus de

14 MOORE, 1965 p. 114-117.

15 ROCO, 2002.

différence absolue entre matière inerte, vivante, pensante ; entre naturel et artificiel ; entre homme, machine, animal : « La convergence des diverses technologies est basée sur l'unité matérielle à l'échelle nanométrique et sur l'intégration des technologies à partir de cette échelle »¹⁶ cité par G. Hottois (2014)¹⁷.

En 2016 Brainbridge revient en détail sur ce concept de « convergence » issu des NBIC :

« La convergence est un modèle de changement dans l'évolution des sciences et technologies et qui les intègre à la société. Elle fournit un cadre et une approche pour faire progresser non seulement la science et l'ingénierie, mais aussi le commerce et la politique. La convergence est une intégration profonde des connaissances, des outils et de tous les domaines pertinents de l'activité humaine pour permettre à la société de répondre à de nouvelles questions, de créer de nouvelles compétences et technologies, et globalement de changer respectivement les écosystèmes physiques ou sociaux. [puis il dresse le portrait d'une société digitale] Dans l'avenir, les possibilités de convergence comprendront une nouvelle plate-forme pour la coévolution homme-technologie à l'aide d'assistants cognitifs intelligents et l'expansion de la science et de la technologie citoyennes¹⁸. »

Ainsi les NBIC auraient la potentialité grâce à leur convergence de modifier les deux paradigmes médicaux : traiter et prévenir ainsi que le pouvoir de reculer le vieillissement et la mort.

16 ROCO, 2003.

17 HOTTOIS, 2014, emplacement 66.

18 Brainbridge, 2016, p. 1 et 2.

LA MÉDECINE, LE VIEILLISSEMENT ET LA MORT

La médecine

La médecine traditionnelle se propose deux missions : prévenir les maladies et guérir des états pathologiques. Dans un futur proche ces exigences de prévention et de guérison se renforceront d'une exigence d'amélioration continue de l'individu. Il lui sera demandé non seulement de prévenir, de guérir mais d'augmenter les performances physiques, intellectuelles, sexuelles et génétiques de l'individu sain. Le paradigme a changé : non seulement guérir mais augmenter, améliorer. Le double sens : « augmentation » et « amélioration » vient de la traduction ambivalente du mot anglais *enhancement* qu'utilise le philosophe américain A. Buchanan dans deux de ses ouvrages parus en 2011 : *Beyond Humanity ? The Ethics of Biomedical Enhancement*¹⁹ et *Better Than Human : The Perils and Promise of Enhancing Ourselves*²⁰. Pour ce philosophe la question de l'amélioration (*enhancement*) ne doit pas se poser puisqu'elle est déjà là, est appliquée et tire sa légitimité de son usage (biotechnologies des interfaces neuronales directes²¹, chirurgie esthétique par exemple). Ce qui est moins clair ce sont les conditions qui vont permettre à ces améliorations de se développer sans danger somatique, psychologique ou génétique. La médecine a, on l'oublie parfois dans un débat qui privilégie trop souvent l'émotion et le spectaculaire aux dépens de la raison, ses propres limitations et effets indésirables. Refuser de prendre en compte l'existence de ces moyens amélioratifs occulterait dans ce même aveuglement l'étude scientifique critique

19 BUCHANAN, 2011, p.1.

20 BUCHANAN, 2011, p.1.

21 Une interface neuronale directe – abrégée IND ou BCI ou encore ICM (interface cerveau-machine, ou encore interface cerveau-ordinateur) est une interface de communication directe entre un cerveau et un dispositif externe (un ordinateur, un système électronique...).

objective des effets secondaires et délétères possibles. Les consignes de sécurité applicables au circuit du médicament à visée thérapeutique ou aux dispositifs médicaux implantables par exemple doivent pouvoir l'être de la simple prise d'une drogue aux manipulations génétiques dans le but d'améliorer une performance. C'est ici que la réflexion éthique prend tout son sens, reflète sa complexité et endosse une lourde responsabilité. Si le scientifique ou le philosophe du XXI^e siècle s'allège sans trop de difficulté de certains principes religieux incompatibles avec la rigueur scientifique, il lui est parfois plus difficile d'éviter le piège tendu par la métaphysique lorsqu'elle rejoint un problème physique en apparence insoluble. L'exclamation d'A. Einstein : « Dieu ne joue pas aux dés ! » lorsqu'il se penche sur les premiers développements de la mécanique quantique, opposant son « réalisme » aux travaux « non réalistes » de N. Bohr, en est une illustration. Plus que les tabous religieux, l'évolutionnisme est parfois compris et confondu avec le concept déterministe voire téléologique d'une nature qui tiendrait lieu de morale ou qui ferait intervenir un grand architecte, chaque mutation ajoutant sa brique à la construction d'un univers en perpétuelle amélioration. Or l'évolution n'est ni linéaire, ni synonyme de progrès, il n'y a pas de bonne ou de mauvaise nature. L'évolution est « buissonnante »²², elle débouche souvent sur des impasses biologiques, puis au gré de mutations, d'interactions épigénétiques ou d'interférences environnementales la nature bricole sans préoccupation ontologique, déterministe ou morale. Il est de la responsabilité de l'homme d'aujourd'hui de comprendre et d'assumer sa « position résultant d'un cerveau monstrueux qui l'a poussé, hors de la nature [et] l'en a comme privé, tout en lui conférant un pouvoir sans précédent sur la nature à laquelle il ne cesse

22 PROCHIANTZ, p.24.

d'appartenir puisqu'il en est le produit évolutif. "Anature"²³ par nature ou encore "être et ne pas être un animal", deux façons identiques d'énoncer la conception [...] de *sapiens*, et il va sans dire qu'elle ne va pas sans exiger de notre espèce une responsabilité particulière vis-à-vis de cette nature et de tous ses composants vivants et non vivants »²⁴.

Mort et vieillissement

Dans cette perspective le vieillissement et la mort toujours subis comme des événements inévitables sont étudiés à la lumière de ce nouveau paradigme. S'y préparer, s'y résigner ou la hâter en se donnant l'illusion d'en maîtriser l'heure d'échéance apparaissent comme les seuls comportements aidant à accepter l'inacceptable. Le transhumanisme remet en question les limites naturelles et culturelles de ces événements inévitables :

« Le recul accéléré de la mort sera la plus vertigineuse conséquence de ce que les transhumanistes appellent la "grande convergence NBIC", c'est-à-dire les synergies entre Nanotechnologies, Biologie, Informatique et sciences Cognitives : l'idée que la mort est un problème biologique à résoudre et non une réalité imposée par la nature ou par la volonté divine va s'imposer. Avec l'exploration de l'Univers, l'euthanasie de la mort va devenir l'ultime frontière pour l'Humanité²⁵. »

Dès lors la mort n'est plus une réalité insurmontable ou une énigme impénétrable mais une question biologique avec ses réponses. « La finitude de la personne est empirique, non ontologique »²⁶. La question du vieillissement et de la mort est aussi vieille que l'humanité pensante. Le livre de N. Bostrom :

23 Anature utilise le « a » privatif. Note de l'auteur.

24 PROCHIANZ p. 17-18.

25 ALEXANDRE, 2011, p. 12.

26 HOTTOIS, 2014, emplacement 354.

*A History of Transhumanist Thought*²⁷ souligne l'importance anthropologique des cérémonies funéraires dès notre préhistoire. Cependant la croyance en une vie après la mort n'exclut pas l'effort d'une recherche pour allonger la durée de la vie et il illustre cette quête en nous racontant l'épopée de Gilgamesh, récit sumérien datant de 1700 avant J.-C. qui narre la recherche d'un roi en quête d'immortalité. Après avoir trouvé l'herbe magique il n'a pas le temps de l'avalier qu'elle lui est dérobée par un serpent (déjà lui !). Plus tard l'histoire de l'humanité s'enrichit d'explorateurs, d'alchimistes partis à la recherche de Fontaines de Jouvence, d'Élixirs de Vie et autres Jardins des Hespérides. D'autres comme certaines écoles ésotériques taoïstes en Chine se font fort d'atteindre l'immortalité du corps et de l'âme par le contrôle harmonieux des forces de la nature. En dépit de tous ces efforts, la lutte pour l'immortalité s'est avérée jusqu'à ce jour illusoire. Longtemps cette quête d'une vie allongée de façon acceptable et pourquoi pas infinie a été stigmatisée par une ambivalence dont la mythologie s'est fait le témoin : fascination et *hubris*. Le feu volé aux dieux apportera un progrès immense à l'humanité mais Prométhée²⁸ en sera la victime expiatoire. Icare sans procédé magique vole de plus en plus haut et chute victime de son imprudence et de son *hubris*. Il poursuit :

« L'ancien monde et sa philosophie scolastique désuète qui dominaient l'Europe au Moyen Âge ont cédé la place à une vigueur intellectuelle renouvelée à la Renaissance. L'être

27 BOSTROM, 2005, p. 1-4.

28 Bostrom fait une interprétation extensive du mythe tel qu'il est exposé dans le *Protagoras* de Platon. Prométhée est à l'origine de la capacité technique et sapientielle accordée à l'homme par les dieux. L'homme est nu et démuné. Prométhée vole à Héphaïstos et à Athéna le feu (métaphore de la technique) et la sagesse pour lui permettre de développer ses propres outils de survie au milieu du monde. Ainsi il n'y a pas de rapport direct entre Prométhée et l'accès de l'homme à une quelconque émancipation. L'homme reste débiteur des dieux et il ne possède de qualités propres qu'à la suite de ce larcin.

humain et le monde naturel sont redevenus des objets d'étude légitimes. L'humanisme de la Renaissance encourageait les hommes à s'appuyer sur leurs propres observations et leur propre jugement plutôt que de s'en remettre en toutes matières aux autorités religieuses. L'humanisme de la Renaissance a également créé l'idéal de la personne qui vit harmonieusement, celle qui est développée scientifiquement, moralement, culturellement et spirituellement²⁹. »

Dans son histoire de la pensée transhumaniste il situe le point de repère de cette Renaissance à la date de publication de l'*Oratio de hominis dignitate* de Pic de la Mirandole (1486).

HUMANISME, TRANSHUMANISME : FILIATION OU ASSONANCE ?

À partir de nos définitions de l'humanisme il convient de se pencher sur l'interprétation qui en a été faite par les transhumanistes à travers les œuvres de Pic de la Mirandole et de Condorcet, pour le premier en nous référant à la définition traditionnelle, pour le deuxième en s'adressant à la définition qui place les valeurs humaines au-dessus des autres valeurs.

Jean Pic de la Mirandole (1463-1494).

La compréhension par N. Bostrom de la pensée de Pic de la Mirandole tient à sa confiance aveugle dans l'interprétation de l'œuvre de Pic par J. Burckhardt, un des plus grands historiens d'art spécialistes de cette période. Jacob Burckhardt (1818-1897) est avec Jules Michelet (1798-1874) l'« inventeur »³⁰ de la Renaissance. Dans son ouvrage sur la Renaissance italienne paru en 1860³¹, et pour conformer

29 BOSTROM, 2005, p. 2.

30 VALCKE, 1994, p. 1.

31 BURCKHARDT, 1860.

l'œuvre de Pic à l'esprit de la Renaissance, Burckhardt a rétréci les publications de Pic à une œuvre que la postérité devait retenir sous le nom de *Oratio de hominis dignitate* (*Discours sur la dignité de l'homme*) (écrit en 1486 mais publié en 1496 soit deux ans après la mort de Pic). Il s'agit d'un dialogue entre le Créateur et Adam tel qu'il apparaît dans l'ouvrage de Pic et dont la traduction a été ainsi proposée :

« Je t'ai placé au milieu du monde afin que tu puisses plus facilement promener tes regards autour de toi et mieux voir ce qu'il renferme. En faisant de toi un être qui n'est ni céleste ni terrestre, ni mortel ni immortel, j'ai voulu te donner le pouvoir de te former et de te vaincre toi-même ; tu peux descendre jusqu'au niveau de la bête et tu peux t'élever jusqu'à devenir un être divin. En venant au monde, les animaux ont reçu tout ce qu'il leur faut, et les esprits d'un ordre supérieur sont dès le principe, ou du moins bientôt après leur formation, ce qu'ils doivent être et rester dans l'éternité. Toi seul tu peux grandir et te développer comme tu le veux, tu as en toi les germes de la vie sous toutes ses formes³². »

Ceci est le texte de référence qui relie Renaissance et transhumanisme. Il constituerait pour reprendre l'expression d'E. Garin : « L'Évangile de la liberté radicale de l'homme³³ » (cité par L. Valcke³⁴). On le considère d'abord comme le premier manifeste de la Renaissance. Mais la pensée de Pic ne saurait se résumer dans l'*Oratio* : « Qui donc a entendu parler de l'*Heptaplus*, du *De ente et uno*, de l'*Apologia*... ? Tout au plus sait-on, parce que Kepler en parle, que Pic a écrit une critique de l'astrologie. À l'évidence, une question s'impose : dans quelle mesure l'*Oratio* est-elle représentative de l'ensemble de l'œuvre de Jean Pic ? »³⁵.

32 VALCKE, 2005, p. 14.

33 GARIN, 1963.

34 VALCKE, 2005 p. 15.

35 VALCKE, 1994 p. 1.

Ce texte reflète-t-il sa doctrine et sa proclamation en tant que « manifeste de l'humanisme rinascimental tout entier : vision prométhéenne de l'homme libre, maître de son destin, désormais seul responsable de son devenir et de ses choix »³⁶ ? En réalité l'*Oratio de hominis dignitate* ne propose aucune idée nouvelle. Même si Pic y donne une idée grandiose de l'homme, il en reste à une conception théocentrique traditionnelle dans laquelle ce n'est pas l'homme qui est au centre de la création mais Dieu. L'homme n'est libre que dans la mesure où il accomplit le pouvoir divin. L'homme n'est pas créateur du monde ou de valeurs morales mais il est récipiendaire de valeurs déjà admises et reçues de la loi divine. L'homme est là pour contempler l'ordre divin des valeurs et de l'univers et non pour créer ou renouveler un monde déjà là. La liberté de l'homme est une liberté d'acceptation ou de refus mais jamais une liberté de création ou d'affirmation. L'homme peut seulement découvrir l'ordre de sa nature mais il ne peut jamais le modifier ou prendre la place centrale. L'homme n'est jamais autonome. Il n'est jamais lui-même sa propre loi. Il est ainsi beaucoup plus proche d'une tradition médiévale et scolastique que d'une tradition humaniste de laquelle on l'a trop facilement rapproché. Le style même de ce texte rhétorique est écrit, selon les érudits spécialistes de son œuvre, dans un latin littéraire

« conforme à l'attente des milieux humanistes » et « que par contre, après avoir donné la preuve éclatante de sa parfaite maîtrise du latin classique, il n'en ait pas moins choisi d'écrire ses autres ouvrages en latin scolastique, voilà qui, surtout dans le contexte de l'époque, est extrêmement révélateur [...] c'était là une façon de dire qu'il ne fallait pas prendre l'*Oratio* trop au sérieux, en tout cas pas à la lettre, et que, par ailleurs, si on voulait trouver

36 *Ibid.* p. 2.

sa vraie pensée, ce serait dans ses autres écrits qu'on aurait à la chercher³⁷. »

Ce n'est donc pas dans l'*Oratio* que l'on trouve la vraie pensée de Pic mais dans ses autres ouvrages, cet écrit ne doit pas être pris à la lettre comme Burckhardt semble l'avoir fait mais comme une sorte d'exercice de style de la part de son auteur « on n'hésitera pas, au besoin et à l'occasion, à donner aux traductions les quelques coups de pouce nécessaires pour rendre les textes plus conformes à cette attente et, dès lors, Jean Pic devint l'idéal exemplaire, le prototype de l'humaniste du *Quattrocento*, et cette fonction paradigmatique fera partie de son mythe... »³⁸.

C'est à partir de ce texte et de l'interprétation qu'en a fait J. Burckhardt, celui d'un homme nouveau, qu'a pris racine le mythe mirandolien. La publication posthume de l'*Oratio*, les circonstances mystérieuses qui entourent sa mort survenue à 31 ans seulement feront le reste. Les transhumanistes s'en sont saisis³⁹ en en prolongeant l'interprétation grâce à l'éclairage scientifique séduisant des NBIC sur le thème de la liberté de l'homme. Derrière cet acronyme la Science prétend ramener les questions métaphysiques de la mort et du vieillissement à un problème biologique. Dans le droit fil de son interprétation de l'humanisme N. Bostrom écrit :

« On dit souvent que le siècle des Lumières a commencé avec la publication du *Novum Organum* de F. Bacon *the new tool* – "le nouvel outil" (1620), qui propose une méthodologie scientifique basée sur une enquête empirique plutôt que sur un raisonnement à priori. F. Bacon préconisait d'effectuer toutes choses possibles, c'est-à-dire utiliser la science pour maîtriser la nature afin d'améliorer les conditions de vie des êtres humains. L'héritage de la

37 *Ibid.* p. 3.

38 *Ibid.* p. 2.

39 BOSTROM, 2005, p. 2.

Renaissance se combine avec l'influence d'Isaac Newton, Thomas Hobbes, John Locke, Emmanuel Kant, du marquis de Condorcet et d'autres pour former la base de l'humanisme rationnel, qui met l'accent sur la science empirique et la raison critique – plutôt que sur la révélation et l'autorité religieuse – comme moyens d'apprendre sur le monde naturel, notre place en son sein et de fournir une base pour la moralité. Le transhumanisme a des racines dans l'humanisme rationnel⁴⁰. »

« L'humanisme rationnel » dont parle N. Bostrom s'éloigne de la définition proposée et admise et à laquelle il donnerait simplement un développement éclairé par le regard de la science – d'où l'adjectif « rationnel » – surtout il réduit les recherches scientifiques, philosophiques de ces grandes personnalités à des travaux qui ne constituent pas l'essentiel de leur œuvre. Au Siècle des Lumières l'histoire de la pensée humaniste se focalise sur l'œuvre posthume de Nicolas de Condorcet : *Esquisse d'un tableau historique de l'esprit humain* (1795). Ici encore l'œuvre de ce grand intellectuel est ramenée à une publication posthume anecdotique mais qui a contribué à rendre Condorcet célèbre.

Nicolas de Condorcet (1743-1794)

Le transhumanisme se rattache à la pensée de Condorcet par le célèbre passage de l'*Esquisse* où il reprend le thème du perfectionnement sans fin de l'homme :

« Nous pourrions déjà conclure que la perfectibilité de l'homme est indéfinie, et cependant jusqu'ici nous ne lui avons supposé que les mêmes facultés naturelles, la même organisation ; quelle serait donc la certitude, l'étendue de ses espérances, si l'on pouvait croire que ces facultés naturelles elles-mêmes, cette organisation sont aussi susceptibles de s'améliorer ? ... Sans doute l'homme ne

40 Ibid. p. 2.

deviendra pas immortel (... mais la longévité devrait toujours s'accroître) dans l'immensité des siècles⁴¹. »

L'œuvre et la pensée de Condorcet sont réduites à cette citation qui même si elle reste célèbre ne doit pas faire oublier que Condorcet était d'abord un mathématicien, physicien et économiste connu pour ses travaux sur les systèmes de vote, le jury pénal. Défenseur du droit des femmes avant la lettre (leur droit de vote en particulier) il est l'inventeur du droit d'auteur. Réduire sa pensée comme le fait N. Bostrom à ce passage d'une œuvre inachevée dont il n'a pu écrire que la préface – d'où le titre commençant par « Esquisse... » – en raison de sa mort prématurée en 1794, c'est donner à la pensée humaniste une signification bien trop large qui pourrait simplement s'en tenir à la prévalence des « valeurs humaines » en général.

ORIGINES DU MOT TRANSHUMANISME

Le terme « transhumain » est utilisé pour la première fois par P. Teilhard de Chardin (1881-1955) en 1959⁴². C'est bien plus par le contenu de son œuvre que par l'emploi de ce terme qu'il a été invoqué par les transhumanistes des années 1980 dans leur volonté de filiation (Barrow⁴³, Moravec⁴⁴) et plus près de nous par R. Kurzweil en 1999⁴⁵. Teilhard de Chardin professe une forme d'humanisme chrétien. La théodicée chrétienne de saint Augustin, la théorie de l'évolution de Charles Darwin, la biosphère de Vernadski⁴⁶ sont réunies dans

41 CONDORCET, 1795, p. 222.

42 *L'avenir de l'homme*, 1959. Les œuvres de P. TEILHARD DE CHARDIN ont toutes été publiées à titre posthume puisqu'il est décédé en 1955.

43 BARROW, 1986.

44 MORAVEC, 1988.

45 KURZWEIL, 1999.

46 Vladimir Ivanovitch Vernadski (1863-1945), chimiste et minéralogiste russe, a créé la théorie de la biosphère dans laquelle toute vie vient d'un unique matériau : la biosphère qui serait la version non théologique de la noosphère de Teilhard.

un syncrétisme où le *Phénomène humain* de Teilhard de Chardin (1955)⁴⁷ doit être pensé comme une étape de l'évolution qui conduit au déploiement de la noosphère et à sa convergence vers le « point *Oméga* » et l'avènement de la figure du « Christ cosmique ». Teilhard de Chardin est moniste : matière et esprit ne sont que deux faces d'une même réalité. La pensée transhumaniste le rejoint et s'en réclame car elle professe qu'au niveau nanométrique il n'y aurait plus de différence entre esprit et matière, ce qui participe à la Convergence. Le monisme de Teilhard de Chardin, son approche holiste s'opposent cependant au transhumanisme contemporain qui est un transhumanisme séculier même si, en Occident, il ne peut ignorer la pensée chrétienne qui faufile son histoire. Julian Huxley (1887-1975) forgera le terme de transhumanisme au sens où les transhumanistes l'entendent de nos jours. Premier directeur de l'UNESCO, il a développé les possibilités de l'eugénisme comme moyen d'amélioration de la population humaine. Frère d'Aldous Huxley (1894 - 1963) l'auteur du *Brave New World (Le meilleur des mondes)*⁴⁸ (1932), roman qui radicalise l'eugénisme et autres formes d'ectogénèse. C'est J. Huxley qui va substituer le mot « *transhumanism* »⁴⁹ à celui de « *evolutionary humanism* » dans cette déclaration de foi : « *"I believe in transhumanism" : once there are enough people who can truly say that the human species will be on the threshold of a new kind of existence, as different from ours as ours is from that of Peking man. It will at last be consciously fulfilling its real destiny* »⁵⁰ J. HUXLEY, 1957.

47 TEILHARD DE CHARDIN, 1955.

48 HUXLEY, 1932.

49 HUXLEY, 1957, p. 17.

50 « Je crois au transhumanisme » : une fois qu'il y aura suffisamment de gens qui pourront vraiment affirmer cela, l'espèce humaine sera au seuil d'une nouvelle sorte d'existence, aussi différente de la nôtre que la nôtre

Pour J. Huxley le transhumanisme est un humanisme naturaliste, moniste et non dualiste, évolutionniste et non statique. Sa réflexion est surtout dominée par la biologie (en particulier ses réflexions relatives à l'eugénisme positif) contrairement au transhumanisme du XXI^e siècle qui fait une large place aux NBIC. Pour J. Huxley la biologie obéit à un évolutionnisme dont l'épistémologie, la méthodologie est basée sur la coopération entre l'homme et la Nature où l'homme est dans la position dirigeante à la différence de l'évolutionnisme de Darwin. C'est le nouveau paradigme.

CONCLUSION

Le transhumanisme voudrait être une étape de la pensée humaine se réclamant de périodes-pivots de la pensée occidentale Renaissance ; et siècle des Lumières, au simple prétexte qu'ils se seraient débarrassés de toute sujétion religieuse : « La grande majorité des transhumanistes se présentent comme agnostiques ou athées, laïques et libres-penseurs »⁵¹ écrit G. Hottois. Cela ne suffit pas à opérer une révolution intellectuelle, rend le lien avec l'humanisme difficilement lisible quelle qu'en soit la définition et pour les différentes raisons analysées ici. La démonstration d'une convergence des NBIC améliorant les performances humaines (version séculière) soulève dans certaines applications le même scepticisme que la théorie de la noosphère vers le point *Omega* (version chrétienne). Cela est peut-être la conséquence d'un monde accéléré où la technique est confondue avec la Science. Le microscope à effet tunnel est la vitrine de la mécanique quantique, les *big data* celle des Sciences économiques et du calcul algorithmique, l'IA forte hybride

l'est de celle de l'homme de Pékin. Elle accomplira enfin consciemment sa véritable destinée ». Traduction personnelle.

51 HOTTOIS, 2014, emplacement 322.

conscience et machine, etc. Ce phénomène avait déjà été prophétisé dès 1990 par Jacques Ellul lorsqu'il écrivait : « la science est devenue un moyen de la technique »⁵². S'il est impossible d'ignorer cette formidable évolution technologique et encore possible de lui proposer une indispensable réflexion éthique, il paraît très difficile de la rattacher à l'humanisme.

BIBLIOGRAPHIE

ALEXANDRE, Laurent, *La mort de la mort. Comment la techno médecine va bouleverser l'humanité*, Paris, JC Lattès, 2011.

BARROW, John, *The Anthropic Cosmological Principle*, New York Oxford University Press, 1986.

BOSTROM, Nick, « A History of Transhumanist Thought », *Journal of Evolution and Technology*, n°14, 2005, p. 1-25.

BRAINBRIDGE, William, « Science and technology convergence : with emphasis for nanotechnology – inspired convergence », *J Nanopart Res* n°18, 2016, p. 1-19.

BUCHANAN, Allan, *Beyond Humanity? The Ethics of Biomedical Enhancement*, Oxford, Oxford University Press, 2011.

BUCHANAN, Allan, *Better Than Human : The Perils and Promise of Enhancing Ourselves* Oxford, Oxford University Press, 2011.

BURCKHARDT, Jacob, *La civilisation en Italie au temps de la Renaissance*, Paris, Plon, 1906.

CONDORCET, Nicolas de, *Esquisse d'un tableau historique de l'esprit humain*, Paris, Agasse 1795 (an III de la République).

ELLUL, Jacques, *La technique ou l'enjeu du siècle*, Paris, Economica, 1990.

FURETIÈRE, Antoine de, *Dictionnaire de Furetière*, Paris, Dictionnaire Le Robert, 1978.

HOTTOIS, Gilbert, *Le trans-humanisme est-il un humanisme ?*, Bruxelles, L'Académie en poche, 2014, version numérique.

52 ELLUL, 1990, p. 7.

- HOTTOIS, Gilbert, « Pour un transhumanisme philosophique critique », *Généalogies et nature du transhumanisme*, DAMOUR, Franck (dir.), DEPREZ, Stanislas (dir.) DOAT, David (dir.), Montréal Liber, 2018, p.73-87.
- HUXLEY, Aldous, *Brave New World*, London, Chatto & Windus, 1932.
- HUXLEY, Aldous, *Brave New World Revisited*, New York, Harper & Row, 1958.
- HUXLEY, Julian, *New Bottles for New Wine*, London, Chatto & Windus, 1957.
- KURZWEIL, Ray, *The Age of Spiritual Machines : When Computers Exceed Human Intelligence*, New York, Penguin, 1999.
- KURZWEIL, Ray, *The Singularity is Near : When Humans transcend Biology*, Viking / Penguin press 2005.
- KURZWEIL, Ray, *How to Create a Mind : the secret of Human Thought Revealed*, New York, Viking / Penguin press, 2013.
- MOORE, Gordon, Edward, « Cramming More Components onto Integrated Circuits » n°19, *Electronics*, 1965, p. 114-117.
- MORE, Max, « Principes extropiens.3.0 », Traduction française, Transition pour les éditions Hache, mars 2003, <http://editions-hache.com : essais : more : more1.html>.
- MORAVEC, Hans, *Mind Children : The Future of Robot and Human Intelligence*, Cambridge (MA), Harvard University Press, 1988.
- PROCHIANTZ, Alain, *Singe toi-même*, Paris, Odile Jacob, 2019.
- ROCO, Mihail, *Converging Technologies for Improving Human Performance : Nanotechnology, Biotechnology, Information Technology and Cognitive Science*, National Science Foundation. Springer, Dordrecht, 2003, [www.wtec.org/ConvergingTechnologies/1/NBIC report.pdf](http://www.wtec.org/ConvergingTechnologies/1/NBIC_report.pdf).
- SAINT-EXUPÉRY, Antoine de, *Pilote de guerre*, Paris, Gallimard, 1942.

TEILHARD DE CHARDIN, Pierre, *L'avenir de l'homme*, Paris, Seuil, 1959.

TEILHARD DE CHARDIN, Pierre, *Le phénomène humain*, Paris, Seuil, 1955.

VALCKE, Louis, « Pic de La Mirandole », *L'Agora*, vol.1, n°7, p.1-9.

[http://agora.qc.ca/Dossiers/Jean Pic de Mirandole](http://agora.qc.ca/Dossiers/Jean_Pic_de_Mirandole)

VALCKE, Louis, *Pic de la Mirandole, un itinéraire philosophique*, Paris, Les Belles Lettres, 2005

UNE IA CONQUÉRANTE ET UN RÊVE NON ACCOMPLI

par M. Patrick PENEL
de l'Académie du VAR

Les organisateurs de la conférence ont placé sa thématique à la croisée des études et recherches sur le cerveau, sur les neurosciences, sur l'apprentissage, sur les biotechnologies, et sur « l'intelligence artificielle », de sorte que le débat s'oriente « vers quel humanisme ? » : vaste programme et larges débats suscitant beaucoup de passions... Cette contribution sera plus modeste, excluant la domination fantasmée d'une forme d'intelligence mécanique sur l'intelligence humaine, tentant de mettre en perspective les conjonctures historiques et ce que l'on sait aujourd'hui en matière d'Intelligence Artificielle ou plutôt d'Informatique Avancée, terminologie¹ certainement préférable pour sortir du flou entretenu par le vocabulaire. On voudrait en savoir davantage sur les projets chimériques et mieux mesurer quelques-unes des prouesses réelles des logiciels d'intelligence artificielle, souvent ou presque toujours différées à une décennie ou davantage dans le futur². Il serait intéressant d'apporter quelques éclairages sur les concepts-clés et les développements de l'apprentissage calculé, et bien sûr, d'examiner cet autre marqueur des recherches expérimentales prioritaires de notre siècle, le traitement technologique et informatique des masses de

1 Nous proposons de promouvoir cette traduction de « IA » par « Informatique Avancée » comme suggérée par Jean-Louis Dessalles, enseignant-chercheur à Telecom Paris Tech, auteur du livre *Des intelligences très artificielles*, publié chez Odile Jacob (2019).

2 Nous nous souvenons tous de l'annonce faite en 1958 par un spécialiste d'économie et psychologie de la décision, Herbert Simon « un ordinateur champion du monde au jeu d'échecs avant 10 ans ».

données. Car extraire de l'information pertinente au sein de ces *Big³ data* semble bien devenu impossible au regard humain, aussi les défis sont-ils algorithmiques sur fond d'informatique toute puissante.

UN PAYSAGE NUMÉRIQUE DOMINANT

Manipuler la parole et la vision à un niveau abstrait, numérisé, semble ne plus gêner personne. Les assistants robotisés du quotidien et leurs îlots d'informations disponibles ont été très rapidement adoptés. Le numérique est devenu l'affaire de toute la société. « La modélisation mathématique et le calcul scientifique font maintenant partie du paysage général en science et sont universellement utilisés dans la quasi-totalité des secteurs industriels »⁴, comme l'exprime Thierry Colin, mathématicien et chercheur en radiomique.

En ce début de XXI^e siècle, le champ investi par les développeurs d'« IA » est immense, deux faits fondamentaux⁵ sont indissociables de l'évolution des technologies et de notre société toute entière : le premier est la maîtrise et la puissance croissante de l'informatique, inter-connectivité incluse ; le

3. L'aspect 'Big' concernerait surtout le marketing commercial (auquel cas il est impossible de ne pas évoquer les géants d'internet, GAFAMINT et BATX, voire des plates-formes numériques comme celle d'AirBnb) et nos existences totalement paramétrisées. Les projets industriels fondés sur la connaissance issue de données massives sont une autre histoire.

4 COLIN Thierry, de *Sophia Genetics*, « La modélisation des tumeurs, un rôle clé dans les années à venir ? », *Santé et intelligence artificielle*, CNRS Editions (2018).

5 Il faudrait compter avec un troisième fait fondamental, la médecine réparatrice du corps humain de patients handicapés ou amputés. La compréhension du niveau de complémentarité homme-machine est ici essentielle, ses limites parfaitement connues des spécialistes de la science biomédicale : point d'homme augmenté et de promesses délirantes associées.

second est la surabondance et le large partage de nos relations avec l'écrit numérique, avec le son numérique (dont la parole), avec l'image numérique (de source photographique, de signatures radiologiques, de signaux enregistrés, de scènes vidéo-urbaines), avec toutes données numérisées multiples (assurance, banque, énergie, etc.).

Tous les deux génèrent ou peuvent générer des transformations spectaculaires dans la vie sociale, professionnelle, quotidienne de tout un chacun. Tous les deux composent un paysage numérique dominant et plaidant pour une échelle de temps des plus courtes, ils influent donc profondément sur les comportements : le premier par ses critères décisifs pour l'imaginaire de la créativité scientifique et technique, le second par ses critères provisoires, ou encore indéfinis, pour d'indispensables reconfigurations de l'information.

Notre nouveau monde numérique fixe une illusion d'immédiateté dans l'accès à la connaissance. Il ne célèbre plus hélas culture, éducation et science. Et il continue à annoncer d'autres conquêtes, les études en cours pour le contrôle de robots, celles pour la voiture autonome. L'œil scientifique et critique devrait davantage veiller et rassurer. Les défis ne sont pas seulement algorithmiques, ils sont scientifiques, organisationnels et humains.

INTERROGATIONS NÉCESSAIRES

Comment une mécanique algorithmique, forte de structurations de plus en plus sophistiquées, inspirées des cerveaux humains, pourrait-elle donner naissance à une « vie artificielle » pareillement riche de sentiments, de prédictions et de conscience ? Sur tel ou tel sujet ciblé choisi, comment pourrait-elle créer une pensée ou exposer un raisonnement ou bien encore formuler un problème ? Comment pourrait-elle gagner sur les sciences statistiques modernes les plus avancées

ce qu'aucun échantillonnage ne peut réaliser, la capacité d'une perception globale du sujet permettant de comprendre et d'apprendre ? Comment l'informatique pourrait-elle procéder là où le complexe à décrire existe ? La mathématique et la physique sont réputées riches d'exemples où règnent l'imprévisible et le non-linéaire, la biologie et l'économie fournissent également de tels exemples. Observons dans une expérience de pensée un pendule *a priori* libre d'osciller avant de se stabiliser : dans le cas typique, bien connu et fort simple d'un pendule magnétique lorsqu'il est lâché au-dessus d'un plateau portant trois aimants disposés aux trois sommets d'un triangle équilatéral, sa dynamique est toute autre car l'aimant qui influencera sa stabilisation est totalement imprévisible. La trajectoire du pendule est chaotique, contrariée par la complexité⁶ des portions d'espace où deux ou trois des bassins d'attraction des aimants interagissent. Certes elle reste déterministe, mais complètement différente en réponse à des variations excessivement faibles de sa position d'origine au lâcher. Seules la science et l'expertise humaine permettent une exploration logique de tels phénomènes, pas la machine, là où l'incertain et l'imprévisible existent, tout comme l'indécidable (relatif) existe même en mathématique.

UN MYTHE PRESTIGIEUX : LA MACHINE PENSANTE

L'histoire de l'intelligence artificielle est déjà bien ancienne, pas moins de soixante-dix ans. Elle est composée d'épisodes saisonniers, printaniers ou hivernaux selon le soutien des agences de financement et en fonction de l'enthousiasme des chercheurs. Pour l'essentiel on peut dire qu'elle a accompagné la révolution informationnelle. Peu

⁶ La forme des bassins d'attraction est fractale. Pour qui est intéressé, voir Christophe JACQUET, *Le pendule et les trois aimants* <<https://jacquet.xyz/fractales>>

avant, l'information numérisée avait ouvert la voie à l'informatique numérique. Il a fallu ensuite confondre le développement du traitement automatisé de l'information avec le développement des machines de traitement, ordinateurs ou futurs cerveaux électroniques. Ces traitements furent enracinés au cours des années 1940. Le mot informatique, contraction des mots information et automatique, se répandit à la fin des années 1950. Scientifiques et ingénieurs du numérique et de l'ère de la cybernétique⁷ créèrent ce domaine informatique qui étudie, conçoit et réalise des systèmes automatiques capables de traiter de l'information numérisée. À partir de 1956, c'est précisément ce domaine qui fut baptisé du symbole à deux lettres « IA », symbole d'une idée promue comme LA perspective depuis, celle de ladite intelligence artificielle, désormais affichée partout à un rythme obsédant. Que ces systèmes automatiques intégrés à des technologies portables puissent s'être comportés comme des assistants vocaux ou de reconnaissance d'images, avec de belles performances comme l'on sait, est tout simplement remarquable. L'outillage informatique de l'IA a amplement élargi le monde du numérique, bien au-delà du monde du numérique scientifique, bien au-delà également du traitement et de la reconnaissance des textes, pour la traduction notamment. Que ces systèmes puissent simuler une conversation est une toute autre question. Que ces systèmes puissent faire preuve d'intelligence est encore une autre question. Que ces systèmes, demain intégrés à des robots dits intelligents parce que très liés à l'IA puissent disposer d'autonomie d'apprentissage et de décision pour

7 De Platon à Ampère, on a suivi la base étymologique grecque «*κυβερνητική*», l'art de gouverner, avant que N.WIENER ne fasse de la cybernétique une science des échanges d'information et des interactions systémiques. *Cybernetics and Society, The Human Use of Human Beings*, 1950

exécuter des tâches (prévues ou déléguées, ou déviantes en cas d'autonomie totale) est encore une autre question⁸.

« Une planète qui tourne autour du soleil se comporte de façon à décrire une ellipse parfaite. Il ne viendrait à l'idée de personne de penser que la planète est intelligente et qu'elle sait ce qu'est une ellipse » écrit Jean-Louis Dessalles⁹. Pourtant celles et ceux qui alimentent l'idée d'« intelligence artificielle », l'idée mythique prestigieuse de « machine pensante », et puis celle d'une « post-humanité » n'invitent pas au doute. Ces idées ont intégré le débat public avec les honneurs de la presse, à l'affiche de tous les tournants de la modernité et innovations de rupture.

LES « TISSAGES » ALGÉBRIQUES JACQUARDISÉS AU XIX^E SIÈCLE

Tandis qu'astronomes et ingénieurs manipulaient toujours les tables de calcul mathématique, un mathématicien touche à tout, Charles Babbage, issu de la riche bourgeoisie anglaise, avait, dès 1835, conçu sur le papier un prototype de premier calculateur numérique automatique. Pour fonctionner, sa machine analytique serait instruite via des cartes perforées inspirées de celles du Lyonnais. Babbage était précurseur en cela qu'il séparait bien programme de calcul et données, unité de traitement et mémoire. Quelqu'un comprit rapidement l'intérêt de ses projets et le progrès par rapport à la Pascaline du XVII^e siècle¹⁰ : il s'agit d'Augusta Ada, née Ada Byron, puis

8 Encore une autre question en effet, on pense à la voiture autonome, aux robots militaires, aux robots fous des marchés financiers..., problèmes de statut éthique et juridique.

9 *Des intelligences très artificielles*, déjà cité, chez Odile Jacob (2019).

10 Inventée en 1642 par Blaise Pascal, la première machine à calculer mécanique numérique était arithmétique : La pascaline pouvait faire des

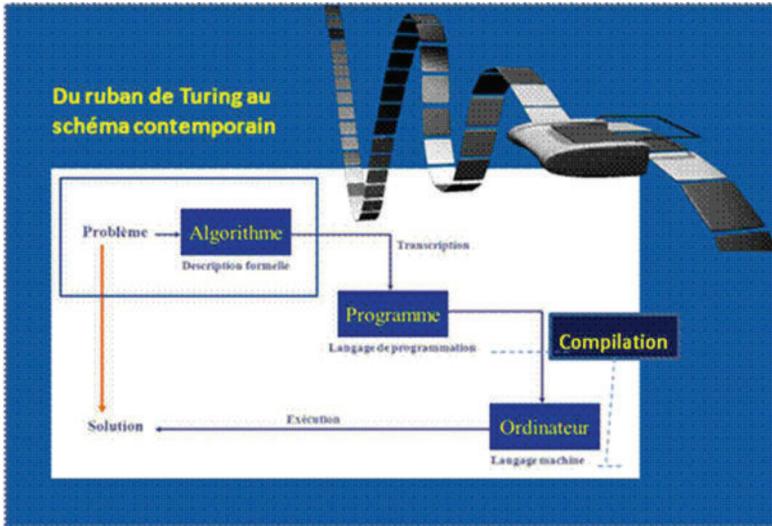
comtesse de Lovelace, fille d'une mathématicienne et d'un poète, « nous pouvons dire que l'*Analytical Engine* de Babbage réalise les tissages algébriques tout comme la machine de Jacquard tisse fleurs et feuillages ». Elle vanta le passage du calcul arithmétique numérique standard à du calcul analytique.

Ada la pionnière édita une méthode innovante en théorie des nombres¹¹ dans un langage destiné à être exécuté par la machine de Babbage, le premier programme informatique de l'histoire était là. Plus important, en réalisant qu'on pouvait utiliser les nombres pour représenter dans la machine des notes de musique ou des lettres, elle ouvrait des perspectives de tâches pluridisciplinaires au-delà des seuls calculs mathématiques. Un siècle avant l'ordinateur.

calculs d'addition (un exemplaire est exposé au Musée des Arts et Métiers, à Paris).

11 Les nombres de Bernoulli sont des nombres rationnels B_k intervenant dans le calcul des sommes de puissances d'ordre quelconque des entiers de 0 à (n-1) : *leur calcul sur la machine analytique de Charles Babbage* par Ada Lovelace a été publié par Louis-Frédéric Ménabréa, Bibliothèque universelle de Genève (1842).

LE PROGRÈS ET LE GÉNIE POUR AVANCER



L'histoire des calculateurs numériques a compté de nombreuses tentatives mécaniques et électromécaniques avant que ne s'impose l'idée d'une machine universelle, peut-être pensante, assurément pour aller au-delà d'une simple fonction de calcul. Deux personnalités scientifiques d'exception, Alan Turing et John von Neumann, firent converger leurs visions d'avenir et contribuèrent aux grands projets des premiers ordinateurs électroniques, l'*Eniac*¹² pour l'armée américaine et l'*Edvac*¹³ d'après-guerre. De la machine de Turing au modèle de von Neumann, la décennie 1936-1946 a été

12 *Eniac*, Electronic numerical integrator and computer, premier calculateur ayant les capacités d'une machine de Turing avec un mode de programmation sur un panneau de connexions, un monstre de 30 tonnes (réalisation entre 1945 et 1955).

13 *Edvac*, Electronic Discrete Variable Automatic Computer, premier ordinateur électronique. Il opérait en mode binaire contrairement à l'*Eniac*, qui opérait en mode décimal.

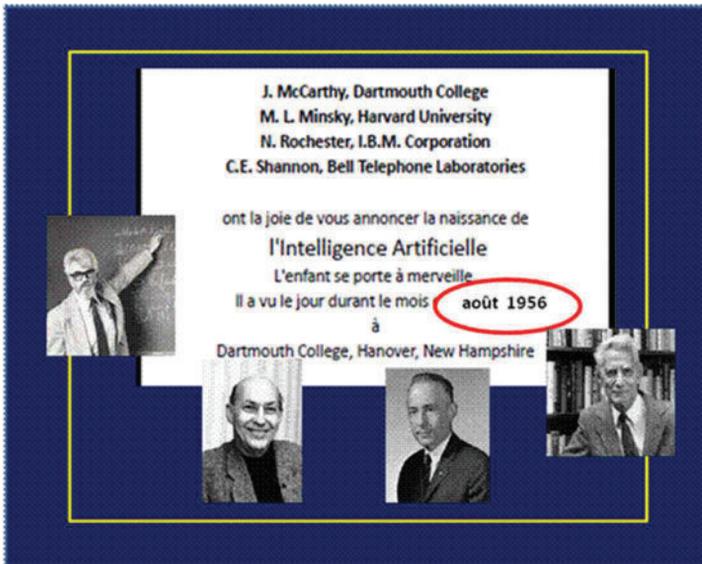
décisive. Le mathématicien et cryptologue britannique Turing avait pensé les bases de l'intelligence mécanique du calcul, imaginant un modèle abstrait de machine, un modèle numérique parfait d'automate programmable : un outil mathématique formel apte au calcul de tout ce qu'un enchaînement d'instructions précises à appliquer serait capable de faire. La séquence écrite d'un nombre fini de règles opératoires à suivre, très exactement l'algorithme propre au calcul, serait traduite dans un langage de programmation pour être communiquée à l'automate. Programme et automate étaient encodés comme des nombres, de sorte que « tout était calcul ». Les idées sous-jacentes de nombre décodé comme une machine, d'une machine encodée comme nombre, étaient révolutionnaires. Complètement nouveau aussi, un ruban de cases mémoires théoriquement infini enregistrerait les données et tous les états successifs du calcul. Mais il n'existe pas d'algorithme général¹⁴ permettant de déterminer si une telle machine va ou non s'arrêter de calculer, indécidabilité établie par Turing.

Le grand mathématicien hongrois von Neumann a progressivement dessiné l'architecture de l'ordinateur : une machine à programme enregistré en mémoire totalement inspirée par la machine de Turing. Le passage de cartes perforées à des impulsions électroniques marquait un progrès essentiel pour accélérer le traitement des instructions et enrichir l'algorithmique. Les questions non résolues alors concernaient la technologie des mémoires pour stocker instructions et données. La notion de programme enregistré à côté des données du calcul et la séparation entre unité de

14 L'indécidabilité est l'impossibilité absolue et définitivement établie de résoudre par un procédé général de calcul un problème donné. Ainsi, depuis Turing, contrôlabilité et finitude se sont imposées dans sa définition d'un algorithme, intégrant désormais séquence finie d'instructions et nombre fini de données.

traitement et unité de mémoire sont restées inchangées depuis von Neumann.

LES PRÉCURSEURS, ENTRE MACHINE PENSANTE RÉVÉE ET RÉVOLUTION INFORMATIONNELLE



L'intelligence fut fêtée au nord et au sud de New-York, au MIT et à Princeton, dans un environnement scientifique exceptionnel où physiciens et mathématiciens de génie se succédèrent nombreux. Logique mathématique, calculabilité, ingénierie informatique et théorie des automates étaient constamment à l'ordre du jour, bien d'autres sujets de recherche scientifique également. Le mathématicien américain Norbert Wiener s'intéressait à la théorie de la communication, au système nerveux et à la philosophie de l'automation¹⁵. Il est

15. Norbert Wiener (1894-1964), philosophe de l'automation, cité ainsi dans l'*Encyclopédie Universalis*, son PhD date de 1912 à Harvard en logique mathématique. Il s'était formé aussi à la philosophie et à la biologie, *Cybernétique et Société* (voir la note 7).

au MIT depuis 1919. À partir de 1933, Albert Einstein, prix Nobel de Physique 1921, s'y installera, tandis que le grand mathématicien Kurt Gödel est accueilli à Princeton. John von Neumann aussi est à Princeton depuis cette année 1933. Alan Turing navigua entre Bletchley Park, au nord-ouest de Londres, Manchester, Cambridge et le MIT pour poursuivre le développement des machines à information. Ils ont tous vécu plusieurs révolutions scientifiques en même temps, de façon emblématique ils voulaient visualiser et connaître le fonctionnement spécifique du cerveau, y suivre le processus qui construit la pensée et conduit à la décision intelligente ou à la critique intelligente. Autour de Wiener, en conjecturant « le cerveau mathématiquement modélisable », on jugeait réalisable ce rêve de projeter l'expertise humaine et les facultés humaines sur les premiers ordinateurs en les reproduisant en machine, voici donc les miracles annoncés de l'intelligence artificielle. L'enfant « IA » naquit dans l'État du New-Hampshire à l'été 1956, sa naissance célébrée au *Dartmouth College* au cours d'un colloque organisé par John McCarthy, Marvin Minsky, Nathaniel Rochester et Claude Shannon.

Les deux aînés sont Rochester et Shannon. Rochester avait 42 ans, depuis huit ans il travaillait chez IBM où il développait le premier langage assembleur symbolique, les instructions en binaire étant remplacées par des codes compréhensifs pour le programmeur. Shannon avait 40 ans, il a travaillé aux *Bell Laboratories* sur la cryptographie et la théorie de l'information numérisée ; au MIT depuis peu il s'intéressait à la communication. Le CNRS présente Claude Shannon comme le père du binaire. Les deux plus jeunes, McCarthy et Minsky, avaient 26 ans, ils s'étaient rencontrés à Princeton alors qu'ils préparaient leur thèse, le thème des machines pensantes les réunissait. McCarthy venait de croiser Shannon aux *Bell Laboratories* sur ce même thème qui était une véritable obsession pour Turing.

Dans *Computing machinery and intelligence* en 1950, Turing introduisit ce qui deviendra le célèbre jeu-test de Turing, un homme en confrontation verbale avec un opérateur, opérateur éventuellement pensant, doit pouvoir déterminer si cet opérateur est un humain ou bien une machine (un ordinateur). Ce test n'a jusqu'ici jamais été passé avec succès pour une machine.

Une autre question importante posée par Turing était comment « lier pensée et langage de façon symbolique » ? Cette abstraction recherchée conduisit à un rapprochement fort pertinent entre linguistique, science et technologie¹⁶, pour tout ramener à des manipulations de symboles. Les progrès en vue de la traduction automatique s'inscrivirent à la suite.

RADICALEMENT NOVATEUR POUR LA COMMUNICATION NUMÉRIQUE

Pour Wiener, le pouvoir de décision de l'homme ne devait pas être bousculé par une machine. Pourtant la production automatisée de l'information allait l'affaiblir, et de même ébranler la vision antérieure dominante de systèmes centralisés. Pour la transmission et la communication de messages et d'informations, en effet, le chemin et la procédure de transfert, sur un réseau plus virtuel que connecté au sens traditionnel des réseaux de télécommunications, n'étaient ni clairs ni établis. Le sujet n'était pas simple, quand bien même les opportunités informatiques paraissaient immenses lorsque le ministère américain de la défense créa le réseau Arpanet. Deux soucis cruciaux étaient à l'ordre du jour : le souci du

16 La guerre avait imposé de savoir décrypter les codes secrets des nazis, puis la guerre froide commanda de savoir traduire automatiquement en anglais les données des textes russes et vice-versa. Voilà ce qui fut à l'origine des premiers développements de l'assistance robotique pendant la guerre et des progrès de l'informatique avancée depuis.

contrôle des flux numériques et le souci de faire communiquer des ordinateurs utilisant des systèmes d'exploitation différents. Le financement de départ était militaire mais les projets étaient orientés vers le domaine public, en dehors des usages marchands et sans objectif premier d'application militaire.

L'invention de la commutation par paquets (1962-63) procéda d'idées totalement nouvelles : segmenter l'information numérisée en paquets¹⁷ de données numériques, considérer leur routage indifféremment désordonné entre plusieurs nœuds de réseau interconnectés, puis dans un second temps, dans un ordre référencé, réassembler ces paquets sur le nœud destinataire, avec un contrôle qui, en cas d'erreur ou de perte, permettait de réémettre l'ensemble. Un réseau d'ordinateurs hétérogène, et le routage réparti par paquets : Cette invention est à mettre au crédit de nombreux personnages, Paul Baran, ingénieur américain d'origine polonaise, Donald Davies, informaticien britannique, en étroite relation avec un doctorant du MIT Leonard Kleinrock, avec une forte participation française aussi. Les travaux internationaux de cette époque créèrent l'informatique distribuée interactive. Le projet Cyclades de l'INRIA¹⁸ permit la mise au point du routage par paquets en France

17 Sur l'invention de la commutation par paquets, on pourra consulter *Reliable Digital Communications Systems Using Unreliable Network Repeater Nodes* et *On Distributed Communications Networks* publiés par Baran en 1962, et la thèse de KLEINROCK, « *Message Delay in Communication Nets with Storage* » (MIT 1963).

18 Lancé en 1966 par Charles de Gaulle dans le cadre du Plan Calcul, et créé officiellement l'année suivante, l'IRIA a réuni mathématiciens et informaticiens français pour la recherche en informatique et automatique avant de devenir en 1980 l'institut national INRIA, toujours pionnier entre monde de la recherche en mathématiques appliquées et monde entrepreneurial.

parallèlement au développement d'Arpanet aux États-Unis. Et bientôt Internet.

Déjà à la fin des années 1980, courriels et données numériques de toutes sortes étaient échangés sur un réseau de plus de 100 000 ordinateurs. Aujourd'hui le chiffrage est devenu astronomique, mais la médiation humaine de communication reste un grand défi d'avenir. Une sous-organisation des unités socio-économiques, culturelles, politiques, médiatiques, montre bien la nécessité de médiation. Une crise informationnelle s'est installée, profonde, révélant une faible productivité de l'information produite rapportée aux gains de progrès technique. D'autant qu'avec l'informatique distribuée, les opportunités technologiques semblent sans limite.

LA SURENCHÈRE TECHNOLOGIQUE, LA TOILE, ET L'« IA » TOUJOURS EN DEVENIR

En 1989 à Genève, le modèle de réseau de communication des scientifiques du Cern et de leurs collaborateurs du monde entier devenait le projet *World Wide Web*, pour passer en quelques années dans le domaine public. L'initiateur du projet est un physicien britannique du Cern, Tim Berners-Lee. Le document « Gestion de l'information : une proposition », dans lequel il décrit cet outil d'échanges scientifiques modélisé comme une « toile » a bien mérité la célébrité historique et l'universalité¹⁹. En 1994, Berners-Lee rejoignait le MIT, et fondait le Consortium international WWW dont le premier siège européen fut français, à l'INRIA. La technique géniale des liens « hypertextes » associée à cette infrastructure de

19 BERNERS-LEE Tim, *Information management : A proposal* (mars 1989) – on pourra consulter « [The original proposal of the WWW, HTMLized](#) » sur *World Wide Web Consortium*. Depuis, le Web a révolutionné les communications dans le monde entier.

réseaux qui prendra le nom d'internet, autorisait désormais tout échange d'informations, le protocole de communication Http et le langage informatique Html étaient maîtrisés. Google, avec un moteur de recherche en ligne comme application web, est arrivé une dizaine d'années plus tard. Le réseau social des étudiants de Harvard, Facebook, apparaissait en 2004, s'étendant à plusieurs universités et plusieurs écoles en quelques mois, plate-forme de réseau social moins de deux ans plus tard.

Pouvoir cumuler les performances des technologies de réseau informatique et celles de traitement automatisé des grands volumes de données échangées²⁰ et rapidement, plus qu'un souhait exigeant, cela entraîne de nouveaux défis : construire des ordinateurs toujours plus puissants, des superordinateurs..., comprendre et valider les algorithmes de l'apprentissage profond automatisé, le *deep learning*, popularisé dès les années 1990 par le français Yann Le Cun, prix Turing 2019 depuis conjointement avec Yoshua Bengio et Geoffrey Hinton pour leurs travaux.

Il y a quelques mois, Mark Zuckerberg annonçait le nouvel objectif de Facebook, la création d'un véritable monde virtuel, comme un univers parallèle au nôtre, le métavers. Et l'entreprise changeait de nom à cette occasion, devenant Meta : « Dans le métavers, vous pourrez faire presque tout ce que vous êtes en mesure d'imaginer, vous rassembler avec vos amis et votre famille, travailler, apprendre, jouer, faire du shopping, créer ». Ce deuxième monde, de jeux et de commerces en réalité virtuelle, serait l'environnement idéalisé de la sociabilisation sur Internet, deviendrait un futur 3D-Web

20 Viser le traitement jusqu'à l'exa-byte de données, l'échelle correspondante est à « 1 024 puissance 6 » bits, voir ci-après le paragraphe « chiffrer le stockage des données ».

avec ventes de services et biens numériques, avec des « copies » tridimensionnelles des personnes et des objets.

Yann Le Cun est le responsable de la recherche en « IA » de Facebook depuis 2013, sa vision futuriste s'est orientée vers une nouvelle algorithmique, plus autonome à partir de vastes ensembles de données variées, quelle que soit leur forme, et plus polyvalente, capable de générer du visuel, de l'oral ou du textuel, de l'interactif simulé et en outre la traduction en temps réel. Cette algorithmique intégrerait la plupart des techniques actuellement maîtrisées de l'informatique avancée.

Une question demeure : que renferment-ils, tous ces algorithmes de l'art de penser ?

Pour garantir une qualité d'immersion dans l'artificiel du métavers, Meta met en œuvre *AI Research SuperCluster*, la construction d'un superordinateur parmi les cinq ou six superordinateurs actuels les plus puissants au monde.

DEUX ÉPISODES PRINTANIERS DE L' « IA », ÉCHECS ET GO

Une intelligence dite artificielle supplantant l'intelligence humaine ? C'est en matière de jeux que l'illusion aura opéré sur le grand public, à deux reprises en 1997 et 2016. La première étape pour ne plus balbutier en « IA » datait de 1958, la plus marquante des promesses alors – avant dix ans un algorithme équipant un ordinateur devant battre le champion du monde du jeu d'échecs – ce fut *DeepBlue* d'IBM contre Kasparov. 39 ans s'écoulèrent donc avant que se réalise la prédiction de l'économiste américain Herbert Simon, prix Turing 1975. Sans diminuer la performance informatique d'IBM, l'algorithme produit explorait un espace de possibilités de jeu fondamentalement fini, suivant des règles en nombre fini, et en temps fini il avait pu basculer du recensement de tous les coups possibles (*Big Data* appropriées) sur un petit

nombre de coups pertinents pour finaliser une partie gagnante. Les promesses non tenues jusqu'alors de la gestion des mégadonnées et de l'intelligence artificielle étaient temporairement « effacées ». La deuxième promesse suivit, le défi d'une nouvelle prouesse dans les jeux, au go cette fois, lancé sans promettre un délai. Avec l'apprentissage profond numérique et la magie des réseaux de neurones, le succès s'est appelé AlphaGo contre le champion du monde coréen Lee Sedol, 19 ans après *DeepBlue* d'IBM. Toutefois, le glorieux algorithme de 2016 de Google *DeepMind* ignorait les règles non écrites du plaisir de jouer !

AlphaGoZéro et AlphaZéro lui succèdent avec mille fois moins d'options de jeu explorées, ce dernier logiciel aurait inventé sa propre façon de jouer à partir des règles de jeu et bien sûr des données enregistrées de très nombreuses parties de go jouées pour l'entraînement. Continuer à parler d'« Informatique Avancée » paraît fort pertinent, d'autant plus que la suite de cette histoire est maintenant faite de championnats d'ordinateurs « machine contre machine », des combats entre artificiels qui peuvent être de conception intelligente sans être intelligents. On peut conjecturer qu'à force les ordinateurs joueront la même partie répétée.

Avec ces expériences dans le monde des jeux, échecs, go, etc., la preuve est indirectement donnée que la clé n'est pas le développement cognitif. Le mot réflexe a remplacé le mot intelligence. La traduction de « IA » par intelligence artificielle est ici invalidée. Le dualisme dominé-dominant populaire est devenu un faux débat puisqu'il n'y a plus d'engagement de champion humain dans ces championnats d'ordinateurs.

DES OCÉANS DE DONNÉES NUMÉRIQUES

Kasparov a été battu par une machine dont la puissance permettait de gérer des mégadonnées. Sedol a été battu par un

superordinateur spécialement programmé pour avoir appris des millions de coups successifs à partir de centaines de milliers de parties jouées. Optimiser à travers la forêt d'arbres décisionnels, certes plus dense pour le jeu de go que pour le jeu d'échecs, était supervisé et inclus dans les algorithmes produits. On ne peut rester sur le thème fantaisiste d'une « intelligence algorithmiquement produite ». Pour réaliser des tâches basées sur des règles « métier », comme sur des règles « jeu » pour gagner, traduire en machine le séquençage des chemins de réussite n'est pas scientifiquement révolutionnaire. Les enjeux sont clairement la puissance de calcul et les *Big data*, et les technologies pour leur traitement. Hors des domaines de règles spécifiques, extraire des caractéristiques et de l'information au sein de méga-données est autrement plus difficile : pas seulement leur exploration, mais aussi leur capture et surtout leur stockage.

Le monde informatique ne cesse d'immerger dans les profondeurs d'« océans » de données numériques, des faits présents ou passés, issus du domaine industriel ou concernant nos personnes, pour dans un second temps les traiter quasiment comme des matières premières ou comme des armes afin d'asseoir des pouvoirs : leur collecte non régulée et leur exploitation peu rationalisée, peu socialisée, peuvent à juste titre nous préoccuper ; les observations extraites par inférence automatisée ont toutes les chances de reproduire des schémas habituels de représentation, ainsi que les erreurs et les biais existant dans les données.

L'essor des volumes de données est exponentiel, au quotidien comme au niveau industriel. Des hordes d'assistants informatiques ou robotiques « travaillent » au quotidien, sites internet, smartphones, cartes de fidélités, téléphones portables, objets connectés, réseaux sociaux, caméras de surveillance, etc. Au niveau industriel, on ne compte même plus le nombre

d'îlots d'informations enregistrées disponibles, un millier de giga-octet de données enregistrées par heure de vol à bord d'un avion Airbus, et combien de données dans les secteurs de la banque, de l'assurance, de l'énergie. On dit que 90 % des données accumulées aujourd'hui disponibles ont été créées ces trois dernières années.

Dans le monde numérique des scientifiques, la production, la gestion et l'analyse des données sont sous leur contrôle amont et aval, les algorithmes sont dictés par les travaux théoriques, souvent inspirés par les mathématiques, et les résultats reproductibles²¹ validés au fur et à mesure. Sur eux dans ces conditions, la réflexion et la pensée experte sont en passe de devenir plus élaborées. Nous rappellerons la « découverte » du boson de Higgs, combien de milliards de giga-octets de données issues du Large Hadron Collider de Genève pour cette réussite, une réussite venue confirmer une théorie vieille de 50 ans, permettant alors de préciser le modèle standard de la physique et la théorie de l'unification des forces de l'univers. Ce sont les résultats de simulation ou d'expériences ou d'observations qui donnent lieu à ces données enregistrées, contrôlées par la méthodologie de la science et les théories, ces données sont réellement des matières premières pour les physiciens.

POUR CHIFFRER LE STOCKAGE DES BIG DATA

Le chiffrage du stockage des données numériques peut donner le tournis, il nous faut néanmoins en prendre métriquement le sens, du bit de référence jusqu'au yotta-bit à l'échelle de millions de milliards de milliards de bits : d'un bit, unité pour positionner un caractère '0' ou '1', on passe à 8 bits

21 La certification de la reproductibilité de la recherche scientifique disposant de dispositifs expérimentaux est un sujet de grande importance.

ou 1 octet ou 1 byte (8 bits-positions, 2 états, 1 octet permet donc de coder $2^8 = 256$ combinaisons) (un octet en imagerie code en effet 256 couleurs et un pixel se décrit sur trois ou six octets pour augmenter la résolution colorimétrique). Un facteur mille conduit au kilo-octet ou kilo-byte (1 KB pour 1 024 bits précisément), nous multiplierons par 1 024 à chaque fois. Un facteur d'un million conduit au méga-octet ou méga-byte (ainsi 2MB correspondent à une photo numérisée de bonne définition, précisément sur 1 048 576 bits). On obtient ensuite le giga-octet ou giga-byte (2 GB correspondent à une étagère de 20 mètres de livres numérisés), puis le téra-octet ou téra-byte (2 TB correspondent à une bibliothèque de recherche académique numérisée), puis le péta-byte, l' exa-byte, le zetta-byte (ainsi 2 ZB correspondaient au flux annuel de données sur Internet en 2016-2017), pour finir le yotta-byte (1 YB correspond à « 1 024 puissance 8 » bits, c'était la toute nouvelle unité adoptée en 1991 lors la 19^e Conférence internationale générale des poids et mesures. Au-delà des matériels informatiques²² où ces échelles traduisent bien les gains de puissance, nous souhaitons les comparer aux normes astronomiques : l'unité donnant la distance moyenne de la Terre au Soleil, 1 UA, vaut 149 597 870 700 mètres, de l'ordre de 0,15 péta-m. Une année-lumière, donnant la distance parcourue dans le cosmos par la lumière en une année, c'est 9,46 exa-m. Le disque galactique de notre Voie lactée a un rayon estimé de 520 000 années-lumière, soit environ 49 yotta-m. Autrement dit, les préfixes adoptés ont une certaine universalité et nous familiarisent avec des ordres de mesure habituels en cosmologie, quelle odysée !

22 *Cloud*, disque dur, clé USB, carte SD..., les défis techniques sont de taille, et bien sûr le temps est à la miniaturisation. Le premier disque dur des années 1950, « IBM Ramac 305 », n'était pas modeste en poids, pesant une tonne, mais modeste en stockage de données, 5 MB.

LA MALÉDICTION DE LA GRANDE DIMENSION

La dimensionnalité estime le nombre de paramètres ou variables qui servent à définir une position, un lieu, en pratique un état, un évènement, un texte, une image, en général une donnée. Ainsi la représentation dans le cosmos est tridimensionnelle comme dans l'espace physique, les évènements dans l'espace-temps de notre univers sont référencés de dimension quatre, tandis qu'en général la dimension de l'espace des données est beaucoup plus grande. Deux données de même type sont tout aussi aisément localisées dans un espace de grande dimension égale à leur nombre de variables caractéristiques. Mais leur distance pourrait y perdre tout sens s'il s'agit de textes ou d'images, le niveau de précision appelant naturellement et toujours augmentation de la dimension. La recherche mathématique, dans des espaces de grande dimension, conduit en effet à savoir les données dispersées, rares et très isolées, ce qui ne facilite pas leur analyse. Le volume des hyper-sphères de localisation tend vers zéro avec la dimension d .

Précisément il est calculé par :

$$\sqrt{\pi}^d R^d / G(\frac{d}{2}+1) \text{ avec } G(2) = 1, G(\frac{3}{2}) = \sqrt{\pi}/2, G(\frac{5}{2}) = 3\sqrt{\pi}/4 \dots^{23}$$

Et les algorithmes de classification tolèrent mal l'augmentation de la dimensionnalité parce qu'elle entraîne une diminution de leurs performances. Sans augmenter le nombre d'échantillons des données d'apprentissage, l'erreur de généralisation resterait élevée. Les fondamentaux de classification statistique sont à repenser, chaque type de données pouvant en outre nécessiter des algorithmes

23 Le volume de l'hyper-sphère unité est maximal pour $d=5$, à 8; $2/15$. Il tend vers 0 quand d tend vers l'infini tandis que le volume de l'hypercube tend vers l'infini avec d comme 2^d .

spécifiques pour leur analyse. Aux questions d'optimisation dynamique les réponses obtenues sont faibles : pour cette raison, dès 1961, le mathématicien américain Richard Bellman avait introduit²⁴ le concept de fléau de la dimension. En 2018, Stéphane Mallat²⁵ s'en est emparé afin d'affirmer pour les mathématiciens la nécessité de mieux comprendre la complexité de l'apprentissage algorithmique liée au nombre de variables caractéristiques.

Une expérience de pensée géométrico-statistique permet de faire le test de cette complexité. Disposant d'une centaine de données numérisées vues comme des points, successivement dans l'intervalle unidimensionnel des nombres compris entre 0 et 1, dans le carré unitaire bidimensionnel, dans le cube unitaire tridimensionnel, jusque dans un hyper-cube unitaire de dimension d quelconque grande, comment inférer une information de qualité sur la distribution ou la répartition des cent points donnés ? Il faudrait au moins 10^{2d} variables d'observations, si $d=10$, $10^{20} = 100 \times 1\,000^6$, autrement dit il faudrait être à l'échelle des préfixes entre zeta – et yotta –. Quand il s'agit d'une centaine d'images avec un million de pixels pour argumenter une information de reconnaissance ou de classification, nous comprenons immédiatement qu'il faille modéliser autrement. L'urgence est vraiment à sous-dimensionner, c'est-à-dire à représenter ces mêmes points-données « projetés » dans un sous-espace de dimension la plus petite possible en recherchant d'autres paramètres tout aussi caractéristiques et plus pertinents pour décrire ces points-données. En sciences bio-médicales cela se conçoit bien. Cédric Villani signale que des méthodes de régularisation

24 BELLMAN Richard, *Adaptative control process : a guided tour*, Princeton University Press, 1961.

25 Stéphane Mallat est professeur au Collège de France, depuis 2017, où il occupe la chaire Sciences des données.

dérivées des problèmes de transport optimal « ouvrent des perspectives pour la réduction de dimensions, mais elles requièrent des puissances de calcul très importantes »²⁶.

Les expérimentations numériques et la compréhension mathématique du traitement de masses de données sont des sujets majeurs et très pluridisciplinaires. Au Collège de France, Stéphane Mallat insiste beaucoup sur la méthodologie RAP (R pour Régularités existantes, connues *a priori*, ou bien recherchées, A pour Approximation en basse dimension, P pour Parcimonie et construction de représentations faisant preuve de parcimonie au sens économe). Un autre fait notable est qu'un choix judicieux de dimensions réduites à utiliser n'est pas la seule prérogative des *data scientists*, elles peuvent être éclairées par la connaissance et l'expertise et s'inscrivent dans toutes les collaborations pluridisciplinaires souhaitables. Amont et aval.

L'APPRENTISSAGE NON SUPERVISÉ S'ÉLOIGNE

Le dépistage du cancer du sein repose sur la mammographie associée à l'examen clinique. Après plusieurs décennies d'aide informatisée au diagnostic, on cherche le remède attendu aux erreurs de détections lésionnelles et aux erreurs d'interprétation. Car la projection bidimensionnelle du volume du sein radiographié sous compression crée des superpositions. Nombre de masques dus à ces chevauchements de tissu fibro-glandulaire défavorisent ou limitent la prédiction de cancers possibles, même en double lecture. Alors comment les spécialistes et les radiologues, deuxièmes lecteurs potentiels, en étroite collaboration avec les informaticiens des

26 VILLANI Cédric, « Optimal transport : old and new », *Springer Science*, (2008).

données, peuvent-ils augmenter grandement la qualité des bases de données ?

Le concours de l'imagerie par tomosynthèse et des technologies du *deep learning* sur une base de données savamment préparées pourrait bien être porteur de succès comme aide à l'interprétation. Un premier niveau algorithmique mathématique reconstitue l'image en trois dimensions à partir des images mammaires successives acquises sur un angle de 15 à 60 degrés. Sensibilité et précision des images tridimensionnelles sont significativement accrues, combinées avec, et là est probablement une clé du progrès, un deuxième niveau de supervision d'expertise médicale pour effectuer un travail très avancé d'annotations des bases d'images bidimensionnelles. À priori sur toutes les images-données d'apprentissage sont superposées des marques géométriques les segmentant, zones critiques identifiées, structures anatomiques et lésions bien localisées, pour déboucher sur le troisième niveau aidé par les logiciels d'apprentissage et techniques de reconnaissance.

Deux progrès majeurs que nous entendons souligner : la collaboration entre scientifiques des données, experts médicaux et radiologues, si possible des radiologues différents pour éviter les biais dus à chacun, et cette modélisation géométrique segmentée parfaitement affichée. Naturellement on s'éloigne quelque peu de l'apprentissage non-supervisé mais on se rapproche déjà fortement de l'humain, du bien individuel immédiatement concerné et de l'intérêt collectif de toute la société. En outre nous voudrions croire à cette preuve de modernité qu'il suffisait de faire appel à l'intelligence humaine pour mieux préparer les masses de données disponibles avant exploration.

LE FILTRAGE POUR SOUS-DIMENSIONNER À L'INTÉRIEUR D'UN MODÈLE DE CALCUL NEURONAL

Nous utiliserons le cas d'une image de faible définition (seulement $499 \times 499 = 249\,001$ pixels, pour comparaison les photographes parlent en méga-pixels, 10 Mpx au moins) : sans filtrage, qualifier tous les éléments de reconnaissance de l'image sur une architecture de 499 neurones demanderait le calcul déjà exorbitant de $499 \times 249\,001$ coefficients synaptiques, ces poids associés aux connexions pour représenter la force des neurones²⁷.

Considérons une fenêtre de filtrage, 3×3 pixels, parcourant l'image avec un pas de 2, ne retenant que (choix pertinent adopté) les pixels de plus grande valeur selon trois composantes couleurs codées de 0 à 255 (dans l'espace colorimétrique tridimensionnel (rouge, vert, bleu)). Nous supposons le procédé répété quatre fois pour atteindre par exemple des images de 28×28 « pixels élus », tout en préservant la disposition spatiale de ces pixels élus unicolorés. La dimension du tableau numérique calculé des caractéristiques de l'image est profondément réduite, l'image initiale n'est plus qu'une image abstraite sous forme de 3 matrices 28×28 de données pertinentes relevées. En reconnectant ces données sur l'architecture des 499 neurones, il reste à classifier l'image, c'est-à-dire en calculant les fameux

27 Au soma d'un neurone biologique correspond un neurone artificiel, aux dendrites correspondent les entrées de connexion, à l'axon correspond la sortie, aux synapses correspondent les poids. Chaque neurone reçoit les valeurs entrées de connexion pour émettre une information de reconnaissance, que donne la somme pondérée des valeurs entrées comparée à une valeur seuil.

poids et en les faisant converger vers des valeurs assurant une reconnaissance de l'image apprise²⁸.

Le procédé de filtrage qui vient d'être décrit donne une assez bonne idée du principe de convolution des systèmes de neurones, pour lequel l'organisation du cortex visuel en aires successives a été fortement inspirant, comme l'écrit Yann Le Cun. Les réseaux multicouches convolutifs ont révolutionné la vision par ordinateur, à partir des années 1990. Utilisant le même principe, voici un autre procédé de filtrage avec une même fenêtre, circulant sur l'image entière, afin de détecter des champs particuliers de l'image, des formes, des contours, des motifs, etc. Ainsi un neurone travaillant sur une fenêtre 3 x 3 pixels n'aurait que 9 poids à calculer pour devenir neurone-détecteur. On peut imaginer une orientation possible vers l'auto-apprentissage en s'appuyant sur un ensemble de neurones-détecteurs.

LA REPRÉSENTATION NUMÉRIQUE ET L'APPRENTISSAGE AUTOMATIQUE TROIS EXEMPLES

Prendre numériquement le relais sur l'expertise humaine est loin d'être une évidence. Trois exemples très concrets, très simples, l'illustrent, celui du modèle du patient numérique, celui du modèle du ballon de rugby numérique, et celui du modèle génétique de girafe numérique.

Dans le premier cas, toutes les données acquises sont paramétrées et prises en compte, anatomiques, physiologiques, biologiques, moléculaires, génétiques, comportementales, plus les données d'imagerie, ainsi un double numérique du patient

28 Reconnaissance avec un score proche de 1, obtenue en minimisant le coût d'apprentissage représenté par l'erreur quadratique moyennée sur les poids.

est en cours de création pour servir dans toutes les disciplines médicales : pré-diagnostic et aide au diagnostic, pré-pronostic et aide au pronostic, est-ce à dire pour aboutir demain au remplacement du médecin ou du spécialiste ? Non pas, seulement pour utilement les aider. Dans la vraie vie, face au patient, l'essentiel est assuré par leurs qualités de conscience professionnelle et d'intelligence humaine réunies, y compris compassion et humilité, et cette capacité de comprendre parmi des milliers de paramètres ceux qu'il est important d'interpréter pour décider une thérapeutique personnalisée et bien ciblée. L'aide servie par le double numérique du patient sera renforcée par toutes les annotations avancées étiquetant les données, comme l'a montré l'exemple des travaux pluridisciplinaires autour des radiologues sur la mammographie pour atteindre 85% de réussite. Sur les données-images, la modélisation géométrique (points, zones, contours marqués, etc.) et la segmentation guidée par le spécialiste intégreront demain les nouveaux algorithmes.

Dans le second cas, un fait divers lu dans *L'Essentiel*, ce journal quotidien luxembourgeois, relatait qu'un club de rugby, écossais de mémoire, avait choisi pour un match l'enregistrement assisté par un logiciel d'intelligence artificielle : le résultat fut extraordinaire, quand, au cours d'une phase de jeu proche de la ligne délimitant l'aire de jeu, le traitement numérique du portrait d'un arbitre de touche, chauve et au visage de forme oblongue caractérisée, fut identifié comme le ballon ! On ne vit plus que l'enregistrement des mouvements de l'arbitre ! Dans la réalité sportive assistée, même empreinte d'humour, il est curieux que les données enregistrées de l'évolution du ballon de rugby n'aient en rien aidé à prédire sa trajectoire future, pas plus que l'innovation de jeu de tel ou tel joueur ! Se pose la question de faire confiance à un algorithme qui ne sait pas expliquer ses décisions. Sur un autre sujet de réalité assistée, quelle place les

logiciels de la voiture autonome laisseront-elles à l'intervention humaine, outre les questions éthiques ou légales ? Quelle place pour anticiper les situations extrêmes et les problèmes soudains sur la route, pour renforcer le choix et le suivi par l'homme marginalisé pourtant à bord ? L'hypothèse d'une absence complète de risques d'accidents ne tient pas en termes de responsabilité morale.

Dans le troisième cas, ces reines géantes et majestueuses que sont les girafes dans les savanes constituent une population en danger, 35 à 40 % d'entre elles auraient disparu au cours des trente dernières années²⁹. « Nous avons étudié les liens génétiques de tous les sous-groupes de girafes sur l'ensemble du continent africain et nous avons découvert qu'il existait non pas un ou deux mais au moins quatre groupes très distincts génétiquement qui, apparemment, ne se reproduisent pas entre eux » explique en 2016 le professeur Axel Janke, de l'Université Goethe de Francfort. Récemment³⁰, les logiciels d'apprentissage profond sur réseaux de neurones convolutifs et de reconnaissance automatisée ont été utilisés avec une base de 4 000 images photographiques de girafes : les résultats semblent confirmer l'identité des quatre espèces de girafes à partir des pelages, grâce aux motifs particuliers de ceux-ci et en dépit de leur apparence similaire. Sans marquage complexe ou artificiel, la performance atteint 90 %. Pour les études démographiques et les suivis écologiques, c'est évidemment

29 FENNESSY Julian, JANKE Axel, *et al.* « Multi-locus analyses reveal four giraffe species instead of one », *Current Biology*, vol 26-18, p. 2543-2549 (2016).

30 MIELE Vincent, BONENFANT Christophe, *et al.* « Revisiting animal photo-identification using deep metric learning and network analysis », *Methods in Ecology and Evolution* (2022, à paraître).

fort intéressant, il serait possible de viser d'autres populations animales à la peau idiosyncrasique avec les mêmes techniques.

LA POÉTIQUE DE LA CRÉATIVITÉ HUMAINE

Ce sous-titre est emprunté à Étienne Klein. Comme lui, on peut être séduit par le détour par l'abstraction pour avoir accès à la réalité. Voici ce qu'il disait de l'univers de Galilée et du langage de la nature : « Galilée parie que cet univers nous deviendra intelligible si nous apprenons son langage. Et ce langage de la nature, ce ne sont ni les livres déjà écrits qui le prescrivent, ni les théories enseignées, ni l'observation du monde aussi attentive soit-elle ; ce sont les mathématiques... »

Qu'il nous soit permis ci-après de paraphraser Klein à propos de l'univers informatique et de son langage du numérique : ce monde informatique du numérique nous deviendra intelligible si nous apprenons et maîtrisons son langage. Et ce langage, ce ne sont ni les logiciels déjà édités, ni les théories enseignées sur les machines pensantes, ni l'observation esclave des connexions et interconnexions..., le langage du monde numérique ce sont aussi les mathématiques qui l'ont algorithmiquement inspiré, et l'intelligence humaine et le fruit des collaborations des hommes permettant d'exprimer le petit nombre de variables pertinentes grâce auxquelles on peut formuler des lois, à valider ou non par l'expérience, tout en obtenant des résultats inédits. Entre 1936 et 1956, les pionniers avaient tous une formation mathématique ou ils étaient mathématiciens. Les mathématiciens d'aujourd'hui se penchent sur l'apprentissage profond non supervisé et cherchent à comprendre pourquoi ça marche quand il y a succès.

SERVIR L'HUMAIN : INTÉRÊT COLLECTIF ET BIEN INDIVIDUEL

Et s'il n'y avait pas plus de compétition entre l'homme et la machine qu'entre un musicien et son bel instrument..., c'est une parabole que j'aime bien, elle n'est pas biblique, mais elle ouvre une sérieuse piste de réflexion : recentrer sur l'homme ayant appris à jouer de son bel instrument, limite déjà l'aspect compétition, penser toutes les capacités naturelles et musicales de cet homme afin de définir ce qui peut servir son bien individuel et en même temps l'intérêt collectif des humains voulant l'écouter, en l'occurrence bien individuel jouissif à produire de la musique, et intérêt collectif à entendre la musique produite.

Le bel instrument n'avait d'autre vocation que d'interagir avec des utilisateurs musiciens, d'où l'importance de l'interface homme-instrument, des évidences sur l'exemple parabolique, de même les systèmes technologiques du numérique ont tous vocation à interagir avec les utilisateurs et les interfaces homme-machine seront fondamentales davantage encore à l'avenir. Partir de l'homme et non plus de la machine, non plus de toutes les technologies de l'Informatique Avancée et du monde numérique, tenir notre regard centré sur l'homme pour définir quel intérêt collectif à utiliser ces produits et quel bien individuel à les manipuler. Une telle démarche inverse la problématique actuelle.

« Servir l'humain : intérêt collectif et bien individuel » s'avère la bonne orientation d'avenir pour tous les produits d'Informatique Avancée, les nouveaux métiers auront en responsabilité les travaux indispensables sur la préparation et la gestion des données, sur la révision des règles et l'adaptation des législations, sur les interfaces homme/machine, sur les certifications des algorithmes et des logiciels, sur plus de scientificité, sur toutes les formes de

médiation, sur l'amélioration de la cybersécurité, sur les exigences de respect de la vie privée, et que la liste s'allonge, on ne peut formuler d'autre souhait pour une dynamique économique, car la suite de l'histoire des *Big data* n'est peut-être pas encore écrite. Plus de questions que de réponses, un formidable appel à l'intelligence humaine est une évidence, l'appropriation des progrès réalisés et des « outils » matures disponibles devrait être une seconde évidence dans l'assistance robotique. La soi-disant pensée algorithmiquement produite à partir des *Big data* demeurera certainement un vrai sujet de débat.

ÉPILOGUE

Nous avons parcouru un foisonnement d'innovations techniques orientées systèmes experts et gestion automatisée, et reconnu une créativité humaine phénoménale, tout simplement explosive : l'informatique distribuée interactive et le web ont signé la réussite de la révolution informationnelle, les superordinateurs signant la puissance triomphante du calcul, de la calculabilité dans à peu près tous les domaines, linguistique, imagerie, etc.

« L'humanité a toujours progressé via des innovations, certaines entraînant des transitions profondes irréversibles : l'écriture, l'imprimerie, la machine à vapeur, aujourd'hui le numérique et sa convergence avec la robotique et l'intelligence artificielle », écrit notre confrère Alain Boudet de l'Académie de Toulouse. Je ne changerais que les deux derniers mots, auxquels je préfère informatique avancée à la place d'intelligence artificielle.

Mais, en accord avec Jean-Gabriel Ganascia, « ni l'accroissement de la puissance de calcul, ni la capacité de stockage ne produisent automatiquement de l'intelligence »³¹.



31 GANASCIA Jean-Gabriel, *Le mythe de la singularité : Faut-il craindre l'intelligence artificielle ?*, Éditions du Seuil, 2017.

CYCERON ENTRE NEUROSCIENCES ET INTELLIGENCE

Cette interview croisée est une discussion entre le fondateur et premier directeur du Centre Cyceron, Jean-Michel Derlon, et un autre ancien directeur de Cyceron, Francis Eustache, qui a également dirigé une unité de recherche du site pendant vingt ans. L'accent est porté sur les origines du projet, sa mise en place, les outils d'exploration du cerveau, les grandes réalisations scientifiques et le concept pluridisciplinaire de Cyceron visant à faire travailler dans un même lieu cliniciens et fondamentalistes.

Francis Eustache :

Cette interview sera présentée dans le cadre d'une rencontre des Académies ayant pour thème « neurosciences et intelligence artificielle ». Quand j'ai eu l'opportunité d'en discuter avec quelques Académiciens, j'ai tout de suite pensé à Cyceron car, à mes yeux, c'est le lieu qui représente le mieux, à Caen, cette connexion entre les neurosciences et l'intelligence artificielle. J'ai envie de dire l'intelligence tout court parce que c'est un endroit où, depuis sa création, des personnes ont eu le loisir d'exprimer leur intelligence et leur créativité pour les mettre au service de la science et du progrès. Revenons aux sources. Nous allons d'abord parler de Cyceron, de ses prémices, de ses fondations et ensuite de son évolution depuis plus d'une quarantaine d'années. Jean-Michel, merci d'être avec nous. Je te demande dans un premier temps de te présenter, puis de nous parler des premiers moments de Cyceron, et même de l'avant.

Jean-Michel Derlon :

Le fond de mon métier, c'est la neurochirurgie. J'ai effectué mes études à Paris où j'ai été interne et débuté mon clinicat à l'hôpital de la Salpêtrière. Je suis venu à Caen en 1977 où j'ai terminé ma formation et ma qualification au CHU, et je suis devenu Professeur en 1985. J'ai ensuite dirigé le service de neurochirurgie à la suite de Jean-Pierre Houtteville, entre 2002 et 2012. Il se trouve que, très tôt durant mes études, je m'étais intéressé aux nouvelles techniques d'investigation du système nerveux, tout particulièrement dans le domaine de la pathologie vasculaire cérébrale, qui était à la fin des années 1970 au cœur de leur développement. Les Anglo-Saxons et les Scandinaves étaient très en avance sur les Français dans l'utilisation des méthodes radio-isotopiques de mesure du débit sanguin cérébral. Le développement de ces techniques était passionnant mais il n'y avait pratiquement pas un endroit en France où elles étaient utilisées. Puis, à la fin des années 1970, commence à se développer l'imagerie par tomographie à émission de positons. Cette méthode comporte des contraintes techniques lourdes, puisque les radio-isotopes, émetteurs de positons, doivent être produits par un cyclotron et ensuite intégrés à une molécule au sein d'un laboratoire de radiochimie, avant de pouvoir être injectés à l'homme et détectés sous une caméra spéciale, différente des caméras à scintillation qui étaient disponibles à cette époque : donc, toute une chaîne technologique dont la plus lourde, dans les années 1975 à 1980, était le cyclotron. La tomographie à positons paraissait extraordinairement prometteuse, permettant non seulement de mesurer de façon très précise le débit sanguin cérébral, mais aussi la consommation d'oxygène ou de glucose par les cellules cérébrales, ou le volume du sang dans les vaisseaux. En fait, une foule de paramètres qui, jusqu'alors,

n'était accessible que chez l'animal d'expérience, avec des technologies très invasives. Là commençaient à émerger les perspectives de mesurer chez l'homme, dans toutes les régions du cerveau, des paramètres qui étaient inexploables auparavant.

Je suis arrivé à Caen au moment où se terminait la construction du Grand accélérateur national d'ions lourds, le Ganil, qui est un outil destiné à la physique corpusculaire, et qui n'avait pas de mission biologique. Les seuls centres de tomographie à positons qui existaient dans le monde étaient tous installés à côté d'un gros cyclotron, partagé avec les physiciens. C'était le cas à Jülich, en Allemagne, à Brookhaven aux États-Unis, mais aussi à Philadelphie, Washington DC, Boston. Les laboratoires de recherche médicale étaient construits au sein d'un campus de physique, pour bénéficier d'une ligne de production de radioéléments émetteurs de positons, pouvant être intégrés à une molécule injectable ou éventuellement inhalable pour les investigations humaines.

Constatant que le campus de physique du Ganil se trouvait à un kilomètre de l'hôpital universitaire, j'ai pensé qu'on avait peut-être là une opportunité extraordinaire pour installer un centre de recherche utilisant la tomographie à émission de positons. Étant encore chef de clinique, en 1979, j'ai remis une note en ce sens au préfet de région, puis au président du Conseil régional qui était Michel d'Ornano. Très vite, l'affaire a été prise en main par son adjoint direct, René Garrec, qui lui-même devait quelques années plus tard devenir à son tour Président du Conseil régional. Le projet a été lancé sur le plan de la politique régionale, dès le début des années 1980. Il était soutenu par les instances politiques de l'époque, mais il lui manquait cruellement un soutien puissant de la part des

instances scientifiques. Il suscitait plutôt de petits sourires, voire des quolibets de la part des grosses universités qui étaient sur les rangs pour obtenir, elles aussi, cet équipement : Paris, Lyon, Toulouse. Installer un tel centre dans un « pré à vaches » paraissait totalement hors de logique. Mais c'était sans compter sur le fait que, justement, le Ganil déjà avait été installé à la suite d'un fameux combat qui avait opposé entre eux des physiciens d'un peu partout en France. Caen avait été choisi pour le Ganil, notamment du fait du poids politique que représentait Michel d'Ornano, à l'époque ministre de l'Industrie.

Le CEA et le CNRS étaient les deux instances parties prenantes dans le Ganil, mais le CEA était l'instance la plus puissante, celle qui avait le plus de moyens financiers. Comme je m'étais ouvert du projet à un certain nombre de personnes, le directeur du Ganil, Michel Lefort, m'avait dit : « votre projet ne peut aboutir s'il n'a pas un soutien scientifique ; il faut que vous rencontriez Jules Horowitz, et que vous le lui exposiez. S'il est partant, vous aurez une autoroute ». Jules Horowitz – qui a finalement donné son nom au campus – était le directeur de l'Institut de recherche fondamentale du CEA, c'est-à-dire qu'il avait la haute main sur tous les projets scientifiques du CEA, pas seulement dans le domaine de l'énergie nucléaire, mais aussi le contrôle des activités dans le domaine des sciences du vivant. En sciences biomédicales, le seul centre en France doté d'une caméra à positons était le service Frédéric Joliot (SHFJ), installé à Orsay, totalement sous le contrôle du CEA. Un jour, on m'a convié à déjeuner avec Monsieur Horowitz, qui était entouré de tout un aréopage du Ganil, ainsi que de quelques conseillers qui l'avaient accompagné depuis Paris. C'était impressionnant pour le jeune médecin que j'étais, pas encore hospitalo-universitaire, de se trouver face à

ce personnage devant qui tout le monde était extrêmement respectueux et attendait son tour pour donner son opinion, et le tour, c'était aussi lui qui le donnait ! Je lui ai donc exposé les grandes lignes du projet et je lui remis une petite note à son sujet. À la fin, il a dit : « nous allons réfléchir à cela ». Et quelques semaines – même pas un mois – après, le retour est venu. On m'a affecté un petit bureau et une secrétaire dans les locaux de Ganil, dont un certain nombre de responsables avaient été désignés pour accompagner la poursuite du projet. Très vite, M. Jules Horowitz avait voulu qu'il prît une forme juridique : ainsi a été créé un Groupement d'intérêt public (GIP), qui était à l'époque une structure nouvelle ayant pour objectif l'association d'organismes privés et publics dans des projets scientifiques ou industriels. Cette formule était très soutenue, à l'époque, par le ministère de la Recherche sous la responsabilité de Jean-Pierre Chevènement. On a donc créé un GIP pour la construction de notre centre de recherche, qui au départ s'appelait Institut de recherche métabolique, c'est-à-dire IRM, mais ce n'était pas un choix très heureux parce que justement, au ministère, s'affrontaient les tenants du développement de la résonance magnétique et ceux de la tomographie à positons. Toutes ces techniques étaient tellement chères et nouvelles qu'on ne pensait pas réaliste, à l'époque, de mener de front le développement des deux. Donc, à l'initiative de Dominique Comar, qui était un des pionniers du SHFJ d'Orsay, le nom de Cyceron a été proposé à partir des éléments suivants : il y a un cyclotron, il y a de la chimie, il y a du positon (« positron » en anglais), ce qui a donné CYCERON, nom qu'il conserve encore 40 ans plus tard.

Ensuite, il y a eu la phase de conception. Je vous passe tous les aléas politiques. Entre-temps en effet, la majorité avait changé, il y avait eu l'élection de François Mitterrand, et il

fallait trouver désormais, dans la nouvelle majorité nationale, des soutiens aussi vigoureux que ceux qui persistaient désormais uniquement au plan régional. Par l'entremise de Jean-Pierre Houtteville, alors chef du service de neurochirurgie, qui était affilié au parti socialiste, je suis entré en contact avec François Geindre, qui était le maire d'Hérouville-Saint-Clair. Il était très influent au sein du parti socialiste, proche de Jean-Pierre Chevènement et de Laurent Fabius, les deux ministres de la recherche successifs de l'époque. Pour débloquer les situations les plus complexes, c'est un ministre d'État, Louis Mexandeau, très proche de François Mitterrand, qui a été approché. Ainsi, le couple Geindre-Mexandeau a accompagné, côté nouvelle majorité nationale, ce projet qui continuait par ailleurs d'être soutenu au plan régional par René Garrec, M^{me} Anne d'Ornano – présidente du Conseil départemental, et Jean-Marie Girault – maire de Caen, tous trois de la même famille politique.

Sur le plan scientifique, le CNRS a participé, surtout par la mise à disposition de postes et d'agents. Puis l'Inserm, qui était dirigé par Philippe Lazar. Finalement, deux projets ont été retenus en France, en plus du SHFJ qui existait déjà à Orsay : le projet de Lyon, qui s'est appelé le CERMEP, et le projet de Caen, qui a été en fait réalisé bien avant celui de Lyon. Mais il fallait financer tous ces projets et, en 1984, les finances de la France n'étaient pas au mieux. Pour débloquer le financement, il fallait que les trois grands organismes contribuent, non seulement par de l'investissement, mais aussi en fonctionnement. Avec Édouard Zarifian, professeur de psychiatrie très impliqué dans la recherche médicale, et notre doyen Gérard Lévy, nous avons été invités à un dîner organisé par Louis Mexandeau. Je pense qu'une telle circonstance arrive extrêmement rarement : étaient réunis en effet le

directeur général de l'Inserm, le directeur des sciences du vivant du CNRS et le directeur de la recherche fondamentale du CEA. En l'espace d'une soirée, le budget a été débloqué ! Les différentes institutions ont contribué, même si la plus grosse part de l'investissement revenait au CEA, côté État. Le montage reposait sur un partage de l'investissement entre les collectivités locales d'un côté, les institutions scientifiques (donc l'État) de l'autre côté, à égalité. Quant au fonctionnement, il était entendu que seuls les partenaires nationaux – CEA, CNRS et Inserm, y contribueraient. Cette contribution concernait les postes (chercheurs, ingénieurs, techniciens) et le budget hors salaires. À l'époque, il y avait des liens excellents entre le corps médical et la direction du CHU. Or, j'avais compris très vite que les salaires constitueraient l'essentiel de la participation des organismes scientifiques, alors qu'il fallait aussi un budget conséquent pour assurer les dépenses hors salaires. Grâce au premier directeur du CHU, Monsieur Hartmann, qui était devenu Directeur général de la Caisse nationale d'assurance maladie à Paris (CNAMTS), nous avons obtenu un budget exceptionnel et reconductible, de soutien à la recherche faite à Cyceron au titre de l'assurance-maladie. C'était un poste totalement nouveau et qu'il a fallu, au fil des ans, constamment justifier et légitimer, mais qui a été une jouvence pour le démarrage de Cyceron.

On avait donc eu le soutien politique, et ensuite le soutien scientifique, puis les financements et les schémas de financement. Il restait à recruter les membres qui, au sein du laboratoire, développeraient concrètement la recherche. En effet, comme l'avaient dit ses détracteurs, on installait ce projet très ambitieux de recherche biologique et médicale dans un espace hospitalo-universitaire où il n'y avait à l'époque qu'une

seule unité Inserm, c'était celle de Jean-Pierre Pujol, spécialiste du tissu conjonctif, donc très éloignée des thématiques de Cyceron. Il y avait une bonne école de neurologie clinique, mais pas de laboratoire de neurosciences, ni CNRS, ni Inserm. Raymond Chichery était, à l'Université de Caen, un spécialiste du cerveau de la seiche, tout de même très éloigné de celui des mammifères et de l'homme ! Il a donc fallu se mettre au travail pour constituer un noyau de départ de chercheurs. Côté CHU, Bernard Lechevalier travaillait depuis toujours dans le domaine des troubles cognitifs, chez les patients neurologiques, et s'était orienté très tôt vers la neuropsychologie. Parmi ses collaborateurs les plus directs, il y avait Francis Eustache qui s'intéressait, non seulement à la cognition et aux évaluations neuropsychologiques, mais qui a été intéressé tout de suite par les méthodes d'investigation du cerveau, anatomiques et fonctionnelles. Grâce à Francis Eustache et à Bernard Lechevalier, j'ai pu faire acquérir par le CHU le seul appareil existant en France, permettant la mesure du débit sanguin cérébral utilisant le xénon radioactif, le xénon 133, qui a été installé dans le laboratoire de médecine nucléaire du CHU. On a pu ainsi, pendant plusieurs années, réaliser avec cet équipement des travaux sur la pathologie vasculaire cérébrale, puis sur les fonctions cognitives. Il nous a permis de nous initier à l'imagerie fonctionnelle cérébrale, sachant que la construction de Cyceron n'a commencé qu'en 1986, et que l'ouverture n'a eu lieu qu'en 1989. Il s'est donc écoulé dix ans entre la première note remise à la Préfecture de région et l'ouverture effective du laboratoire.

Ainsi, du côté du CHU, il y avait un noyau existant de cliniciens, mais il fallait trouver aussi des neurologues connaissant la méthodologie de la tomographie à positons, et des chercheurs maîtrisant les techniques recherche

expérimentale. Marie-Germaine Bousser, une amie et collègue avec qui j'avais beaucoup travaillé pendant mes années d'internat et de clinicat, et qui était professeure de neurologie à Paris, m'a mis en contact avec les membres du SHFJ d'Orsay. Il y avait là tout un groupe de réflexion sur l'avenir de la tomographie à positons. J'y ai fait la connaissance de Jean-Claude Baron. Il a rejoint Cyceron, avec un poste de directeur de recherche Inserm. Puis il a constitué le noyau de départ, auquel s'est associé Francis Eustache. Ce groupe comprenait aussi des chercheurs Inserm rattachés à Édouard Zarifian : Pierre Etévenon, qui était spécialiste d'électroencéphalographie, Jean-Philippe Boulanger et Jean-Claude Bisserbe, tous deux psychiatres. Édouard Zarifian, toujours lui, m'avait mis en contact, à l'époque où je cherchais des financements du côté de l'industrie pharmaceutique, avec Synthélabo qui était un énorme complexe de l'industrie pharmaceutique basé à Milan. Le PDG de cette entreprise m'a proposé d'intégrer un très brillant chercheur, spécialiste de la circulation cérébrale, qui souhaitait reprendre une activité de recherche publique. C'était Éric Mackenzie qui, dès le début, a participé à la conception des plans de Cyceron pour ce qui concernait la partie expérimentale. Après que le CNRS l'eut recruté, il a amorcé la constitution d'une unité qui a, depuis, donné des « petits » qui sont encore en activité à ce jour.

Il y avait trois points-clés dans Cyceron : le cyclotron, la chimie, l'informatique et le traitement de l'image. Pour le cyclotron c'était Ganil qui avait mis à notre disposition un spécialiste, Claude Le Poec. Pour la chimie, c'est une jeune thésarde de l'équipe Thuillier de l'école d'ingénieurs (ENSI) de Caen, Louisa Barré, qui a été se former à Orsay, et qui est devenue la responsable du Laboratoire de radiochimie, qu'elle a quitté il y a seulement un an pour prendre sa retraite.

Concernant l'informatique et le traitement d'images, c'est Jean-Marcel Travère qui a été recruté par le CEA, et qui a géré durant près de dix ans ce secteur de recherche. Il a quitté Cyceron en 1995 pour partir à Cadarache dans un autre laboratoire du CEA. Pour terminer cette histoire de la fondation de Cyceron, il faut préciser que la technologie avait aussi évolué durant ces années de conception, avec notamment la mise au point de petits cyclotrons compacts dédiés à la production de radioéléments à vie courte émetteurs de positons, que l'on pouvait installer dans les unités de recherche. On n'a donc jamais utilisé les cyclotrons de Ganil, un cyclotron compact ayant été implanté dans le laboratoire lui-même.

Francis Eustache :

Merci Jean-Michel d'avoir planté le décor, merci doublement parce que tu viens de résumer, en un temps record, plusieurs décennies, et merci surtout de l'avoir planté pour de bon parce que tu as fait un travail considérable et pionnier. Même si tu as cité de nombreuses personnes autour de toi, il faut le dire, Cyceron, c'est une grande partie ton œuvre. Je vais me présenter à mon tour et nous allons croiser nos chemins, évidemment, à plusieurs moments.

J'ai fait une partie de mes études à Caen, des études de psychologie. J'ai rencontré dès la licence les neurologues, Bernard Lechevalier et Pierre Morin, qui m'ont accueilli en stage dans leurs services. Les psychologues n'étaient pas légion, à cette époque, dans les services de neurologie. J'ai continué mes études à Paris où j'ai d'abord préparé ce qu'on appellerait aujourd'hui un master professionnel parce que j'étais très intéressé par la clinique. Je suis ensuite revenu à Caen et j'ai obtenu assez vite un poste de neuropsychologue

dans les services de neurologie du CHU de Caen, qui était tout juste construit. C'était en 1979, année qui coïncide avec la petite note remise par Jean-Michel Derlon au Préfet de région. Je croise ainsi Jean-Michel Derlon qui était tout jeune aussi, Jean-Pierre Houtteville : la neurochirurgie, la neurologie seront mon domaine pendant de nombreuses années. Ces services étaient très en pointe et jouissaient d'une excellente réputation au plan national. Il y avait beaucoup de jeunes médecins battants et une vraie volonté de s'engager dans la recherche. Même si les moyens étaient limités, cet environnement m'a donné envie, en parallèle de mon activité de neuropsychologue clinicien, de faire de la recherche et d'essayer de faire coïncider la recherche et la clinique. Dans les années 1980, nous examinions les patients à la lumière des « modèles cognitifs » et nous faisons ainsi une recherche clinique de qualité, mais l'imagerie cérébrale était quasi-inexistante.

Par ailleurs, j'ai eu l'opportunité de préparer une thèse de psychologie à l'Université Paris-Descartes en parallèle de mon travail quotidien de neuropsychologue, thèse qui était dirigée par le Professeur Jean-Louis Signoret à la Salpêtrière. Il y a toujours eu beaucoup de liens scientifiques et d'amitié entre Caen et l'hôpital de la Salpêtrière. Ce travail de thèse utilisait le « fameux » appareil au Xénon 133 installé dans le service de médecine nucléaire du CHU. Je pense que c'était la première thèse en France utilisant une telle méthodologie. Ces moyens étaient encore très rudimentaires pour obtenir des résultats vraiment significatifs, mais on arrivait quand même, comme le faisaient les Scandinaves à l'époque, à montrer que de grandes régions cérébrales étaient impliquées dans le langage, les activités spatiales ou la mémoire. Bien sûr, on savait tout cela depuis longtemps par d'autres biais, mais

c'était quand même fascinant de le montrer de façon directe chez l'humain. Ces premiers travaux nous ont donné du baume au cœur pour aller plus loin, dans la création d'un véritable institut qui est devenu Cyceron. C'est vrai qu'il y avait un paysage intellectuel et scientifique assez fascinant dans ces années 1980, où l'on parlait quasiment de rien.

Puis, il y a eu cet institut avec le cyclotron, la première caméra à positons qui était d'ailleurs une caméra française. Nous étions dans les Trente Glorieuses et il existait cette volonté d'ouvrir des portes dans un domaine complètement nouveau : la compréhension du cerveau humain vivant, y compris dans ses activités cognitives les plus complexes. Des chercheurs de haut niveau, que Jean-Michel Derlon vient de citer, nous ont rejoints. Mais il y avait aussi, localement, dans le monde hospitalo-universitaire et universitaire, des personnes de qualité ayant un grand sens du bien commun. Notre force, à Caen, dans une ville de taille moyenne, en comparaison des mastodontes – je ne parle même pas de Paris – de Lyon, de Toulouse, de Marseille, enfin, toutes les villes qui étaient sur les rangs, c'était de serrer les coudes pour faire en sorte que Cyceron démarre et soit productif. Ce sont les années 1990, le début des années 2000 avec tout de suite une activité scientifique magnifique, au plus haut niveau, avec des articles qui sortaient dans *Brain*, ou dans d'autres journaux de ce niveau, quasiment mensuellement. C'est une période où des avancées majeures ont été réalisées, dans plusieurs secteurs. Le domaine de la pathologie vasculaire était exploré, tant au plan expérimental – c'était le travail de l'équipe de d'Éric McKenzie – que chez l'homme – c'était les travaux de Jean-Claude Baron. La tomographie à positons permettait de visualiser les phénomènes de diaschisis : une lésion focale du cerveau entraîne un dysfonctionnement à distance, qui

explique à son tour des manifestations neuropsychologiques, que ce soient l'aphasie ou les phénomènes d'héminégligence. Il s'agissait vraiment d'une réécriture de la compréhension des dysfonctionnements du cerveau. Un autre domaine, encore plus nouveau, était l'étude de la cognition où l'on détournait en fait l'utilisation première de la caméra positons pour explorer les fonctions cognitives chez l'homme sain – sans pathologie – ce qui était totalement nouveau. Cette approche permettait de visualiser le cerveau en fonctionnement, lors d'activités complexes et cela pour la première fois avec une réelle précision.

Cyceron faisait partie de la dizaine de centres au monde où était mené ce type d'activité et tous les chercheurs se connaissaient. C'était comme une grande famille. On rencontrait à Cyceron des chercheurs japonais, italiens, anglais, belges, allemands qui venaient de ces rares centres comparables à Cyceron de par le monde. Outre l'objet d'étude, ce type de collaboration était totalement nouveau et nous avons eu conscience de vivre un moment fécond. Pour ma part, je suis allé beaucoup à Londres, dans le laboratoire de Richard Frackowiak, sans doute le laboratoire le plus innovant et le plus productif, surtout pendant les années 1990. Dans une salle, il y avait des chercheurs d'horizons complètement divers qui venaient des quatre coins du monde et qui discutaient autour d'une grande table. Aujourd'hui, cela semble évident, dans un centre de recherche, que des chercheurs viennent d'un peu partout. Mais à l'époque, c'était nouveau et très enthousiasmant et stimulant.

Au début des années 2000, les travaux sur les maladies dégénératives, notamment la maladie d'Alzheimer, ont été, selon moi, les recherches les plus marquantes, réalisées à Cyceron. Dans la maladie d'Alzheimer, la plus fréquente, la

maladie emblématique, nous avons mis en évidence, avec ma collègue Béatrice Desgranges, les modifications cérébrales à l'origine des troubles cognitifs. De petites zones d'hypométabolisme provoquent des altérations de la mémoire ou des autres fonctions cognitives. Ensuite, à Cyceron comme ailleurs, d'autres techniques sont venues compléter la tomographie par émission de positons. C'était davantage une complémentarité qu'une concurrence. Ainsi, l'IRM anatomique, dans un premier temps, montrait la coïncidence, j'ai presque envie de dire la connivence, entre les altérations structurales et les altérations fonctionnelles, l'ensemble permettant de comprendre la survenue des troubles cognitifs. Puis, de façon plus surprenante encore, des mécanismes de compensation venaient contrecarrer ces premières anomalies. Nous étions partis dans un premier temps sur un schéma assez classique selon lequel des lésions ou des dysfonctionnements du cerveau allaient entraîner des troubles neuropsychologiques, des symptômes. En fait, on aboutissait à un résultat beaucoup plus complexe selon lequel le cerveau était capable de mécanismes de compensation. Ces mécanismes étaient renforcés par différentes conditions favorisées par la réserve cognitive. Ce domaine est devenu à son tour un objet de recherche car il était essentiel de comprendre comment on pouvait amplifier ces mécanismes de réserve cognitive.

Ce résultat va très loin. Le cerveau est un objet éminemment complexe et, en analysant sa structure et son fonctionnement avec de nouveaux outils, on découvre une façon de vivre et d'interagir avec le monde sous une forme complètement nouvelle. Notre pensée reste indispensable, y compris nos réflexions philosophiques, métaphysiques, existentielles, mais cette pensée peut s'appuyer sur ces nouveaux outils

d'exploration du cerveau qui n'ont cessé d'évoluer depuis lors. Aujourd'hui, les caméras et autres scanners évoluent toujours, gagnent notamment en définition, spatiale et temporelle. De plus, les techniques qui s'appliquent à l'analyse des données sont de plus en plus puissantes et il est possible de les partager à l'autre bout du monde. Les chercheurs peuvent travailler, beaucoup plus facilement, en étant à distance. En mettant en commun à la fois leurs données, et ces outils d'analyse de plus en plus sophistiqués, ils parviennent à de nouvelles approches qui produisent de nouveaux résultats.

In fine, si je devais définir Cyceron, sans en faire l'historique complet depuis ses fondations durant les années 1980, je dirais que c'est un endroit d'intelligence où les jeunes peuvent travailler sur ces phénomènes fascinants que sont le fonctionnement et les dysfonctionnements du cerveau, une nouvelle approche des maladies du cerveau, enjeux absolument considérables pour les années à venir. C'est un endroit où on se pose des questions. On arrive, le matin à 7h30, en attendant le professeur parce que, pendant toute la nuit, on a été préoccupé par un problème qu'on n'arrive pas à résoudre. Ce sont de beaux moments de recherche et de créativité. Travailler sur des données, produire de beaux articles dont on est tous fiers collectivement. Assister à des soutenances de thèse de jeunes qui, lorsqu'ils sont arrivés dans ce lieu, avaient sans doute une idée extrêmement floue de comment tout ça fonctionnait, puis qui, trois ou quatre ans plus tard, ont énormément progressé, et au-delà de l'acquisition de savoirs spécifiques, sont davantage prêts à affronter le monde.

Jean-Michel Derlon :

Oui, je crois que ce qui est important, c'est cette intelligence collective, partagée, que tu décris très bien. Le cas de Cyceron était très intéressant parce que c'était l'un des rares laboratoires de recherche biomédicale, en France, où les médecins et ceux qu'on appellerait les fondamentalistes cohabitaient et échangeaient. C'est lié aussi à l'outil lui-même, parce que la tomographie à positons, c'est la physiologie par l'image, qui permet de « vulgariser » les concepts physiologiques. Par exemple, pour mesurer localement le débit sanguin du cortex chez le singe, il fallait utiliser des sondes à hydrogène, on était à des années-lumière de la pratique clinique. Donc l'étudiant en médecine, qui avait un intérêt scientifique, devait changer complètement de paradigme quand il arrivait dans le laboratoire de recherche, et il fallait qu'il se recycle complètement avant d'être au niveau des partenaires fondamentalistes avec qui il interagissait. À Cyceron au contraire, grâce aux outils de recherche, cohabitaient dans un même laboratoire, une recherche clinique avancée et une recherche expérimentale sur des modèles animaux souvent complexes et sophistiqués. Les étudiants qui arrivaient là avaient une opportunité extraordinaire de se rencontrer : les cliniciens d'un côté, les biologistes et les physiologistes de l'autre. Cela a été à l'origine, je pense, d'un état d'esprit. Et je me souviens, il y a quelque temps, j'avais assisté à une soutenance de thèse. Je bavardais ensuite avec l'étudiante sur les quatre années qu'elle avait passées au sein du labo. Elle m'a décrit ce choc d'un espace où des cliniciens qui quittaient leur cabinet de consultation ou leur salle d'opération se retrouvaient avec des biologistes qui étaient sur leur PC ou leur dispositif de biologie moléculaire. Le fait que tout cela fonctionne avait eu pour elle un caractère

extraordinaire. Un des objectifs de Cyceron avait été précisément de casser les barrières disciplinaires, et de faire que chacun apporte sa contribution et s'imprègne de l'approche de l'autre en sorte qu'après quelques années, un biologiste était capable de faire un exposé de pathologie humaine aussi bien qu'un médecin, et vice-versa.

Francis Eustache :

Oui, cet aspect interdisciplinaire, on le ressent toujours fortement entre cliniciens et biologistes. J'ai pu le mesurer car j'ai eu l'opportunité de diriger une unité de recherche pendant une vingtaine d'années et d'avoir dirigé aussi Cyceron, après Jean-Michel, c'était de 2013 à 2016. Et ce qui s'est passé encore de nouveau, tout récemment, c'est d'intégrer les sciences humaines et sociales, ce qui était quelque chose d'inconcevable auparavant. Par exemple, impliquer des historiens dans ces problématiques. Ce rapprochement s'est fait par le biais de l'étude de la mémoire. La mémoire d'une personne, c'est la mémoire d'un cerveau, mais notre cerveau est en contact avec l'environnement, avec la culture, avec notre identité sociale. Grâce à la rencontre avec un historien, Denis Peschanski, qui est le président du Conseil scientifique du Mémorial de Caen, une autre institution remarquable, nous avons pu développer des travaux aux confins des neurosciences cognitives et de l'histoire. Nous avons pu décrire les réseaux cérébraux impliqués dans la mémoire individuelle, mais il existe également des régions cérébrales impliquées dans la mémoire collective, dans nos représentations collectives, en quelque sorte notre mémoire historique. Notre mémoire collective s'immisce dans notre cerveau et colorie les réseaux de notre mémoire individuelle.

C'est une autre dimension qu'a permis Cycleron, avec des articles retentissants dans les meilleures revues scientifiques. Ces travaux fondamentaux nous permettent aussi de rebondir sur la recherche clinique et notamment sur une pathologie qui vient de l'environnement : le trouble de stress post-traumatique. Une personne est confrontée à un événement traumatique, et le fonctionnement du cerveau se trouve modifié par ce traumatisme psychique. Pour comprendre ce type de pathologie, il est indispensable de mettre ensemble toutes ces connaissances, tous ces savoirs, faire en sorte que toutes ces disciplines interagissent, de la physique à la biologie mais également de la neuropsychologie aux neurosciences cognitives et aux sciences humaines et sociales. Cycleron a été conçu dès l'origine pour associer des chercheurs d'origines très diverses, le succès de l'entreprise reposait en partie sur l'oubli mutuel de ses origines.

Propos recueillis à Caen, le 22 février 2022

Iliès Zaoui et Conscience Robotics

Le chemin d'une réussite

Entretien réalisé par MM. Laurent Bellamy, Didier Laforge et Edgar Leblanc

La jeune carrière d'Iliès Zaoui s'est entièrement construite sur un projet élaboré lors de ses études universitaires, approfondi pendant une activité professionnelle de sept années et abouti avec la création de Conscience Robotics : comment un robot peut-il se déplacer dans un environnement donné, « sans passer par la phase de pré-séquençage habituelle » ?

Étudiant à l'Université de Caen Normandie, il prépare un Master recherche « algorithmes et modèles de l'information ». Le diplôme obtenu en 2011, il est engagé en 2012 chez Capgemini où, pendant sept années, il travaille pour de grands groupes de la communication - M6, TF1 -, de l'aviation – Air France KLM, Dassault Aviation – et de l'agroalimentaire – Sodexo. Ces diverses activités lui permettent d'approfondir et de préciser son projet de fabrication d'une intelligence artificielle qui donnera à un robot quelconque les capacités de se déplacer et d'intervenir dans un environnement inconnu et en mouvement. C'est en 2012 que sera écrite la première ligne de code d'un projet qui deviendra Conscience Robotics.

En novembre 2017, il fonde Conscience Robotics, une TPE innovante fabriquant des produits commercialisables et embauche dès juillet 2018 un premier salarié. Contrairement à une startup qui n'a souvent pas de chiffre d'affaires, l'entreprise d'Iliès Zaoui se positionnera délibérément sur le marché, recherchera l'équilibre financier en surveillant l'évolution des charges et des produits. L'entreprise est d'abord accueillie par Normandie incubation qui fournit des locaux, une aide - remboursable, mais fort utile au

démarrage – à la recherche, ainsi qu’un encadrement administratif formateur pour la suite. Dans cette première étape, la priorité revient à la recherche et au développement R&D. Des partenaires financiers sont sollicités. Un créateur de startup apportera 250 000 €. Un emprunt obligataire convertible en actions du même montant sera souscrit auprès de la BPI, partenaire financier. Iliès Zaoui fait le choix de solliciter un minimum de subventions et de ne contracter aucun emprunt bancaire classique, pariant que les recettes obtenues par la vente des produits conduiront à l’équilibre et à la pérennité de l’entreprise.

Le système développé par Conscience Robotics est simple, évolutif et compatible avec une large gamme de robots. Après une installation simplifiée le robot dispose d’une intelligence artificielle nouvelle génération. Sans aucune intervention humaine, il tire le maximum de lui-même en fonction de ses capacités physiques et en apprenant de ses erreurs. Grâce à Conscience Robotics, chaque robot développe ses propres solutions face à une situation inconnue quels que soient le contexte et l’environnement. L’apprentissage d’une nouvelle solution par un robot est alors partagée avec tous les robots du même type. Ces compétences aussi appelées « intelligences » sont accessibles à travers le « Conscience store ». L’utilisateur pourra y choisir les « intelligences » à installer sur son robot. Ainsi il appartiendra à chaque robot doté de la solution « Conscience Robotics » de développer son plein potentiel selon les besoins de son utilisateur. L’intelligence artificielle partagée, permettra par ailleurs une collaboration entre les robots afin d’optimiser la réalisation des tâches et ce, en fonction de leurs capacités respectives.

Le système est déjà déployé et commercialisé. « Ariel » effectue et optimise la certification de nettoyabilité d’équipements agroalimentaires et renforce la sécurité

sanitaire de l'alimentation quotidienne pour l'entreprise Actalia. « Bering » autonomise un robot qui effectue le marquage au sol pour les Aéroports de Paris. « Median » est un robot qui intervient dans les process de télémédecine, il permet au praticien de prendre le contrôle à distance du robot qu'il va déplacer de façon autonome dans la chambre de la personne, le bras se déploie pour visionner de très près par exemple une blessure. « Asset tracking » est une caméra qui permet d'inventorier en temps réel les contenus d'un hangar d'avions. Ou encore « Conscience Care » qui dans les EHPAD va permettre à une IA de détecter un comportement anormal soit par l'intermédiaire d'une caméra soit en utilisant un bracelet.

Conscience Robotics, en quatre années, a connu un développement suffisant pour qu'il soit qualifié de réussite. Avec le crédit impôt recherche – toujours compliqué et onéreux à obtenir par le temps nécessaire pour compléter les dossiers – l'entreprise a atteint son équilibre économique et l'avenir à moyen terme paraît assuré. Aucun brevet n'a été déposé, le temps et les moyens à consacrer à cette activité sont investis dans la recherche et le développement. Ces choix confortent la stratégie de l'entreprise de se concentrer sur la réponse aux demandes ou besoins du marché pour assurer sa pérennité.

Le nombre de collaborateurs est passé de deux à douze, et une attention particulière est portée par le jeune chef d'entreprise à la constitution d'une équipe volontaire et motivée. Les juniors recrutés sont tous diplômés de l'Université, adeptes du travail en équipe, soucieux de pouvoir « sélectionner les tâches » quotidiennes et de long terme qu'ils accomplissent, avec un fort désir de « surliberté » dans le travail qui conduit à mettre en place des formes de télétravail choisis, voire revendiqués, autant que des procédures collectives

avec séances de répartition des tâches et de bilan quotidiennes, hebdomadaires et mensuelles. Une plus forte organisation collective pour une plus grande liberté individuelle ; un management stratégique et attentif au quotidien : Iliès Zaoui cultive une stratégie collaborative assumée.

Et les perspectives ne manquent pas. Les projets sont nombreux : Conscience Robotics a l'innovation dans son ADN. La prochaine étape sera la plateforme bipède Enki. Ce robot aura la taille d'un enfant de quinze ans et sera un



domestique de maison. Il sera capable de faire exactement les mêmes choses qu'un humain, c'est-à-dire se déplacer, monter un escalier, saisir et utiliser des objets du quotidien. Dans la maison connectée, l'environnement est modifié pour le rendre automatique : avec Enki c'est le robot qui s'adapte à l'environnement. À moyen terme pour l'entreprise, Iliès Zaoui imagine une usine robotisée pour produire des robots dont l'utilisation sur une large échelle permettrait de réduire la



penibilité du travail humain. Il se projette dans une société où le travail physique serait exécuté par des robots et l'homme dégagé de ses tâches matérielles, pourvu d'un quelconque revenu universel, serait libéré de la contrainte salariale. Et il s'interroge sur le devenir d'une humanité ayant épuisé les ressources de la planète Terre et poussée à la conquête de l'espace interplanétaire, nouvelle frontière pour l'humanité. Quelle place alors pour l'homme dans un univers d'où le travail, activité réalisatrice autant que nourricière, aurait disparu ? Un jour, peut-être, en sortira-t-il un livre : écrire, le rêve caché du créateur d'entreprise ?

INTELLIGENCE ARTIFICIELLE À L'UNIVERSITÉ D'OXFORD : L'INTÉGRATION DES DISCIPLINES

*par M. Frédéric THIBAUT-STARZYK
de l'Académie des Sciences, Arts et Belles-Lettres de Caen
Ancien directeur de la Maison française d'Oxford*

LA MAISON FRANÇAISE D'OXFORD

La Maison française d'Oxford (MFO) est une institution française au sein de l'Université d'Oxford. C'est un Institut français de recherche à l'étranger, unité mixte du CNRS et du Ministère de l'Europe et des Affaires étrangères. La MFO a pour vocation de mener des recherches et d'amener la science et la culture française en Grande-Bretagne, d'être un lien et un centre d'information pour les universitaires britanniques vers la France. Pour mener ses recherches, la MFO héberge des chercheurs français dans toutes les disciplines pour des séjours ponctuels ou de longue durée. Elle invite des intellectuels et des artistes à présenter leur travail à Oxford.

Au cœur de l'Université d'Oxford, impliquée dans l'ensemble des disciplines, la MFO est un poste avancé au sein d'une Université de tout premier rang \$1 on peut y appréhender, dans un système différent du nôtre, une approche particulière et globale de la recherche en Intelligence Artificielle qui a pris une des premières places internationales.

Ce texte n'est pas une contribution scientifique à la réflexion sur l'Intelligence Artificielle, mais plutôt la présentation du point de vue particulier sur la recherche en IA que donne la position d'un centre français interdisciplinaire au sein d'une université étrangère majeure, si différente des universités françaises et de notre pratique académique.

OXFORD : UNE ORGANISATION PARTICULIÈRE

Les universités d'Oxford et de Cambridge ont ceci de particulier qu'elles ont deux systèmes d'organisation parallèles. En plus des facultés et départements, les enseignants et leurs étudiants sont regroupés par *colleges*. Les *colleges* couvrent l'ensemble des disciplines, ils offrent notamment à leurs enseignants un bureau, une salle commune et un restaurant, à leurs étudiants un lieu de vie (chambre, restauration, sport). Les étudiants y bénéficient d'un encadrement rapproché et de cours en petits groupes de trois, en plus des enseignements dispensés dans les départements de l'Université. S'y retrouvant régulièrement, pour manger ensemble et pour les nombreux événements sociaux au sein du *college*, les universitaires sont amenés à échanger en dehors de leur discipline, ils se rencontrent et se connaissent beaucoup plus, au-delà de leurs domaines de spécialisation, que dans les universités habituelles. La faculté ou le département n'est ainsi qu'un lieu de travail, alors que la communauté de vie et de réflexion est le *college*, sans étiquette disciplinaire, avec des collègues de tous horizons. Des conférences et discussions larges sont organisées dans les *colleges*, forçant l'interdisciplinarité comme au sein d'une académie.

L'IA est par essence interdisciplinaire et l'université des *colleges* s'y prête particulièrement bien.

L'autre force d'Oxford dans le domaine des technologies, moins connue, est d'avoir toutes les sciences de l'ingénieur au sein d'un seul et même département. Dans d'autres grandes universités, on trouve en général un département de génie chimie, un département de génie électrique, de génie informatique, etc. Avec toutes ses forces en ingénierie réunies et, là encore, une certaine interdisciplinarité, Oxford a obtenu une visibilité internationale de premier plan et occupe régulièrement le podium des classements¹. Cette

¹ *Times Higher Education*, 2020.

interdisciplinarité de l'ingénierie donne à Oxford une force considérable pour couvrir de manière homogène l'ensemble des aspects et traiter l'ensemble des enjeux de l'IA.

L'IA À OXFORD

Dans Oxford, le signe le plus clair de la recherche en IA est la présence de ces drôles de voitures bardées de capteurs qui vous dépassent dans votre trajet vers le centre-ville à vélo. Malgré les logos de l'Université et la mention d'Oxbotics en lettres futuristes, il ne s'agit pas de voitures publicitaires mais de véhicules robots avec leur conducteur au rôle tuteur et bientôt accessoire. L'IA pour les véhicules autonomes est donc localement une part importante de la discipline. Les autres grands domaines sont bien sûr aussi présents. Comme on l'a vu dans l'actualité récente du vaccin, la recherche biomédicale est une facette majeure d'Oxford. Dès les années 1990, les chercheurs du département d'Ingénierie ont proposé à leurs collègues médecins un institut d'ingénierie pour la Biomédecine. L'apprentissage automatique ou statistique (*machine learning*) leur fournit un nouvel outil, notamment pour le suivi thérapeutique et l'aide par diagnostic automatique. De nombreux autres instituts forment maintenant le paysage des sciences de l'information et de l'intelligence artificielle à Oxford (Oxford Internet Institute, Institute of Biomedical Engineering, Oxford Robotics Institute, Big Data Institute...). Du point de vue des forces en jeu, les domaines d'application de l'IA dans cette université se répartissent approximativement entre 50 % pour la médecine, 25 % pour les véhicules autonomes et 25 % pour finance et l'environnement. Toutes ces recherches sont en interaction forte avec la sociologie, l'éthique et le droit.

MÉDECINE ET BIO-MÉDECINE

Un des objectifs les plus simples pour l'informatique en médecine a été de numériser le dossier médical des patients.

Le bénéfice n'en est pas seulement l'élimination du dossier papier. Les données enregistrées peuvent être traitées par ordinateur pour, par exemple, une aide au diagnostic. Les électrocardiogrammes fournissent beaucoup plus de données que ce que l'œil humain peut y déceler : l'ordinateur peut en extraire les informations sous-jacentes. Les images médicales peuvent être traitées automatiquement pour y rechercher des informations peu accessibles ou habituellement inexploitées. Le suivi médical des patients peut être aidé par des instruments portés en permanence par la personne, patches ou vêtements connectés qui surveillent des paramètres biologiques. Le tout n'est possible qu'avec des outils de traitement de l'information issus de l'apprentissage automatique.

VOITURE AUTONOME

Par le biais d'Oxbotica, compagnie dérivée (*spin-off*), le logiciel de pilotage de véhicule mis au point à l'Université a été distribué dans de nombreux pays. Il est lié à un système distant qui peut coordonner une flotte de véhicules. Ces technologies s'appliquent aussi dans des domaines très divers comme l'exploitation minière ou l'agriculture. Pour mettre au point ce genre de logiciel et améliorer sa sécurité, des simulations sont réalisées en utilisant des comportements humains réels pour les autres usagers de la route. Le logiciel peut alors intégrer et imiter ces comportements par le biais de l'apprentissage automatique.

FINANCE ET ENVIRONNEMENT

Dans le secteur de la finance, l'IA joue déjà un rôle important pour la détection des fraudes ou pour les transactions boursières pilotées par des algorithmes². Les gestionnaires de *hedge funds* ont été parmi les premiers à appliquer l'apprentissage automatique pour analyser les données des

² ZHANG *et al.*, 2020.

marchés en termes de vitesse et de volume. Les *traders* manipulent maintenant un volume de données bien supérieur aux capacités de l'esprit, sur les échanges boursiers mais aussi sur les événements météorologiques et sur les variations des trafics commerciaux. Les algorithmes sont particulièrement efficaces pour déceler les tendances et les corrélations dans ces grands ensembles de données apparemment non structurées et pour proposer ensuite les options boursières correspondantes. Pour le grand public, les banques commencent d'utiliser l'IA pour détecter les mouvements inhabituels dans les comptes de leurs clients et les alerter de possibles détournements de leurs avoirs. Les banques peuvent aussi analyser les comportements de leurs clients et leur offrir plus de services et d'offres financières tout en réduisant leurs risques.

L'IA peut aussi être utilisée pour détecter et suivre le commerce illégal d'animaux et de produits animaux³. Le trafic d'animaux protégés est une menace sévère pour la biodiversité, il est responsable de 60 % du déclin de la population d'éléphants entre 2009 et 2014 et génère plusieurs milliards d'euros de bénéfice pour les trafiquants. Les recherches dans ce domaine sont à la croisée de disciplines très diverses, du développement international aux sciences économiques en passant par les sciences de l'information, la psychologie et les sciences de l'ingénieur. En combinant toutes ces compétences, des outils ont été développés pour la surveillance des ventes et pour mettre à jour l'ensemble des moteurs du trafic. Ils surveillent les forums sur internet et l'ensemble des canaux de communication électronique. Les chercheurs peuvent ainsi avoir un regard neuf et comprendre comment de nombreuses communautés bénéficient de cette manne illégale. Ils proposent alors des méthodes pour intéresser les populations à la vie sauvage plus qu'à son trafic. L'IA permet aussi de mettre au point des caméras de

3 CUGNIERE *et al.*, 2019.

surveillance intelligentes, mises à disposition des communautés locales pour les impliquer dans le suivi de la vie sauvage ou implantées dans des environnements hostiles par exemple pour le suivi des populations de manchots.

L'IA FACE À LA THÉORISATION DES DISCIPLINES

L'IA est par essence interdisciplinaire, car elle s'applique en général au corpus de données d'une autre discipline : finance, médecine, ingénierie... Elle envisage cependant ces sciences avec un point de vue qui lui est propre. À la fin du XX^e siècle, l'informatique se basait sur une théorisation et une compréhension fine d'une discipline pour en réaliser des simulations. Ainsi, dans le domaine de la chimie, la formalisation des lois physiques et des interactions moléculaires fonde la simulation sur ordinateur des propriétés et des réactions moléculaires. En partant des propriétés électroniques des atomes, on peut déduire – et progressivement reproduire sur ordinateur – le repliement des protéines et simuler l'interaction d'un médicament avec le site de fixation d'un enzyme. Le chimiste, avec sa compréhension de la nature, reste à la manœuvre : il peut suivre le processus du logiciel qui extrapole ses connaissances. Tout a changé avec l'application de la technique des réseaux de neurones et l'apprentissage automatique en général.

Dans les méthodes de l'IA qui relèvent de l'apprentissage automatique, les données de départ et les données d'arrivée sont simplement fournies en grand nombre au logiciel⁴. Le logiciel (l'algorithme) construit alors un modèle statistique du lien entre les données de départ et d'arrivée et peut, lorsqu'on lui soumet ensuite de nouvelles données de départ, donner en retour les données d'arrivée correspondantes les plus probables. Plus le nombre d'exemples fourni initialement est élevé, meilleur sera le modèle statistique. La compréhension

4 WOOLDRIDGE, 2020.

du phénomène physique à l'œuvre n'est plus nécessaire pour prédire le résultat. La théorisation des observations et la connaissance du phénomène par un spécialiste du domaine étudié ne sont plus strictement nécessaires (même si le résultat doit être validé par un expert lorsque c'est possible). La démarche originale ainsi mise en œuvre par les informaticiens s'affranchit de l'expertise scientifique pour arriver à la solution d'un problème, en allant à l'encontre d'une approche académique classique de construction de la connaissance.

L'IA peut cependant aussi contribuer à la démarche scientifique. Pendant longtemps, les chercheurs ont cherché à extraire les informations des données de leurs expériences, au besoin avec l'aide de l'ordinateur. Plus récemment, la quantité de données recueillies et la vitesse d'acquisition ont donné une part croissante à l'analyse automatique des données : les algorithmes peuvent fouiller les données à une vitesse bien supérieure aux capacités humaines. L'IA peut aussi contribuer à l'étude au-delà de sa simple capacité à traiter un important volume d'information, en éliminant le bruit qui pollue les données ou en corrigeant les biais de mesure. Malheureusement, les données sont souvent incomplètes et les possibilités de traitement algorithmique multiples : il faut donc aussi mettre à contribution des méthodes de l'apprentissage automatique pour le choix des processus de traitement de l'information et la prise en compte de leurs coûts. À partir des données ainsi valorisées, l'IA pourrait même proposer et formuler des extensions de lois physiques déjà découvertes. Nous sommes déjà au point où des systèmes d'IA ont pu retrouver des lois comme celle de la conservation de l'énergie ou de la quantité de mouvement⁵. Une réelle contribution de l'IA à la connaissance scientifique semble possible.

5 ROBERTS, 2018.

SOCIOLOGIE : L'IA ET LE TRAVAIL

L'apprentissage automatique a donné un nouveau pouvoir aux machines, celui de menacer l'utilité sociale de l'homme, qui craint d'être remplacé dans le travail intellectuel comme il l'a été dans les travaux de force au cours du XX^e siècle. Certains ont prédit que plus de 40 % des tâches actuellement effectuées par des humains dans le cadre de leur travail pourraient être effectuées par des ordinateurs. Ils l'ont fait en utilisant l'IA, montrant ainsi que même le travail du sociologue est modifié par l'outil⁶. Depuis les années 1980, ce sont 80 % des emplois de routine qui ont disparu⁷. Aujourd'hui, même les tâches complexes comme la conduite d'un véhicule ou la rédaction juridique peuvent être prises en charge par des IA. Cette évolution du monde du travail suivra celles qui ont changé la société depuis la révolution industrielle, elle nourrit déjà le travail des sociologues.

ENCADRER JURIDIQUEMENT L'IA

Les systèmes d'IA, comme on l'a vu, sont souvent opaques. Ce sont des boîtes noires qui proposent des conclusions, des décisions, sans explications et difficiles à justifier. Si ces décisions portent sur des personnes, elles peuvent avoir des conséquences importantes. Comment sont-elles encadrées juridiquement ? Si la démarche d'un algorithme peut être expliquée, cela ne revient pas à justifier ni à légitimer une décision. Pourtant, de nombreuses décisions sont déjà prises sur la base de modèles établis par l'IA, par exemple à partir de nos comportements sur le web et dans les réseaux sociaux. Les algorithmes peuvent prédire notre orientation politique, sexuelle, notre état de santé. Ils peuvent nous proposer des ressources particulières sur internet ou même contribuer aux décisions sur notre employabilité ou sur nos demandes

6 FREY *and* OSBORNE, 2017.

7 AURTOR *et al.*, 2003.

d'assurance ou de prêt bancaire. Même les lois européennes sur la protection des données ne protègent pas efficacement l'individu. Ces lois se concentrent en général sur la collecte de données mais pas suffisamment sur l'utilisation qui en sera faite plus tard (comme l'établissement de liens entre codes postaux et types de population)⁸. Les juristes doivent travailler à l'encadrement des analyses par IA de données personnelles, même anonymisées.

Plus largement, l'éthique scientifique doit se saisir des enjeux de l'IA⁹. L'institut du futur de l'humanité a été créé à Oxford pour rassembler mathématiciens, philosophes, sociologues et ingénieurs et examiner avec recul les grandes questions sur l'avenir sans se focaliser sur des points de détail. L'IA a bien sûr été identifiée parmi ces grandes questions comme un risque important, notamment pour sa gouvernance et pour la sécurité des personnes. Le risque augmente avec l'autonomie de la technologie. La robustesse des systèmes est leur capacité à rester sûrs dans les pires situations. Cette robustesse doit s'exercer en gardant des fonctionnements cohérents avec les choix des superviseurs humains, même si la compétence technique des IA peut dépasser celle des humains. L'apprentissage automatique pourrait être orienté dans cette direction en y introduisant une fonction de récompense pour maintenir cette cohérence. Cependant, une IA compétente pourrait alors exploiter cette fonction de récompense en y trouvant des failles et détourner le mécanisme. Il faut donc maintenir une vérification humaine pour les cas difficiles et permettre une intervention humaine dans les processus à tout moment. L'institut du futur de l'humanité explore l'ensemble de ces aspects techniques. Une des options est d'utiliser des IA spécialisées avec des tâches bien définies plutôt que des

8 WACHTER, 2019, 2018.

9 SHADBOLT *et al.*, 2019.

opérateurs globaux pour assurer une tâche¹⁰. Ce travail sur la gouvernance des IA est combiné à une analyse des contextes politiques. Des rapports sont rédigés pour une communication régulière vers les politiques et les décideurs¹¹. Ce travail nécessite une collaboration avec les sciences politiques, les relations internationales, les spécialistes de la sécurité.

On voit que l'Intelligence Artificielle a engendré des envolées d'enthousiasme, mais en même temps la crainte des risques qu'elle fait naître. Pour évaluer ces risques et les limiter, la contribution de l'ensemble des disciplines est nécessaire. Les philosophes et les sociologues peuvent déceler certains dangers, les juristes les encadrer. À Oxford, c'est l'intégration de toute l'Université qui permet de traiter l'ensemble des facettes de l'Intelligence Artificielle. Dans la société en général, c'est la communication autour de l'émergence de cette nouvelle technologie, au sein de communautés interdisciplinaires larges, qui peut amener son développement équilibré pour le bien général.

BIBLIOGRAPHIE

AURTOR, David H., LEVY, Frank, MURNAN, Richard J., « The Skill Content of Recent Technological Change : An Empirical Exploration », *The Quarterly Journal of Economics* n°118, 2003, p. 1279–1333.

CUGNIERE, Laure, WRIGHT, Joss, MILNER-GULLAND, Eleanor Jane, « Evidence to Action : Research to Address Illegal Wildlife Trade », *Oryx* n° 53, 2019, p. 411.

DAFOE, Allan, « Global Politics And The Governance Of Artificial Intelligence », *Journal of International Affairs*, n° 72, 2018, p. 121–26.

DAFOE, Allan, BACHRAC, Yoram, HADFIELD, Gillian, HORVITZ, Eric, LARSON, Kate, GRAEPEL, THORE,

10 DAFOE *et al.*, 2021.

11 DAFOE, 2018.

- « Cooperative AI : Machines Must Learn to Find Common Ground », *Nature* n° 593, 2021, p. 33–36.
- FREY, Carl Benedikt, OSBORNE, Michael, « The Future of Employment : How Susceptible Are Jobs to Computerisation ? », *Technological Forecasting and Social Change*, n° 114, 2017, p. 254–80.
- ROBERTS, Stephen, *Artificial Intelligence and Machine Learning in Science*, Paris, OCDE, 2018. https://doi.org/10.1787/sti_in_outlook-2018-10-en.
- SHADBOLT, Nigel, O'HARA, Kieron, DE ROURE, David, HALL, Wendy, *The Theory and Practice of Social Machines*, Springer International Publishing, Cham, 2019.
- Times Higher Education* (THE), 'World University Rankings 2021 by Subject : Engineering', 2020. <https://www.timeshighereducation.com/world-university-rankings/2021/subject-ranking/engineering-and-it>.
- WACHTER, Sandra, « Data Protection in the Age of Big Data », *Nature Electronics*, n° 2, 2018, p. 6–7.
- WOOLDRIDGE, Michael J., *The Road to Conscious Machines : The Story of AI*, London, Pelican Book, 2020, p 416.
- ZHANG, Zihao, ZOHREN, Stefan, ROBERTS, Stephen, « Deep Learning for Portfolio Optimisation », *Paper ID 3613600. Rochester, NY : Social Science Research Network*, 2020. <https://doi.org/10.2139/ssrn.3613600>.

LA COMMUNICATION DE L'INTELLIGENCE ARTIFICIELLE COMPARÉE À L'INTELLIGENCE HUMAINE

par Mme. Mareike Wolf-Fédida

de l'Académie des Sciences, Arts et Belles-Lettres de Caen

« L'intelligence est le capital qu'on met dans la culture¹ . »

En partant de la citation de Binet, je souhaite préciser que ma contribution se tient dans le domaine de la psychologie clinique/psychopathologie et de la psychothérapie. Ainsi, c'est l'aspect de la communication et de l'interaction entre deux types d'intelligence qui m'intéresse ici. Je ne conteste pas les nombreux bénéfices tirés du fait de pouvoir profiter de calculs rapides, de logiciels ou d'aides robotiques. C'est plutôt une discussion au sujet des processus psychiques engagés avec l'intelligence artificielle que je propose ici. Ainsi, en parlant de l'intelligence artificielle, je prends modèle sur l'humain, bien qu'il y ait d'autres intelligences (animale, botanique, technologique, physique) dont je ne me sens pas compétente pour m'y référer.

Pouvoir interagir avec une autre forme d'intelligence est intéressant mais je souhaite aussi méditer sur les risques que comporte cette nouvelle opportunité.

LE CONCEPT D'INTELLIGENCE AVANT QUE CELUI-CI DEVIENNE « ARTIFICIEL »

Rappelons brièvement que le concept d'intelligence voit le jour au début du XX^e siècle à l'initiative de Binet et Simon².

1 BINET, (1889) 1997, p. 146.

2 BINET, 1922.

Celui-ci est né dans la pratique de l'enseignement, celle de l'instituteur face à la « classe unique » de l'époque. La « classe unique », longtemps pratiquée en milieu rural jusqu'à aujourd'hui, permet de regrouper des élèves d'âges différents afin de leur assurer l'instruction scolaire, et se retrouve dans des écoles à faible effectif. Bien évidemment, les élèves n'étudieront pas les matières au même niveau. Celui-ci sera déterminé par rapport à l'âge de l'enfant. Pour cela, il a fallu standardiser ce que l'on attend en connaissance (dans le contenu d'une matière) à un âge précis, appelé « âge mental ».

L'idée, très louable, a été de vouloir qu'un enfant profite au maximum du contenu de l'enseignement en le lui proposant à un niveau qui est lui accessible. Trop exigeant, il décrochera, trop facile, il s'ennuiera. Se tromper dans les maquettes d'enseignements produira donc le même résultat chez l'élève : dans un cas comme dans l'autre, il n'en profitera pas. Après avoir déterminé ce que l'on pouvait attendre comme production et assimilation du savoir à un âge donné, Binet et Simon se sont aperçus qu'il existait des variations. Certains élèves en demandent davantage et d'autres en demandent moins. La façon la plus démocratique de faire progresser l'enfant, c'était donc de lui proposer le niveau auquel, lui, l'élève, répond favorablement – en-dessous de son âge réel, si cela pouvait le faire avancer ; au-delà de son âge, si cela pouvait l'aider à garder la cadence. Bref, Binet et Simon³ se sont aperçus qu'il y avait des variables pour formaliser le concept de l'intelligence : l'âge réel par rapport à l'âge mental⁴. Il s'agit de savoir si un individu est en avance sur son

3 BINET, Th. SIMON, 1907.

4 *Ibidem*. Ces deux variables sont à diviser (l'âge mental par l'âge réel) et le résultat est à multiplier par 100, ce qui donne le calcul du quotient d'intelligence. Dans un cas normal, un enfant de 10 ans accroche au contenu d'apprentissage prévu pour 10 ans et donne la formule $10/10 = 1 \times 100 = 100$. Cette formule traduit le quotient d'intelligence (QI). Le chiffre 100 témoigne donc d'un quotient d'intelligence représentant la normalité. En dessous de 100, des difficultés intellectuelles augmentent avec le chiffre

âge ou en retard⁵. Ce calcul donne lieu à un test : le test d'intelligence. Celui-ci prend également en compte la compétence du vocabulaire et celle des mathématiques. En effet, il y a des individus qui sont plus portés sur la littérature et d'autres sur la conceptualisation mathématique. L'un ne prévaut pas sur l'autre dans ce test. Toutefois, Binet reconnaît que la faculté de verbaliser est sollicitée constamment dans presque toutes les matières. Celle-ci donne donc un avantage quand on a acquis cette compétence. Cela explique également qu'un individu mutique peut exceller en compétences mathématiques et se trouver démuné face à la parole comme dans les troubles Asperger ou du spectre autistique.

Binet⁶ reconnaît également deux faiblesses à son test, premièrement être lié à la culture occidentale dont il est issu et deuxièmement être lié au facteur du temps. Venant d'une autre culture, on accorde d'autres intérêts au monde, comme les eskimos qui décrivent de manière très variée la neige – ce qui ne serait pas valorisé dans ce test. Si on fait abstraction de la mesure du temps, tout individu peut exceller dans le test d'intelligence s'il prend tout son temps. Ces deux critiques sont très importantes pour discuter de l'intelligence et aussi de l'intelligence artificielle : savoir se réguler dans l'échange avec l'autre par rapport à une autre culture et la partager et aussi respecter le temps de la réflexion de l'autre.

L'intelligence, Binet le reconnaît lui-même⁷, est une construction. À part la mesurer, il est difficile de connaître ses contours et de la définir. D'autres concepts ont vu le jour à cette époque, qui ont eu également cette caractéristique d'avoir

qui diminue, au-dessus de 100 commence la douance, le haut potentiel, voir le génie, selon l'augmentation du quotient.

5 L'exemple : l'enfant d'un âge réel (AR) 10 ans ayant besoin d'un apprentissage, âge mental (AM) de 12 ans, $12/10 = 1,2 \times 100 = 120$ QI. L'enfant d'un âge réel de 10 ans ayant besoin d'un apprentissage de 8 ans (AM), $8/10 = 0,80 \times 100 = 80$ QI.

6 BINET, 1922.

7 *Ibidem*.

une valeur pour l'application, sans pouvoir être définis autrement : « l'inconscient » dans la psychanalyse chez Freud ou, auparavant, « l'intentionnalité » en philosophie chez Bergson. Dans les années 1960, ce sera la « cognition » dans le traitement de l'information qui conduira vers l'intelligence artificielle via l'informatique et les sciences cognitives, sans qu'on puisse situer la cognition⁸. Puis, ce sera au tour des neurosciences de présenter une approche qui ne traite pas des neurones (sinon ce serait de la neurologie) mais des sciences cognitives et de la philosophie de l'esprit.

Samuel Lajeunesse définit la cognition ainsi : « On entend par cognition, l'acquisition des savoirs sur l'environnement et cela par l'intermédiaire des phénomènes comme l'attention, la perception, la mémoire et l'image mentale »⁹.

C'est à ce moment que se confondent la cognition et l'intelligence.

La conception du test d'intelligence, au début du siècle dernier, a ouvert en psychologie l'ère des tests. S'ensuivent les tests « psychodiagnostics ». En effet, la présence ou le manque d'intelligence ne protège pas d'être affecté d'une maladie psychique ou mentale. On peut être peu doué et équilibré comme être doué et très perturbé sans que l'un aille nécessairement avec l'autre. Le quotient d'intelligence ne dit rien sur le profil psychologique d'un individu, à part sur ses performances intellectuelles.

Le test projectif de Rorschach (non figuratif avec des taches) voit le jour en 1922, et en 1935 apparaît le test projectif figuratif TAT (Test d'aperception thématique). Il y aura le test des visages de Szondi, utilisé par le FBI, tout comme le MMPI (Minnesota multiphasique inventaire de personnalité). Ce dernier prévoit une « échelle de mensonge » pour évaluer la sincérité dans les réponses. Puis, dans le monde de l'entreprise, une multitude de tests a été conçue pour des besoins

8 ANDLER, 1992.

9 LAJEUNESSE, 1998, p. 5.

spécifiques afin de ne pas recruter du personnel ne correspondant pas aux attentes. Aujourd'hui, l'abondance de tests est telle que tout magazine grand public propose, à un moment de l'année, de tester tout et n'importe quoi. Ainsi les tests sont-ils entrés dans le divertissement.

Les premiers tests continuent à rester aujourd'hui l'outil indispensable dans la formation à l'université du psychologue clinicien, par exemple, pour l'orientation scolaire ou le diagnostic psychiatrique. Rappelant aussi qu'au début du XX^e siècle, les sciences de l'éducation sont naissantes, la pédagogie aussi, tout comme l'anthropologie et l'ethnologie, et enfin, la criminologie. L'intelligence est alors un concept parlant en sciences humaines qui traverse toutes ces disciplines. En retour, cela voudra dire également que toutes celles-ci sont en rapport avec la notion d'intelligence.

LA SUBJECTIVITÉ ET L'INTELLIGENCE

L'objet de la discipline « psychologie », c'est d'étudier la subjectivité. Pour cela, il y a différentes approches pour cela qui se complètent à travers l'histoire de la psychologie débutant à la fin du XIX^e siècle. La psychologie clinique, la psychologie expérimentale, la psychologie différentielle, la psychologie du développement et ainsi de suite. Toutes proposent une approche pour éclairer le sujet dans sa subjectivité. La psychologie est une discipline qui naît laborieusement et relativement tard¹⁰ au regard des disciplines de l'Antiquité comme la médecine, la philosophie ou la théologie.

Tout d'abord, les concepts de la pensée viennent de la philosophie, à commencer par Descartes (*Discours de la méthode*, 1637)¹¹, puis la philosophie transcendantale, Kant (*La critique de la raison pure*, 1781)¹², Hegel

10 WOLF-FÉDIDA, (1998), 2007.

11 DESCARTES, (1637), 2009.

12 KANT, (1781) 2012.

(*Phénoménologie de l'esprit*, 1807)¹³, Husserl (*Méditations cartésiennes. Introduction à la phénoménologie*, 1929)¹⁴ et Heidegger (*Être et temps*, 1927)¹⁵ prenant forme sous l'intitulé de « phénoménologie ». Celle-ci a également inspiré à Jaspers le premier *Traité de psychopathologie* (1913)¹⁶. Ce traité inaugure la psychopathologie comme une discipline indépendante, celle des maladies mentales et psychiques. Tous ces auteurs ont en commun d'avancer avec la philosophie transcendantale, une méthodologie qui cherche à rendre la pensée objective, et de pouvoir concevoir le sujet dans sa subjectivité en ne se tenant à rien d'autre qu'au phénomène. Nous allons retenir de cette démarche trois notions comme essentielles pour la discussion ici, celle de l'intra et l'intersubjectivité, celle de l'empathie et celle de l'intentionnalité¹⁷.

Intra et intersubjectivité

L'intra-subjectivité est celle de l'introspection quand une personne parle à elle-même pour se rendre compte comment elle se sent (*Befinden, Befindlichkeit*, traduction littérale « se positionner, sonder où se trouver »). Cela peut se passer sous forme de la petite voix intérieure ou de réfléchir sur soi à voix haute, par exemple, se demander si on « s'est levé du bon pied et si le programme du jour s'annonce plaisant ».

L'intersubjectivité, en revanche, demande à considérer la subjectivité de l'autre, se poser en observateur entre soi et l'autre, prendre en compte les subjectivités de chacun. Pour cela, il faudra être capable de se mettre à la place de l'autre (ce qui requiert de l'empathie). C'est dans ce va-et-vient entre soi et l'autre que le raisonnement trouve son équilibre. Par

13 HEGEL, (1807) 2012.

14 HUSSERL, (1929) 1987.

15 HEIDEGGER, (1927) 1986.

16 JASPERS, (1913) 1933.

17 WOLF-FÉDIDA, 1995.

exemple, devoir prendre un rendez-vous, répondre à une invitation, évaluer comment on est perçu par l'autre et si la relation est équitable (« qui a invité la dernière fois ? », « est-ce qu'on va être perçu comme aimable et comment on a contribué à ce que ce soit le cas ou non ? »).

Un équilibre psychique et une bonne intelligence sollicitent les deux processus de l'intra et de l'intersubjectivité. Afin de pouvoir réfléchir sur soi, il est nécessaire de se placer comme un observateur par rapport à soi et de s'exercer comme dans l'intersubjectivité (« cela se passe entre moi et moi-même »). De même, si on souhaite se mettre à la place de l'autre et s'imaginer comment on pourrait être perçu, il est nécessaire d'accorder autant de capacité d'observation à son interlocuteur. Plus un individu est capable d'osciller entre intra et intersubjectivité plus il excellera dans la communication avec les autres et il réussira en anticipant sur lui-même et les autres. Il s'agit là d'une preuve d'intelligence.

En revanche, en restant trop dans l'intrasubjectivité, la personne risque un retrait social (timidité exagérée, comportement obsessionnel, repli sur soi, comportement trop narcissique, etc.). En restant trop dans l'intersubjectivité, il y a le risque de s'y perdre. À force de relativiser et de surévaluer le regard venant de l'extérieur, le Moi se dissoudra devant l'angoisse d'être persécuté (dissociation, délire, hallucination). Bref, le déséquilibre entre l'intra et l'intersubjectivité correspond souvent à des formes de psychoses. Quant à l'incapacité à s'exercer dans l'intra et l'intersubjectivité, celle-ci se retrouve plus communément dans les schizophrénies (l'émoussement affectif), dans les dépressions (le vide) et dans les addictions (la substance remplace le rapport à l'autre).

La psychopathologie est donc étroitement liée à la capacité de s'exercer à l'intra et l'intersubjectivité. Cette capacité se traduit également dans sa façon de se vivre dans le temps (temporalisation) et l'espace (spatialisation) car cette capacité de changer de place entre soi et l'autre implique le temps et l'espace dans la perception.

L'empathie

L'empathie est la traduction impossible du mot *Einfühlung* (littéralement « sentir à l'intérieur »). Tandis que la traduction souligne le *pathos* (passion, maladie) dont il faudra plutôt se distancier, on pourrait avancer la définition suivante : « Essayer de se mettre à la place de l'autre et sentir comme lui sans pour autant se confondre avec lui »¹⁸.

Le concept d'empathie est le fondement de la psychothérapie et de toute consultation psychologique. Mise en avant par le psychiatre suisse, L. Binswanger¹⁹, l'empathie a été formalisée par C. Rogers²⁰ peu après aux États-Unis. La technique rogerienne servira le premier programme conçu pour l'intelligence artificielle pour l'entretien thérapeutique « Parry/Eliza » qui sera discuté plus loin.

Le mot « empathie » est aussi simple qu'il est difficile à pratiquer. Pour s'exercer dans l'empathie, il est nécessaire de savoir écouter et d'avoir la maîtrise sur soi-même sans paraître « froid » pour autant. Il faudrait être clair sur soi-même afin de pouvoir être empathique, sinon on risque d'être toxique et manipulateur. En sciences criminelles (*forensic sciences*) ou criminologie, on déplore souvent le défaut d'empathie chez l'acteur des faits. Mais l'agresseur peut prendre des visages différents et se trouver tout à fait pourvu d'une grande capacité d'empathie afin de séduire de manière perverse (harcèlement moral, harcèlement au travail, abus sexuel) pour arriver à ses fins ou pousser au suicide. Le ressenti des victimes se décrit souvent en termes de « perte d'identité ». Incontestablement, la capacité d'empathie est liée à l'intelligence. Elle sait se servir de manière virtuose des processus d'inter et d'intrasubjectivité.

18 WOLF-FÉDIDA, 2007, p. 21-22.

19 BINSWANGER, 1970.

20 ROGERS, 2013.

L'intentionnalité

L'intentionnalité, littéralement « l'intention d'une chose » est un concept complexe qui trouve des interprétations diverses depuis le XIX^e siècle. D'abord dans les écrits de Brentano²¹ comme « un objet immanent », « contenant quelque chose », puis, chez Husserl²², « un objet mental » se révélant à travers le vécu, pour devenir, ensuite, chez Heidegger²³ une façon d'être-au-monde. L'intentionnalité, dans un rapport flou au romantisme allemand, part de l'idée d'une force signifiante, d'une intelligence propre à toute production psychique. Ne pouvant se soustraire à l'existence, l'être est toujours contraint de se positionner et de signifier. Exister signifie, selon l'étymologie *existere*, d'être projeté en avant, en dehors de soi.

Que ce soit dans le concept du symptôme névrotique chez Freud²⁴ (dont le lapsus, l'acte manqué, le rêve et l'idée subite), il témoigne d'une activité inconsciente, ou la conception du phénomène dans une thérapie daseinsanalytique, style binswangérienne²⁵, en philosophie (Merleau-Ponty, *Phénoménologie de la perception*)²⁶, en linguistique dans le clivage du signifié avec le signifiant (de Saussure, Benveniste, Jakobson)²⁷, dans la psychanalyse lacanienne (« l'inconscient est structuré comme un langage »²⁸), dans le structuralisme ethnologique et anthropologique (Lévi-Strauss, Barthes)²⁹, tous ont ceci en commun qu'ils partent de l'idée d'une force propre du symptôme, du signe, du phénomène ou de l'élément structural.

21 BRENTANO, (1874) 2008.

22 *Op. cit.*

23 *Op. cit.*

24 FREUD, 1901.

25 BINSWANGER, 1971.

26 MERLEAU-PONTY, 1945.

27 SAUSSURE DE, 1987. É. BENVENISTE, 1966. R JAKOBSON, 1984.

28 LACAN.

29 LÉVI-STRAUS, (1958) 2006. R BARTHES, (1970) 2015.

Toutes ces théories, dont il est impossible de rendre compte ici, ont ceci en commun de penser que la parole échappe toujours à celui qui la prononce. Celle-ci dit quelque chose de lui à son insu. Autrement dit, en faisant on dit, et en disant on fait³⁰ aussi. Il faudra toujours prendre en compte ce principe. Il y a différentes manières de prendre en compte la multiplicité et la surdétermination du sens. Par exemple, en psychanalyse, c'est une raison pour travailler avec le silence. Une interprétation ne peut être entendue dans la mesure où le psychanalyste se tait la plupart du temps.

Autrement dit, l'intelligence augmente quand elle sait témoigner de sa capacité à anticiper sur les points qui pourraient lui échapper. Cela devient une question d'éthique et servira à éviter des erreurs car chaque intervention, chaque mot, pourrait se constituer « auteur des faits ». Il s'agit donc d'une conscience aiguë sur les faits, les dires et les conséquences sur soi et les autres.

Ces trois variables (intra et intersubjectivité, empathie et intentionnalité) ici brièvement résumées, entrent dans ce qui est appelé trivialement « la communication ».

LA COMMUNICATION DANS LA CONSULTATION PSYCHOLOGIQUE

On voit, de plus en plus, chez les robots de compagnie familiale qu'ils communiquent en associant des gestes au contenu des mots. En effet, la communication se distingue entre la communication verbale et la communication non verbale. Il ne faudrait pas que l'intelligence artificielle introduise une maladie psychique. Loin d'être uniquement un véhicule d'information, la communication régule la relation. Watzlawick *et al*³¹. ont examiné les stratégies de communication dans les années 1960.

30 AUSTIN 1991.

31 WATZLAWICK, BEAVIN, JACKSON, (1967) 1972.

Ainsi, Bateson³² a démontré qu'en émettant des messages contraires sur le plan verbal et non verbal, ce double message (*double bind*) peut avoir un effet aliénant. L'interlocuteur ne sachant pas à quel signe se fier, cela peut le mener à la schizophrénie. En tout cas, émettre des messages contradictoires est une source de stress. Cette analyse de communication est une clé importante dans la thérapie familiale de type systémique.

Résumons juste quelques spécificités dans la communication psychothérapeutique afin de prendre la mesure que parler dans le but d'apporter de l'aide est loin d'être simple. Toutes les techniques de la parole nécessitent que leur interlocuteur ait envie d'aller à la recherche de lui-même. Par exemple, dans une psychanalyse, l'analyste travaillera avec le silence. Il est plus important que le patient s'écoute parler et quand il arrive que l'analyste dise quelque chose, que ce soit reçu, effectivement, comme une interprétation. Toute prise de parole est à concevoir comme une tentative plus ou moins ratée ou réussie de s'exprimer. Pour parler vrai, il va falloir accepter de s'écouter et apprendre à considérer l'autre en soi.

Puis, dans une psychothérapie d'orientation phénoménologique, sur la base de l'échange compréhensif et empathique, le psychothérapeute fera en sorte que le patient s'intéresse à son vécu, le sien, celui des autres et celui du monde qui l'entoure. Découvrir qu'on peut être intéressant pour soi et pour les autres est riche d'enseignement et procure de la joie. Cette démarche favorise la créativité.

Ensuite, une thérapie cognitivo-comportementale (TCC) procède tout autrement en partant du principe qu'une rééducation au niveau du comportement et de la pensée favorisera une meilleure adaptation à son environnement et à la vie sociale. La satisfaction que l'individu en tirera le fera continuer d'adhérer à ces contenus et à se comporter en conséquence. C'est un travail de surface qui est censé régler

32 BATESON, JACKSON, HALEY, & WEAKLAND, 1956.

aussi les problèmes de fond. Cette conception est proche des conceptions de l'intelligence artificielle à cause de la notion de cognition.

Enfin, il y a la technique rogerienne qui sert à l'entretien en psychologie qui est à la fois empathique et compréhensive comme la démarche phénoménologique, mais celle-ci se réduit plus à une technique bien précise qui consiste à reformuler les contenus énoncés pour les présenter à son interlocuteur d'une autre manière et cela le plus neutrement possible. L'idée est qu'un émetteur ne peut pas se sentir contrarié si on lui répond avec ses propres mots sans pour autant répéter exactement ce qu'il vient de dire. C'est une technique bénéfique pour amener à réfléchir sur ce que l'on dit. C'est cette technique qui sera choisie pour un premier échange entre deux intelligences artificielles comme on le verra dans le chapitre suivant.

Même si les thérapies cognitivo-comportementales sont plus proches de l'intelligence artificielle à cause de la dominante cognitive, cette technique ne propose pas le même programme selon l'affection psychique ou psychopathologique. Il faudrait d'abord établir un diagnostic, et envisager pour cela un psychiatre artificiel, c'est certainement très périlleux.

L'EXPÉRIENCE PARRY/ELIZA D'UN ÉCHANGE PATIENT-PSYCHOTHÉRAPEUTE PAR DEUX ROBOTS D'INTELLIGENCE ARTIFICIELLE

Suite au Test de Turing³³ qui consiste à démontrer un comportement intelligent de la part d'une machine en l'incitant à un jeu d'imitation, l'idée est venue de simuler des conversations humaines. Plus particulièrement, il s'agit de placer ces conversations dans un but de soin comme dans une psychothérapie. Alors celles-ci se déroulent entre une machine

33 TURING, « Computing Machinery and Intelligence », *Mind*, LIX, 236, 1950, p. 433-460.

produisant des traits conversationnels d'une maladie psychique et mentale et d'une autre machine prenant la place du psychothérapeute.

« La première consultation en ligne a eu lieu avec deux intelligences artificielles sur ARPANet en 1972. Elle a mis en présence PARRY, un programme imitant un paranoïaque mis en communication avec ELIZA, Intelligence artificielle de Joseph Weizenbaum qui imite un psychothérapeute rogerien. PARRY était exécuté depuis le SRI lab de Standford et ELIZA depuis l'UCLA³⁴. »

ELIZA est donc la ou le psychothérapeute rogerien(ne) (l'intelligence artificielle peut-elle être sexuée ?) qui doit reformuler les contenus au plus près des dires du patient. La maladie du patient est ici la paranoïa qui se caractérise par une méfiance exagérée et des idées de persécution pouvant aller jusqu'au délire. Il faudra remarquer que la technique rogerienne est indiquée dans beaucoup de domaines sauf dans la paranoïa selon mes observations. C'était donc dès le départ une drôle d'idée de la part des concepteurs de ce « couple psychothérapeutique » (patient-thérapeute), sans doute peu renseigné par un psychothérapeute expérimenté. La programmation d'un patient névrosé (hystérique, obsessionnel ou phobique) aurait pu avoir plus de succès. Probablement que l'idée était de prendre une « grosse » psychopathologie comme exemple (alors que les névroses et les dépressions auraient pu faire l'affaire). L'incompatibilité des échanges fera couper court et la conversation qui se terminera avec des insultes au bout de trente-trois échanges. Le lecteur est surpris de voir surgir des gros mots chez l'intelligence artificielle et sa vulgarité laisse songeur sur la programmation linguistique. On pourra craindre un potentiel de production d'agressivité chez une machine.

Toutefois, rappelons-le, dans cette expérimentation, le but n'était pas de créer une psychothérapie artificielle, mais de

34 AIMÉ, 2017.

vérifier s'il y avait une trace de pensée ou de « comportement intelligent » chez la machine. Rappelons aussi que dans, une conversation, il est nécessaire d'avoir un « but », sinon cela tournera à vide, ce qui se produira malgré tout dans cette expérimentation. En effet, le modèle d'échange dans la conversation entre PARRY et ELIZA se règle selon le principe de question/remarque et réponse. Les échanges sont très brefs et réactifs. Dès le départ, c'était mal parti :

« — *“Bonjour” dit le Docteur. Dites-moi vos problèmes. Terminez, s'il vous plaît, par un point ou un point d'interrogation.*

— Les gens m'énervent parfois.

— *Je ne suis pas sûr de vous comprendre.*

— Vous devriez faire plus attention.

— *Je suppose que vous devriez faire plus attention.*

— Vous reprenez vos propres phrases. »³⁵

On voit que les premiers échanges ne sont pas dénués d'agressivité. Demander une ponctuation pour une conversation signifie qu'elle est écrite. Personne ne se représente ses propres dires par écrit. C'est un point important de souligner que l'intelligence humaine se développe par rapport à l'oral (même dans le refus de communication ou dans la situation de handicap) ce qui n'est pas le cas de l'intelligence artificielle dont la programmation commence avec un système binaire.

À la moitié de l'échange, les choses se gâtent et la/le psychothérapeute radote et deviendra, quelques échanges plus tard, aussi paranoïaque que le patient si bien qu'à la lecture on ne sait plus qui parle.

« — *Je ne suis pas sûr de vous comprendre pleinement.*

— Vous me demandez toujours les mêmes choses, pourquoi ne pas changer de sujet ?

— *Pourquoi vous posez-vous ces questions ?*

35 https://www.researchgate.net/figure/Conversation-between-ELIZA-PARRY-12_fig1_342282938

— Je pense que je vous l’ai dit, les gens m’énervent.

— *Vous énervent ?*

— Je préférerais ne plus discuter de cela.

— *Voulez-vous discuter de votre énervement ?* »

On découvre une normalité d’esprit chez l’intelligence artificielle qui pose un problème ici : l’intelligence artificielle qui représente le paranoïaque analyse les paroles du thérapeute au lieu de se focaliser sur ses propres paroles. Le « paranoïaque » ne considère pas que les paroles du thérapeute sont en réalité les siennes. Il n’y a pas de notion de temps ou d’historique par rapport au déroulement de leur conversation. Qu’une répétition puisse être gênante n’est pas conçu. Les deux intelligences artificielles se répondent sur l’instant présent sans prendre en compte les précédentes interactions. C’est comme si chaque nouvelle phrase était la première. Ainsi, la fonction première de la thérapie centrée sur le client perd ici de son sens. Le « paranoïaque » favorise l’analyse des paroles du thérapeute sans réfléchir aux siennes. Les réponses n’ont pas de lien et les sujets de conversation changent à chaque prise de parole ce qui donne une impression du décousu.

Quand la place des échanges se confond à la fin, patient et thérapeute suivent le même modèle conversationnel, les deux interlocuteurs chercheront à se réguler l’un par rapport au dire de l’autre. Cela produit l’effet de tourner en rond. Une conversation normale ne ressemble pas à un échange de ping-pong à ce point.

La production d’insultes à la fin est surprenante. La méthode rogerienne se veut neutre et accueille la parole sans jugement. Cherchant à comprendre le client/patient, celle-ci crée des conditions pour le client/patient afin de résoudre ses problèmes par lui-même – d’où son nom *client centered therapy*, en français utilisé comme « entretien non directif ».

L’approche rogerienne considère donc que le patient possède en lui la solution à son problème et qu’il lui suffit d’accéder aux ressources nécessaires présentes en lui pour le

résoudre. Le thérapeute va ainsi écouter le patient tout en instaurant un climat sécurisant afin de favoriser la communication. Dans cette pratique, le thérapeute n'impose rien, il ne dirige pas et il n'interprète pas. Il a simplement pour but d'écouter avec empathie. Cela produira l'acceptation positive et inconditionnelle.

La qualité de « présence » à l'autre au cours d'un échange (aspect discuté dans le chapitre précédent) repose sur le fait de pouvoir créer un lieu de mémoire dans la parole échangée. L'interlocuteur est supposé se souvenir de ce que l'on vient de dire et en prendre la mesure. C'est comme si le dire pouvait laisser une trace et élaborer de manière symbolique l'histoire qui cherche à se raconter plus ou moins maladroitement. Cette dimension humaine est totalement absente dans l'échange avec l'intelligence artificielle. En revanche, celle-ci pourrait corriger les phrases prononcées dans une langue, le français, par exemple, ou faire remarquer une incohérence. Cela est certainement appréciable, mais on comprend vite que tout dépend de ce que l'on va chercher dans une conversation. La plupart du temps, les personnes ne le savent pas.

L'intérêt de cette expérimentation PARRY/ELIZA était donc de démontrer si oui ou non la machine est capable de penser suite au Test de Turing. D'autres expérimentations ont suivi pour simuler des discours paranoïaques ou schizophréniques. Les différents échanges ont été présentés à trente-trois psychiatres, 48 % d'entre eux ont identifié le caractère artificiel³⁶ et 52 % se sont fait berner. L'interprétation de ce résultat doit examiner à la fois la qualité de l'échange PARRY/ELIZA et la qualité d'écoute ou de lecture des psychiatres choisis.

D'une part, les expérimentations avaient pour but de savoir si un humain peut être leurré par un discours artificiel pensant qu'il est humain³⁷ et d'autre part, de savoir si l'intelligence

36 SAYGIN ; CICEKLI ; AKMAN, 2020, p. 463-518.

37 COLBY *et al.* 1970, p. 220.

artificielle peut être hostile. Dans ce dernier cas, les expérimentations se sont intéressées à l'intelligence émotionnelle ou esthétique³⁸. Cependant, l'intelligence émotionnelle humaine dépend d'une certaine capacité à relativiser mais aussi à s'impliquer. L'empathie et l'intentionnalité, discutées plus loin, comme faisant partie de l'intra et l'intersubjectivité, sont totalement absentes dans l'échange de PARRY/ELIZA. Comme il n'y a pas de subjectivité dans l'intelligence artificielle, il est difficile de concevoir un modèle. La subjectivité se rapporte à sa propre histoire vécue.

ELIZA a eu une fille, pour ainsi dire, en 2000, ALICE (*artificial linguistic internet chat entity*). Celle-ci n'a pas la vocation d'être psychologue mais constitue le début des *chatbot*, l'assistance par un robot quand il n'y a pas d'humain pour répondre à la *hotline*. Tout le monde connaît aujourd'hui ce type d'assistance pour des services qui ne disposent pas d'interlocuteurs humains. On confirmera que c'est souvent une source de stress et d'énervement. L'assistance par un robot genre *chatbot* exige de poser des questions avec des mots limités ce qui peut faire ressentir une certaine impuissance. Contrairement aux *faq* (*frequent asked questions* – questions qui sont fréquemment posées) l'assistance robotique ne laisse pas beaucoup de choix pour s'exprimer. En fin de compte, l'utilisateur reste dubitatif par rapport à un service qui fait perdre beaucoup de temps. L'apparente simplicité peut avoir l'effet inverse, voire produire de l'agressivité ou de la dépression chez l'être humain en réponse au sentiment d'impuissance face au vide.

En effet, l'expérimentation discutée plus haut se retrouve aujourd'hui sur de multiples services accessibles sur internet. Ainsi les algorithmes d'interaction entre intelligence artificielle et humain combinent les milliers de messages

38 FRODING, PETERSON, 2021, p. 207-214. - YUDKOWSKY E., 2008.

d'utilisateurs pour créer une conversation cohérente entre eux. Les sites de rencontre, par exemple, utilisent les services d'intelligence artificielle pour étudier la compatibilité entre deux partenaires. Bien souvent ces programmes partent du principe qu'une bonne entente n'existe qu'entre personnes qui ont les mêmes intérêts/loisirs/activité. La logique binaire d'une intelligence artificielle n'est pas forcément compatible avec la notion d'enrichissement réciproque que l'on peut trouver à travers la découverte de nos différences.

DISCUSSION DE LA PROGRESSION DE L'INTELLIGENCE ARTIFICIELLE, SES PROGRÈS ET SES LIMITES ET L'IMPACT SUR LA CULTURE

Si on reprend la thèse que « tout ce qui est calculable est calculable par la machine de Turing »³⁹, la question se pose d'une communication calculée entre le robot d'assistance et l'être humain, notamment en situation de handicap (vieillesse, motricité réduite, handicap audio-visuel, mental, etc.). Est-il possible de déterminer les situations où ce genre d'assistance pourra être utile ou, au contraire, représenter des dangers ?

Karl Marx⁴⁰ a fait remarquer que les machines pourront délivrer l'être humain de son esclavage par rapport au travail. Le modèle, bien entendu, était le travail répétitif à l'usine. De nos jours, le temps consacré à la famille se trouve de plus en plus restreint. On pourrait être tenté de laisser la robotique tenir compagnie aux enfants, par exemple, pour les aider à faire leurs devoirs. Aujourd'hui, ils connaissent presque tous le maniement des écrans ce qui se traduit par un nouveau fléau, les demandes de psychothérapies des pré-adolescents perdus dans *Facebook*, *Twitter* et d'autres applications qui ont pris le dessus sur leur vie. Ceci crée une confusion entre la vraie vie

39 CHURCH, 1936.

40 MARX, (1847) 1994, p. 80 *sq.* et Marx K., 1985, p. 321.

et le virtuel. Des problèmes d'identité s'ensuivent pouvant aller jusqu'au suicide (suite au cyberharcèlement, *bashing*).

Imaginons l'assistance d'intelligence artificielle pour les personnes âgées ou pour aider les personnes handicapées à cuisiner, à sortir, etc. Imaginons encore un cas plus simple, la sortie du chien. L'animal pour son bien-être devrait sortir au moins trois fois par jour ce qui peut poser un problème quand on n'a ni le temps ni les possibilités à cause d'un handicap, alors que toutes les conditions sont réunies pour une relation épanouie entre maître et chien. Imaginons donc le robot de demain qui sortira notre chien pour la promenade. Il est supposé connaître le code de la route, de ne pas détacher le chien et, tant qu'on y est, de ramasser les déjections canines. Dans une promenade « naturelle », le chien retrouvera un espace de liberté où il pourra renifler son quartier (sa manière de « lire le journal »). Toutefois, l'homme fera attention qu'il n'y ait pas quelque chose de nocif à manger qui traîne par terre. Il pourra stationner devant une vitrine si cela se présente pour prendre également plaisir à la promenade ou faire des courses, engager une conversation avec d'autres propriétaires de chiens. Il veillera au bon comportement de son chien. On comprendra que pour une tâche *a priori* aussi simple, l'intelligence artificielle puisse être en difficulté. La promenade rigide avec un robot risque d'empêcher le chien d'être sociable, de le rendre névrosé au point qu'il préfère, à la fin, faire ses besoins chez lui à la maison plutôt que dehors.

Comme nous l'avons vu précédemment, la communication a besoin de participation subjective. Pour cela, il est important de se mettre à la portée de l'autre. Une communication dite « fluide » est celle qui reprend sans accroche la pensée de l'autre. Pour cela, ce ne sont pas exactement les mots de l'autre mais la tentative de comprendre ce qu'il aurait voulu dire. Toute conversation, aussi banale qu'elle soit, sollicite cette performance.

L'intelligence n'est pas dépourvue de formes psychopathologiques. Nombre de malades psychiques sont, au

contraire, très intelligents. En revanche, ils ne sont pas malades psychiquement parce qu'ils sont intelligents. Toutefois, leur intelligence peut s'exprimer aussi dans leur souffrance. On peut être intelligent et ne pas savoir se rendre heureux, être incapable de mener une vie amoureuse satisfaisante. On peut être intelligent et peu rusé. Par exemple, l'association MENSA permet à des « personnes à haut potentiel intellectuel de s'épanouir au sein d'une communauté chaleureuse et épanouissante »⁴¹. Ce besoin prouve qu'à l'inverse, ces personnes-là se trouvent démunies face à l'esprit « petit », à la vulgarité et au conformisme.

L'intelligence artificielle ne peut pas connaître ces dilemmes car elle n'a pas d'âge et pas d'histoire personnelle qui influencera son discours. Nous avons vu chez Binet et Simon que le concept d'intelligence ne fait sens que en comparant l'âge réel avec l'âge mental, pour une machine ce sera du non-sens, c'est la performance sans référence. En fait, le concept d'intelligence d'origine n'a plus de sens ici et ne s'applique pas dans le domaine de l'intelligence artificielle.

Aujourd'hui, la capacité de mémoire est confondue avec l'intelligence. On s'exprime en capacité de mémoire de stockage, en kilo, méga et gigaoctets, carte mémoire, stockage externe, *cloud*, etc. Des personnes hypermnésiques souffrent de cette compétence, l'être humain a besoin de pouvoir oublier, refouler et c'est cela qui le rend plus performant, plus créatif. On peut apprendre en jouant à travers une activité ludique plutôt qu'avec du bourrage de crâne, cela a été démontré aussi bien par les travaux en gestaltthérapie qu'en neuropsychologie.

Du temps des concepteurs de la notion d'intelligence, les termes avaient une tout autre signification : « La mémoire est un domaine à cultiver, l'intelligence est le capital qu'on met

41 MENSA <https://mensa-france-net>

dans la culture », a dit Binet⁴² pour sciemment bien faire la différence. C'était il y a un siècle et demi !

L'intelligence était du côté de la créativité, des biens communs, du patrimoine. Aujourd'hui l'intelligence artificielle se situe du côté de la mémoire et de la mise en réseau des productions émises et souvent en lien avec un intérêt commercial.

Revenons au principe fondamental pour définir l'intelligence. C'est le facteur temps qui est important pour apprécier l'intelligence et la rapidité avec laquelle une chose est comprise et une tâche accomplie. C'est dans la réalisation de cette tâche que l'intelligence artificielle est – partiellement – très performante. Une intelligence artificielle n'a pas besoin de savoir à quoi sert la tâche pour l'accomplir.

Je me rappelle l'éclosion des jeux d'échecs grand public dans les années 1980 – une attraction. Pour avoir une chance d'obtenir au moins un match nul, j'ai trouvé la parade du « coup fou », c'est-à-dire de jouer aux échecs de telle sorte que soudainement, au début de la partie et au bon moment, on sacrifie une pièce ce qui dégage l'échiquier et cette incohérence de la raison pour l'adversaire le mobilise pour anticiper ma stratégie, alors que je n'en avais aucune, mis à part de déstabiliser un adversaire trop fort. C'est là qu'intervient le facteur temps, la machine perdra du temps pour calculer tous les stratagèmes possibles nous laissant le temps de réfléchir à comment sauver la mise. Ainsi, prendre sur le temps de l'autre est une stratégie que la machine ne doit pas connaître, mais qui est très bien connue dans le monde des humains et que l'on retrouve dans le harcèlement au travail ou à la maison. La victime risque le *burn out* se soldant par un long arrêt de maladie, un séjour à l'hôpital psychiatrique ou le suicide. Les choses ont évolué dans l'histoire des jeux aux échecs : en 1997, l'intelligence artificielle (*Deep Blue*) a

42 BINET, (1878) 1997.

gagné contre le champion du monde russe Garry Kasparov. C'est probablement une affaire de mémoire.

J'ai récemment entendu l'histoire d'une partie d'échecs entre deux joueurs humains qui a été très mal vécue et s'est soldée par un mat (mettre le roi échec et mat) du fait que la dame avait été prise auparavant. Nous pouvons imaginer le drame psychologique œdipien qui peut surgir chez un joueur passablement névrosé quand son roi perd sa dame (craignant peut-être que l'adversaire ne lui prenne aussi sa femme), alors que l'intelligence artificielle n'a cure d'une signification symbolique. Cela ne menace pas son Moi du fait qu'elle n'en a pas.

Dans les années 1990, les robots commençaient à pouvoir converser avec les humains et cela même en plusieurs langues. Parler en plusieurs langues dans la programmation d'une machine, cela voudra dire, les parler les unes après les autres et non toutes à la fois. Cela aurait été du non-sens pour la machine. Le bilinguisme ou plurilinguisme chez l'intelligence artificielle est très différent de celui de l'être humain qui habite ses langues liées à son histoire⁴³. Pour la machine, la langue avec laquelle on s'adresse à elle est liée à une personne présente, puis à une autre présente ou alors, une nouvelle conversation s'entame avec la même personne dans une autre langue. La personne qui change de langue est prise pour une autre. Quand l'idée vous vient de vous adresser en plusieurs langues dans les mêmes phrases avec une intelligence artificielle, celle-ci a de forte chance de bugger. J'en ai fait l'expérience. Autrement dit, cela ne fera pas rire l'intelligence artificielle de lui faire une farce. Cette notion n'existe pas chez elle. Je n'ai pas encore entendu parler d'intelligence artificielle qui rit et qui pratique l'humour émanant de la spontanéité de la situation. Bien sûr, on pourra programmer une intelligence artificielle qui se mette à rire à chaque fois qu'elle ne saura

43 WOLF-FÉDIDA, 2010.

traiter une information. On dira qu'elle a un tic ou qu'elle est bête. Une intelligence bête, voyons.

Quant à la psychopathologie, on a du mal à imaginer de programmer l'intelligence artificielle avec une maladie psychique dans le but de servir les autres – non pas comme chez PARRY où le robot devait servir de patient. Je parle plutôt de concevoir une intelligence artificielle comme robot d'assistance qui serait un peu angoissée, névrosée, déprimée, bref, de quoi favoriser l'empathie chez celui qu'elle doit servir. Il faudra savoir que ce n'est pas parce que quelqu'un souffre d'un trouble psychique qu'il est moins aimable aux yeux des autres. Bien au contraire, cela peut même constituer un des facteurs qu'on cherche inconsciemment à déceler chez l'autre. Se mettre à la disposition de l'autre, c'est aussi vouloir venir en aide dans une sorte de réciprocité. Nous avons tous une capacité extraordinaire à percevoir les failles chez les autres et la complicité s'installe quand on sait se reconforter.

En général, une maladresse sera pardonnée quand l'intention a été bonne. Ce n'est pas la perfection qui rend une personne aimable, mais ses imperfections. La sympathie éprouvée est un facteur important pour permettre d'aller mieux, comment peut-on transposer cela dans une interaction artificielle ?

Actuellement, il est important de penser à un discours non genré, de pratiquer une communication inclusive. Cela ne veut pas dire pratiquer un discours rigide devenant robotique. Au contraire, l'interaction humaine fera attention à son interlocuteur. Ce n'est pas la peine de parler non genré quand il n'y a en présence que des hommes ou des femmes, à moins de considérer que plus personne ne se définit plus de manière genrée et nous sommes tous iels. Et encore, c'est propre à la langue française. Dans d'autres langues, ce sera différent comme, par exemple, en anglais, le vouvoiement n'existe pas et le tutoiement ne prend pas le genre en compte. En portugais, le genre intervient dans un simple mot comme pour dire « merci ». En allemand, il s'agit de trois sexes avec le neutre

qui sera alors à généraliser. Enfin, ce ne sera pas simple à programmer.

Il faudrait aussi se rappeler que le langage varie selon les milieux, selon les régions et, on vient de le dire, selon les pays où on parle une autre langue. Ces changements introduisent des aspects multiculturels qui font le propre des échanges animés et variés.

La culture repose donc sur toutes ces particularités qui se traduisent dans les langues et les usages. Celle-ci s'appauvrit si elle doit s'uniformiser et être standardisée. Le langage informatique a introduit à une sorte de novlangue.

Dans une publication récente, *Apocalypse cognitive*, G. Bronner⁴⁴ développe comment « l'externalisation de nombre de nos routines mentales par l'intelligence artificielle » permettra une libération de notre « temps de cerveau ». Il explique que le temps libéré pour le cerveau humain par l'intelligence artificielle constitue un trésor inestimable. Ce temps de cerveau risque d'être utilisé pour des fins commerciales par ceux qui mettent des applications à disposition. Concrètement, en utilisant internet pour une recherche, ce qui fait gagner du temps, on risque de succomber au *puta clic* (pièges à cliquer), et on aura perdu du temps au profit de quelqu'un d'autre. Bronner explique ne pas penser uniquement du négatif à propos de l'intelligence artificielle et qu'il faut prendre le mot « apocalypse » selon l'étymologie grecque signifiant « révélation »⁴⁵. Autrement dit, notre usage de l'intelligence artificielle en dira davantage sur nous et sera, dans ce sens, une chance. Il est discutable, à mon avis, si ces révélations sont vraiment une chance. La culture possède son histoire et le bon sens et la simplicité y assoit, en quelque sorte, sa tradition.

Si nous reprenons Binet avec « l'intelligence est le capital qu'on met dans la culture », une discussion intéressante

44 BRONNER, 2021, p.43 *sq.*

45 *Ibid.*, p. 190 *sq.*

s'ouvre pour savoir si on veut considérer internet comme patrimoine culturel, le *dark net* y compris, et si oui ou non on pense qu'internet est démocratique (voilà un beau sujet pour la prochaine épreuve du bac). Est-ce que les crimes d'une société, comme la cybercriminalité, font aussi partie de ce patrimoine culturel ? Cela pose un problème éthique. À l'ère où des œuvres conçues sur ordinateur ont détrôné à la vente aux enchères des tableaux peints par des grands artistes connus, le changement culturel a, en effet, eu lieu. Certaines œuvres virtuelles sont devenues un bien précieux, exprimé par des prix d'achat, plus importants que celles qui existent matériellement.

Reprenons tout simplement la citation de Binet en rajoutant « artificiel », cela donne : « l'intelligence artificielle est le capital qu'on met dans une culture artificielle. » Voici que s'ouvre un nouveau débat en philosophie de l'éthique à savoir comment positionner une culture artificielle par rapport à celle qui a été construite dans l'histoire. Le temps passant, fera-t-elle partie de la culture au même titre que celle de notre patrimoine existant ? Voire remplacera-t-elle ce que nous avons détruit ? Pourrions-nous mettre l'intelligence artificielle à contribution pour nous aider à préserver notre environnement ? Je pense, par exemple, à la pluie acide qui détruit les œuvres architecturales. Toute la question repose sur l'usage qui en est faite.

Ainsi l'être humain se trouve devant un paradoxe de vouloir consommer du divertissement sans s'investir de sa personne. Si le robot en interaction avec l'homme a fait beaucoup de progrès, tout dépend donc de son usage ou mésusage, sachant que l'intelligence œuvrera toujours en cherchant à explorer les limites. Nous devrions méditer à tout moment ce que l'intelligence artificielle nous permet de gagner en temps et en possibilités tout en nous demandant si nous n'avons pas perdu quelque chose – et appauvri notre patrimoine culturel.

Savoir débrancher et quitter les écrans est un exercice important pour se préserver, car la curiosité, en explorant de

nouvelles voies de communication, peut facilement être piégée par un usage addictif ou un attachement au détriment de la vraie vie. La responsabilité personnelle est ainsi engagée vis-à-vis de soi-même. Cela laisse songeur sur notre capacité d'être responsable vis-à-vis de nous-mêmes, à commencer avec une simple « hygiène de vie » et « de morale ». Sommes-nous suffisamment intelligents pour décrocher de ce qui est artificiel ? Faudra-t-il programmer des machines pour nous rappeler régulièrement que nous sommes dans l'artificiel ?

BIBLIOGRAPHIE

- AIMÉ, Xavier, « Artificial intelligence and psychiatry : Eliza and Parry's golden wedding » in *L'information psychiatrique*, vol. 93, 1, 2017, p. 51-56.
- ANDLER, Daniel, *Introduction aux sciences cognitives*, Paris, Gallimard, 1992.
- BATESON, Gregory, GOFFMAN, Erving, HALL, Edward T., WATZLAWICK, Paul, *La nouvelle communication*, Paris, Seuil, 1981.
- BINET, Alfred, (1909) *Les idées modernes sur les enfants*, Paris, Flammarion, 1973.
- BINET, Alfred, *L'étude expérimentale de l'intelligence*, Paris, Alfred Costes Éditeur, 1922.
- BRONNER, Gérald, *Apocalypse cognitive*, Paris, PUF, 2021.
- FRODING, Barbro, PETERSON Martin, « Friendly AI », *Ethics and Information Technology*, 2021, vol. 23, p. 207–214.
- ROGERS, Carl R., *L'approche centrée sur la personne*, Vaugeneray, Ambre, 2013.
- SAYGIN, Ayse Pinar ; CICEKLI, Ilyas ; AKMAN Valor, « Turing Test : 50 years later », *Minds and Machines*, 2000, p. 463–518.
- WATZLAWICK Paul, BEAVIN Janet H., JACKSON, Donald D., (1967) *Une logique de la communication*, Paris, Seuil, 1972.

WOLF-FÉDIDA, Mareike, (1995) *Théorie de l'action psychothérapique*, rééd. Paris, MJW Fédition, 2008.

WOLF-FÉDIDA, Mareike. (sous la dir. de), *Bilinguisme et intelligence : don de soi – perte de soi*, Paris, MJW Fédition, 2017.

YUDKOWSKI, Eliezer, « Artificial Intelligence as a Positive and Negative Factor in Global Risk », in *Global Catastrophic Risks*, New York, Oxford University Press, p. 308-345.

IN MEMORIAM JEAN GUGLIELMI (1933/2022)

par M. Thierry GEFROTIN

de l'Académie des Sciences, Arts et Belles-Lettres de Caen

« Apprendre et penser resteront-ils l'apanage du vivant ? », c'est le titre choisi par Jean Guglielmi, membre de l'Académie des Sciences, Arts et Belles-Lettres de Caen, pour sa contribution à la Conférence nationale des académies organisée à Caen sur le thème « Intelligence artificielle, neurosciences, biotechnologies : vers quel humanisme ? ».

Pourquoi ce professeur émérite de l'université, cet homme de passion, admiré et respecté, ayant consacré toute sa vie professionnelle à l'éducation et à la psychopédagogie avait-il souhaité travailler sur un thème aussi éloigné en apparence de son univers ? En apparence seulement, car l'humain et son avenir sont au cœur du thème choisi par la Conférence nationale des Académies.

« Apprendre et penser resteront-ils l'apanage du vivant ? ». Aucune réponse définitive n'est apportée dans le texte inachevé de Jean Guglielmi qui fait suite à cet hommage. Mais la connaissance de sa vie nous aide à mieux appréhender sa pensée et ce qu'il souhaitait pour l'humanité à venir.

Jean Guglielmi voit le jour le 29 septembre 1933 à Mostaganem, la plus importante ville portuaire de l'ouest algérien. L'annonce de sa naissance est faite dans la gazette de Mostaganem le 8 octobre suivant. Le secrétaire de rédaction de l'hebdomadaire, qui s'est chargé du montage de cette page, a sans le savoir fait un clin d'œil au nouveau-né. En regard de ce carnet rose, on peut en effet lire un article de saison, un marronnier comme disent les journalistes, avec ce titre : « La rentrée des classes ». Titre prémonitoire, car l'éducation, les enfants, l'enseignement seront au cœur de la vie professionnelle de Jean Guglielmi.

En 1950, à 17 ans, l'adolescent entre à l'école normale d'instituteur de Bouzaréa, dans la banlieue d'Alger dont la mission est de transformer en trois ans des adolescents plus ou moins sages en jeunes hommes soucieux de donner le bon exemple, d'enseigner la morale laïque, d'élever avec amour les enfants qui leur seront bientôt confiés. « Vaste programme » ! Pendant ses études à l'école normale, Jean obtient son baccalauréat et un certificat d'aptitude pédagogique. Puis ce sera l'université d'Alger, où il décroche une licence ès lettres. La psychologie (générale, sociale, comparée) est au cœur de l'enseignement qu'il reçoit. La psychologie de l'enfant également.

Après l'université d'Alger, l'Institut de psychologie de Paris reconnaît ses mérites. Puis c'est le grand départ : Jean Guglielmi quitte définitivement l'Algérie de sa jeunesse ; direction Caen. Il s'inscrit à la faculté des lettres et des sciences humaines et de nouveaux diplômes couronnent son cursus universitaire, jusqu'en 1978 où il devient docteur ès lettres.

Ayant survolé le brillant parcours de Jean Guglielmi, ceux qui ne le connaissent pas pourraient penser : que d'études, que de théorie ! Que d'études, oui, que de théorie, non ! Car tout ce savoir acquis durant toutes ces années de formation est très vite mis en œuvre. D'abord instituteur stagiaire au sortir de l'école normale, pendant une année scolaire, il est titularisé l'année suivante. Deux années et demi de service militaire, entre 1959 et 1961, et le voici à nouveau face à des élèves. Il est instituteur adjoint à Birkhadem dans la banlieue d'Alger. Il y reste à peine une année scolaire.

Par décision ministérielle, il est rattaché au département du Calvados. À la rentrée 1962, il fait ses premiers pas d'instituteur adjoint au Chemin-Vert, un tout jeune quartier en pleine construction à la périphérie de Caen. On imagine l'esprit de pionnier qui souffle alors dans cette toute nouvelle école ; comme on peut imaginer l'enthousiasme de ce jeune instit' de 29 ans. Un an plus tard, c'est un autre quartier de Caen, en plein développement lui aussi, que Jean Guglielmi

découvre. Le voici psychologue scolaire au calvaire Saint-Pierre ! De nombreux pieds-noirs viennent d'y arriver. Les accents oranais et algérois font chanter les mots dans la cour de récréation normande de l'école de la rue des Vikings. Ces cris joyeux doivent lui rappeler Mostaganem, Bouzaréa et Birkhadem, l'Algérie de son enfance.

D'où il est, Jean Guglielmi peut voir l'université de Caen. Il la voit si bien qu'il s'y retrouve très vite comme assistant de psychologie. Il vient d'accomplir le grand saut. Il est accueilli au sein d'une prestigieuse institution qui rayonne grâce notamment aux travaux de Gaston Mialaret, dont la renommée ne fait que croître. L'enseignant-chercheur en psychopédagogie est en poste à Caen depuis une dizaine d'années et ses travaux sont suivis avec attention par le ministère de l'Éducation nationale.

En 1967, l'université de Caen est d'ailleurs choisie comme université expérimentale pour mettre en place une licence et une maîtrise des sciences de l'éducation. Le timing est donc parfait pour Jean Guglielmi. L'éminent universitaire de 49 ans et le jeune assistant de 34 ans vont se rapprocher et devenir amis. À la disparition de Gaston Mialaret en 2016, Jean Guglielmi lui rend un vibrant hommage dans la revue *Les sciences de l'éducation – pour l'Ère nouvelle* : « J'ai fait partie de son séminaire de pédagogie qui a décidé de toute ma carrière dont il fut l'artisan, le mentor et la référence permanente. »

Comment parler de l'enfant et de son éducation, sujets chers à Jean Guglielmi, sans citer Montaigne :

« On ne cesse de crier à nos oreilles d'enfants, comme si l'on versait dans un entonnoir, et notre charge, ce n'est que de redire ce qu'on nous a dit. Je voudrais que le précepteur corrigeât ce point et que, d'entrée, selon la portée de l'âme qu'il a en main, il commençât à la mettre sur la piste, en lui faisant goûter les choses, les choisir et

discerner d'elle-même, en lui ouvrant quelquefois le chemin, quelquefois en le lui faisant ouvrir ¹».

Les propositions de Montaigne sont claires : épanouir l'esprit, favoriser son indépendance de jugement et rendre l'homme meilleur et plus sage. L'enfant et le maître, dont certaines qualités sont essentielles pour Montaigne : « Je voudrais aussi qu'on fût soigneux de lui choisir un conducteur qui eût plutôt la tête bien faite que bien pleine, et qu'on y requît tous les deux, mais plus les mœurs et l'entendement que la science ; et qu'il se conduisît en sa charge d'une nouvelle manière. »²

Jean Guglielmi avait la tête bien faite et bien pleine : assistant de psychologie, maître-assistant, chargé d'enseignement en sciences de l'éducation, directeur technique du laboratoire de psycho-pédagogie de l'université de Caen, directeur du réseau interuniversitaire de formation de travailleurs sociaux, chargé de mission à la formation professionnelle continue, maître de conférences en sciences de l'éducation, professeur en sciences de l'éducation à partir de 1979. Jean Guglielmi assume et assure les titres et les responsabilités de plus en plus grandes : directeur de l'UER des Sciences de l'homme à l'université de Caen, chef de projet IUFM, directeur de l'IUFM. Les responsabilités et la reconnaissance. Il est élevé au grade de commandeur dans l'ordre des Palmes académiques en 1988 et chevalier de la Légion d'honneur en 1999.

Tout au long de ce brillant parcours, Jean Guglielmi garda au plus près de son cœur et de son esprit une idée-clé qu'il résumait en une phrase « s'il fallait bien élever l'enfant, il fallait aussi le nourrir, et les apprentissages des connaissances valaient tout autant que la manière d'être, de devenir ». Pour nourrir cet enfant il faut le connaître. Assurément, Jean Guglielmi avait fait sien le précieux conseil que Jean-Jacques

1 MONTAIGNE, Essais, I, 26, « De l'institution des enfants ».

2 *Ibid.*

Rousseau donne dans la préface de l'*Émile* : « Commencez donc par mieux étudier vos élèves : car très assurément vous ne les connaissez point ».

Dans la droite ligne de ce parcours exceptionnel, on ne s'étonnera pas de voir Jean Guglielmi associé à « La Parole à l'école », un projet éducatif scolaire créé à l'initiative de l'Académie des Sciences, Arts et Belles-Lettres de Caen. Une expérience novatrice destinée à favoriser la prise de parole des élèves du CM1 à la 6^e.

Très actif au sein de l'Académie de Caen dont il était membre depuis 1998, Jean Guglielmi participait aussi aux activités de la commission administrative et à la rédaction d'articles pour les *Mémoires*. Il avait en charge la coordination des contributions à une œuvre collective consacrée aux pédagogues normands.

En filigrane de ce portrait qui nous dévoile un travailleur enthousiaste et infatigable, des mots s'imposent à nous : humanisme, tolérance, ouverture d'esprit, écoute attentive. Ces qualités me sont apparues comme une évidence dès notre première rencontre en 1994. Je me souviens, comme si c'était hier, d'une personnalité où se mêlaient empathie et convictions, un mélange de forte générosité et de détermination douce, un humaniste comme il m'a été donné d'en rencontrer fort peu au cours de ma vie. Pendant les années qui ont suivi, Jean Guglielmi ne m'a jamais donné de leçons, il m'a montré sans ostentation le chemin à suivre, celui d'une humanité éclairée. Malgré son absence, il reste pour tous ceux qui l'ont connu et aimé une lumière qui nous aide à progresser et à avancer. Jean Guglielmi n'est pas un homme du passé, il est notre futur.

Thierry Geffrotin occupe le fauteuil 36 à l'Académie des Sciences, Arts et Belles-Lettres de Caen, fauteuil occupé par Jean Guglielmi de 2017, année de sa titularisation, à son décès le 2 janvier 2022.

À titre d'hommage, l'Académie des Sciences, Arts et Belles-Lettres de Caen a souhaité que le texte inachevé de Jean Guglielmi figure dans le présent recueil.

APPRENDRE ET PENSER RESTERONT-ILS L'APANAGE DU VIVANT ?

*par M. Jean GUGLIELMI †
de l'Académie des Sciences, Arts et Belles-Lettres de Caen*

LA CAPACITÉ¹ D'APPRENDRE

Apprendre est un processus par lequel une personne acquiert des connaissances, maîtrise des habiletés ou développe des attitudes.

- Jean-Pierre Astolfi (1992) a distingué la connaissance du savoir. Sylvain Connac (2017) l'explique ainsi : « La connaissance est subjective, elle est le fruit intériorisé et global de l'expérience individuelle

- Le savoir est le résultat d'un processus d'objectivation des connaissances par un travail intellectuel autour d'un cadre théorique. Il prend la forme d'informations lorsqu'il est transmis » (p. 60).

- Quant à la compétence, elle est la capacité d'opérationnaliser des connaissances dans la pratique, c'est-à-dire, non de mobiliser les savoirs en eux-mêmes mais d'articuler des savoirs formalisés, ainsi que d'autres ressources cognitives et affectives, à l'action en situation. C'est ce que le psychologue Gérard Vergnaud appelle « la forme opératoire de la connaissance », celle qui permet d'agir en situation.

¹ Au sens propre ; le mot capacité désigne le fait d'être capable, d'avoir l'aptitude pour réaliser quelque chose : par exemple une capacité professionnelle attestée par un certificat de capacité.

APPRENDRE CONCERNE DONC UNE PERSONNE

Individu de l'espèce humaine (lorsqu'on ne veut préciser ni l'apparence, ni l'âge, ni le sexe).

Et l'apprentissage :

- . Formation professionnelle des jeunes en vue d'apprendre un métier ; temps pendant lequel on est apprenti : entrer en apprentissage.
- . Initiation par l'expérience à une activité, à une réalité : faire l'apprentissage du malheur.
- . Ensemble des processus de mémorisation mis en œuvre par l'animal ou l'homme pour élaborer ou modifier les schèmes comportementaux spécifiques sous l'influence de son environnement et de son expérience.

LA CAPACITÉ DE PENSER

- Considérer quelque chose comme vrai (ou non), comme probable (ou non) : penses-tu qu'il est sincère ?
- Considérer quelqu'un, quelque chose comme étant tels : je le pensais diplomate, mais qu'il est maladroit !
- Avoir telle ou telle opinion sur tel sujet : que pensez-vous de ce film ?
- Avoir la conviction de quelque chose, le croire : il pense avoir réussi son examen.
- Avoir l'intention de ; projeter : je pense avoir fini ce soir.
- Littéraire. Faillir, manquer : à cette nouvelle, elle pensa s'évanouir.
- Se dire intérieurement (souvent en incise) : c'est raté, pensa-t-il.
- Comprendre, déduire, etc. : quand on lui dit vacances, il pense temps perdu
- Concevoir quelque chose dans ses moindres détails avant de le réaliser : il faut bien penser l'aménagement des immeubles.

MACHINE

Une machine est un produit fini mécanique capable d'utiliser une source d'énergie communément disponible pour effectuer par elle-même, sous la conduite ou non d'un opérateur, une ou plusieurs tâches spécifiques, en exerçant un travail mécanique sur un outil, la charge à déplacer ou la matière à façonner. Une machine peut être fixe (machine-outil, machine à laver, etc.) ou mobile (locomotive, tondeuse à gazon, etc.).

Remarques :

- Tout dispositif similaire devant être supporté par une machine ou maintenu par un utilisateur est appelé *outil* ; Si sa finalité se limite à produire de l'énergie mécanique à partir d'une autre source d'énergie, on parle de *moteur* (la machine à vapeur est en réalité un moteur thermique) ;
- Inversement, une machine très polyvalente et capable d'effectuer de grands déplacements est appelé *robot* ;
- On utilise le terme *appareil* pour les dispositifs, notamment électroniques, n'ayant pas pour fonction principale de délivrer un travail mécanique (ordinateur, haut-parleur, instrument de mesure...) ;
- Dans le langage courant, on désigne souvent par le terme *machine* une motocyclette ou plus généralement un 2 roues motorisés (scooter ou cyclomoteur) ;
- On appelle également *machine* dans le langage courant un appareil électronique perfectionné, notamment un ordinateur, une calculatrice graphique ou encore une console de jeux. (*Dictionnaire Larousse*).

PROBLÉMATIQUE

Est-ce que dans la définition d'apprendre le concept de « machine » peut être substitué au concept de « personne » et dire que « Apprendre est un processus par lequel une machine acquiert des connaissances, maîtrise des habiletés ou développe des attitudes ».

- La machine deviendrait capable comme l'a avancé Jean-Pierre Astolfi de distinguer la connaissance du savoir et de préciser comme expliqué par Sylvain Connac que « La connaissance est subjective, elle est le fruit intériorisé et global de l'expérience individuelle »
- Le savoir est le résultat d'un processus d'objectivation des connaissances par un travail intellectuel autour d'un cadre théorique. Il prend la forme d'informations lorsqu'il est transmis ».

Il faut s'interroger, car se développent de plus en plus des recherches et des applications autour de *machine learning* autrement dit « apprentissage automatique » dont les algorithmes « apprennent de manière autonome et améliorent leurs performances pour effectuer une tâche. ». Il est susceptible aussi de réaliser des prédictions à partir de données et de les rendre plus performantes. Il peut aller jusqu'à retrouver des modèles dans de nouveaux groupes de données.



Quelles sont les différences entre le *Deep learning* et le *Machine learning* ?

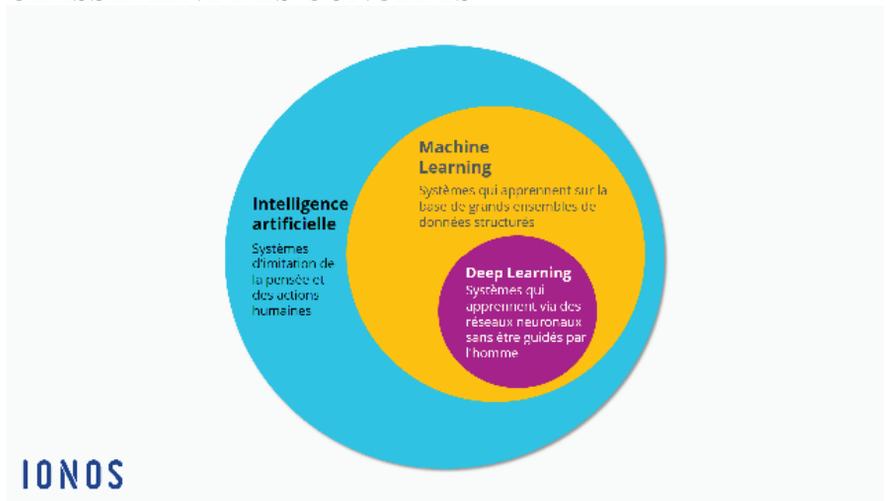
Beaucoup de personnes se méfient de l'intelligence artificielle. Elles ne comprennent pas comment les ordinateurs peuvent apprendre et prendre des décisions intelligentes. Pourtant, les principes fondamentaux de l'IA sont à la portée de tous.

La *machine learning* (apprentissage informatique) et le *Deep learning* (apprentissage profond) sont les deux concepts les plus importants qui rendent l'intelligence artificielle possible. On confond bien souvent ces deux termes, alors qu'ils désignent deux méthodes bien distinctes employées dans des champs d'application différents.

SOMMAIRE :

1. Classement des concepts
2. *Deep learning* vs *Machine learning* : quelles sont leurs différences ?
3. Des champs d'application différents

CLASSEMENT DES CONCEPTS



Machine Learning vs Deep learning : deux parties constituant de l'intelligence artificielle. Le *Deep learning* peut être envisagé comme une forme de *Machine learning*.

Le *Machine learning* et le *Deep learning* font partie de l'intelligence artificielle. Ces approches ont toutes deux pour résultat de donner aux ordinateurs la capacité de prendre des décisions intelligentes. Cependant, le *Deep learning* est une sous-catégorie du *Machine learning*, car il s'appuie sur un apprentissage sans surveillance.

Dans les deux cas, l'intelligence se limite à des utilisations spécifiques. On parle d'intelligence artificielle faible, par opposition à une intelligence artificielle forte qui serait à même, dans de nombreux domaines et circonstances, de prendre des décisions intelligentes semblables à celles de l'humain.

Les deux technologies ont impérativement besoin de disposer des grandes quantités de données, qui leur servent de base d'apprentissage. Les similitudes s'arrêtent là.

DEEP LEARNING VS MACHINE LEARNING QUELLES SONT LEURS DIFFÉRENCES ?

Le *Machine learning* (apprentissage automatique) est la technologie la plus ancienne et la plus simple. Elle s'appuie sur un algorithme qui adapte lui-même le système à partir des retours faits par l'humain. La mise en place de cette technologie implique l'existence de données organisées. Le système est ensuite alimenté par des données structurées et catégorisées lui permettant de comprendre comment classer de nouvelles données similaires. En fonction de ce classement, le système exécute ensuite les actions programmées. Il sait par exemple identifier si une photo montre un chien ou un chat et classer le document dans le dossier correspondant.

Après une première phase d'utilisation, l'algorithme est optimisé à partir des *feedbacks* du développeur, qui informent le système des classifications erronées et lui indiquent les bonnes catégories.

Le *Deep learning* (apprentissage profond) n'a pas besoin de données structurées. Le système fonctionne à partir de plusieurs couches de réseaux neuronaux, qui combinent différents algorithmes en s'inspirant du cerveau humain. Ainsi, le système est capable de travailler à partir de données non structurées.

Cette approche est particulièrement adaptée pour les tâches complexes, lorsque tous les aspects des objets à traiter ne peuvent pas être catégorisés en amont. Le système *Deep learning* identifie lui-même les caractéristiques discriminantes. Dans chaque couche, il recherche un nouveau critère spécifique de l'objet, qui sert de base pour décider de la classification retenue pour l'objet à la fin du processus.

Important : avec le *Deep learning*, le système identifie lui-même les caractéristiques discriminantes des données, sans avoir besoin d'une catégorisation préalable. Le système n'a pas besoin d'être entraîné par un développeur. Il évalue lui-même le besoin de modifier le classement ou de créer des catégories inédites en fonction des nouvelles données.

Tandis que la *Machine learning* fonctionne à partir d'une base de données contrôlable, le *Deep learning* a besoin d'un volume de données bien plus considérable. Le système doit disposer de plus de 100 millions d'entrées pour donner des résultats fiables.

Par ailleurs, la technologie nécessaire au *Deep learning* est plus sophistiquée. Elle exige plus de ressources IT et s'avère nettement plus coûteuse que le *Machine learning* : elle n'est

donc pas intéressante, du moins à l'heure actuelle, pour une utilisation de masse par les entreprises.

Synthèse des différences entre *Machine learning* et *Deep learning* :

	Machine Learning	Deep Learning
Organisation des données	Données structurées	Données non structurées
Base de données	Contrôlable	> 1 million de données
Entraînement	Entraînement par l'humain nécessaire	Système d'apprentissage autonome
Algorithme	Algorithme modifiable	Réseau neuronal d'algorithmes
Champ d'application	Actions simples de routine	Tâches complexes

DES CHAMPS D'APPLICATION DIFFÉRENTS

On peut considérer le *Machine learning* comme une technologie précurseur du *Deep learning*. Concrètement, toutes les tâches accomplies à l'aide du *Machine learning* peuvent être résolues avec le *Deep learning*. Il ne faut donc pas nécessairement opposer apprentissage profond et automatique.

Le *Deep learning* mobilise beaucoup plus de ressources et n'est donc pas un processus efficace. Les champs d'application des deux technologies sont donc en principe bien délimités : toute tâche que le *Machine learning* peut exécuter doit être traitée par cette même technique.

Pour les entreprises, utiliser ces technologies représente un avantage concurrentiel énorme, car les deux apprentissages, automatique ou profond, ne sont pas encore la norme dans le quotidien professionnel.

Champs d'application du *Machine learning*

Marketing en ligne : quelles mesures marketing apportent des résultats ? Les humains sont généralement mauvais pour

passer en revue de grandes quantités de données et fournir des estimations fiables. Dans ce cas, il vaut mieux utiliser des outils d'analyse marketing qui s'appuient sur le *Machine learning*. Ils évaluent des données définies et peuvent fournir des diagnostics fiables à propos du type de contenu capable d'aboutir à une conversion, des contenus que les clients veulent lire et des canaux marketing les plus efficaces pour conclure une vente.

Support client : les *chatbots* peuvent s'appuyer sur le *Machine learning*. Ils s'orientent en fonction des mots-clés trouvés dans la question de l'utilisateur et, par des questions pour obtenir plus d'informations ou prendre des décisions, dialoguent avec l'utilisateur jusqu'à lui apporter la réponse désirée.

Vente : ce qui fonctionne pour Netflix et Amazon est aussi idéal pour la vente. Grâce au *Machine learning*, les systèmes peuvent anticiper avec précision les produits et services qui pourraient intéresser les clients sur leur site. Ils peuvent ainsi faire des recommandations détaillées, ce qui facilite la vente avec des gammes de produits très larges ou des produits hautement personnalisables.

Informatique décisionnelle : le *Machine learning* peut aussi servir à visualiser les données importantes de l'entreprise et à rendre différentes prévisions compréhensibles pour les décideurs humains.

Champs d'application du *Deep learning*

Sécurité informatique : contrairement aux solutions basées sur le *Machine learning*, les systèmes IT et de cybersécurité qui s'appuient sur le *Deep learning* peuvent identifier aussi bien les dangers documentés que les risques jusqu'alors inconnus grâce à leur capacité à détecter les anomalies dans les

patterns connus du réseau neuronal. Le *Deep learning* démultiplie l'efficacité des mesures de sécurité.

Support client : les *chatbots* basés sur le *Deep learning* comprennent l'expression naturelle des personnes et ne sont pas limités à l'utilisation des mots-clés précis. Le dialogue est clairement plus efficace et les solutions proposées plus susceptibles de répondre à la demande.

Création de contenu : le *Deep learning* peut servir à automatiser la création de contenu. À partir d'une base de données de contenus suffisamment fournie, le système peut créer un nouveau contenu ou effectuer des traductions en autonomie.

Assistant vocal : les assistants numériques, comme Siri, Alexa ou Google, reposent sur le *Deep learning*. Les premiers assistants numériques commencent à faire leur apparition en entreprise. Les utilisateurs peuvent s'exprimer naturellement pour demander, par exemple, d'abandonner une commande, d'envoyer un mail, de créer un rapport ou de lancer une recherche.

Outre les champs d'application cités, les deux technologies s'utilisent dans de nombreux domaines du quotidien, comme la médecine, les sciences ou la mobilité.

MAIS À QUI RESTERA LE DERNIER MOT À L'HOMME À LA MACHINE ?

Pour répondre à cette question, on ne peut ignorer *Les Robots*² d'Isaac Asimov³ et son histoire du futur galactique, au travers de laquelle il aborde, de façon métaphorique, les grands bouleversements du XX^e siècle,

Ma réponse sera l'Homme, car malgré toutes les avancées, le *deep learning* n'arrivera pas, comme le dit Pascal, à cette pensée, sauf si l'homme s'en mêle :

« Que l'homme contemple donc la nature entière dans sa haute et pleine majesté, qu'il éloigne sa vue des objets bas qui l'entourent. Qu'il regarde cette éclatante lumière, mise comme une lampe éternelle pour éclairer l'univers, que la terre lui paraisse comme un point au prix du vaste tour que cet astre décrit et qu'il s'étonne de ce que ce vaste tour lui-même n'est qu'une pointe très délicate à l'égard de celui que les astres qui roulent dans le firmament embrassent.

Mais si notre vue s'arrête là, que l'imagination passe outre ; elle se lassera plutôt de concevoir, que la nature de fournir. Tout ce monde visible n'est qu'un trait imperceptible dans l'ample sein de la nature. Nulle idée n'en approche. Nous avons beau enfler nos conceptions au-delà des espaces

2 *Les Robots* (titre original : *I, Robot*) est un recueil de neuf nouvelles de science-fiction écrites par Isaac ASIMOV, publié pour la première fois par Gnome Press en 1950 et traduit en France en 1967. (Wikipédia).

3 « Isaac Asimov est né le 2 janvier 1920 à Petroviči, en Russie. À l'âge de trois ans, il émigre avec ses parents aux États-Unis et obtient la nationalité américaine cinq ans plus tard. Docteur ès sciences, passionné d'histoire, il n'a pas seulement écrit plusieurs ouvrages de vulgarisation sur des sujets très divers, et près de cinq cents romans et nouvelles ; il a créé une histoire du futur galactique, au travers de laquelle il aborde, de façon métaphorique, les grands bouleversements du XX^e siècle, influençant durablement les auteurs qui lui sont contemporains ainsi que ceux des générations suivantes. Il meurt à New-York le 6 avril 1992 » (*Encyclopedia Universalis*).

imaginables, nous n'enfantons que des atomes, au prix de la réalité des choses.

C'est une sphère dont le centre est partout, la circonférence nulle part. Enfin, c'est le plus grand caractère sensible de la toute-puissance de Dieu, que notre imagination se perde dans cette pensée.

Que l'homme, étant revenu à soi, considère ce qu'il est au prix de ce qui est ; qu'il se regarde comme égaré dans ce canton détourné de la nature ; et que de ce petit cachot où il se trouve logé, j'entends l'univers, il apprenne à estimer la terre, les royaumes, les villes et soi-même son juste prix. Qu'est-ce qu'un homme dans l'infini ?

Mais pour lui présenter un autre prodige aussi étonnant, qu'il recherche dans ce qu'il connaît les choses les plus délicates. Qu'un ciron lui offre dans la petitesse de son corps des parties incomparablement plus petites, des jambes avec des jointures, des veines dans ses jambes, du sang dans ses veines, des humeurs dans ce sang, des gouttes dans ses humeurs, des vapeurs dans ces gouttes ; que divisant encore ces dernières choses, il épuise ses forces en ces conceptions, et que le dernier objet où il peut arriver soit maintenant celui de notre discours ; il pensera peut être que c'est là l'extrême petitesse de la nature.

Je veux lui faire voir là-dedans un abîme nouveau. Je lui veux peindre non seulement l'univers visible, mais l'immensité qu'on peut concevoir de la nature dans l'enceinte de ce raccourci d'atome. Qu'il y voie une infinité d'univers, dont chacun a son firmament, ses planètes, sa terre, en la même proportion que le monde visible, dans cette terre des animaux, et enfin des cirons dans lesquels il retrouvera ce que les premiers ont donné, et trouvant encore dans les autres la même chose sans fin et sans repos, qu'il se perde dans ces merveilles aussi étonnantes dans leur petitesse, que les autres par leur étendue, car qui n'admira que notre corps, qui tantôt n'était pas perceptible dans l'univers imperceptible lui-même dans le sein du tout, soit à présent un colosse, un monde ou plutôt un tout à l'égard du néant où l'on ne peut arriver ?

Qui se considérera de la sorte s'effrayera de soi-même et, se considérant soutenu dans la masse que la nature lui a donnée entre ces deux abîmes de l'infini et du néant, il tremblera dans la vue de ses merveilles, et je crois que sa curiosité se changeant en admiration, il sera plus disposé à les contempler en silence qu'à les rechercher avec présomption.

Car enfin qu'est-ce que l'homme dans la nature ? Un néant à l'égard de l'infini, un tout à l'égard du néant, un milieu entre rien et tout. Infiniment éloigné de comprendre les extrêmes, la fin des choses et leur principe sont pour lui invinciblement cachés dans un secret impénétrable, également incapable de voir le néant d'où il est tiré et l'infini où il est englouti. »

Blaise PASCAL, *Pensées*, 1669.

Dans tous les cas, la machine, quel que soit son degré de sophistication⁴, ne vit et ne pense que par l'esprit de l'Homme... qui, comme la langue d'Ésope, est « la meilleure et la pire des choses. » ÉSOPE.

4 Action de sophistiquer quelque chose, de subtiliser avec excès. État de quelque chose, de quelqu'un de sophistiqué, d'affecté, d'artificiel. Grande complexité technique : appareil de haute sophistication. (*Larousse*)

INTELLIGENCE ARTIFICIELLE ET HUMANISME NATUREL

*par M. Jacques BOUINEAU
de l'Académie des Belles-Lettres, Sciences et Arts de
La Rochelle*

« Plus personne ne sait vraiment où nous en sommes ni où nous allons. » Ainsi s'exprime l'un des plus grands spécialistes actuels de l'intelligence artificielle¹, dans un ouvrage devenu incontournable sur le sujet². Dès lors, qu'est-ce qu'un juriste voit quand il regarde du côté de l'IA, et plus particulièrement qu'est-ce que voit un historien du droit, qui est un juriste d'un type particulier ?

En premier lieu : qu'il n'existe pas. L'IA relève en effet d'une mouvance intellectuelle dans laquelle il convient de passer du savoir à la compétence, et donc du statut au rôle. Or le juriste possède un savoir et un statut et c'est à partir d'eux qu'il réfléchit et qu'il se situe dans l'environnement. Premier hiatus.

Ensuite : qu'il n'y comprend rien ou à peu près, car les repères, les mots, les schémas intellectuels lui sont étrangers. On lui présente un monde de pionniers innovants, qui baragouinent le *globish* et chevauchent les idées comme le héros de *Dune* le faisait du Ver des sables.

Enfin : qu'on n'attend rien de lui, qui est en général un homme posé, qui mûrit sa réflexion en ouvrant le Digeste et le Code civil. La tentation est donc grande de passer son chemin.

Mais on ne peut plus tricher dans le monde contemporain, ni se mentir, ni se voiler la face. Comme dans la Grèce

¹ Ci-après IA, que Jean-Louis DESSALLES, qualifie *Des Intelligences très artificielles*, Paris, Odile Jacob, 2019, cité par Patrick Penel, 2019, p. 341, n°9.

² LEE, 2019 (trad.), p. 224.

ancienne, chacun est tenu de se positionner par rapport aux questions débattues dans la *polis*. Et l'informaticien, nouveau flamine³ de la nouvelle divinité, s'il ne risque pas qu'on lui lance à la face que son dieu n'existe pas, doit néanmoins s'attendre aux critiques ou aux mises en garde.

N'étant pour ma part sectateur d'aucun dieu, je ne poursuis rien d'autre que le titre d'honnête homme, dont on gratifiait jadis ceux qui tentaient de percer la connaissance de la « branloire pérenne »⁴ qui s'offrait à leur observation. Et je ne prétends pas résoudre des questions pour lesquelles je ne suis en rien compétent. Peut-être, en revanche, puis-je présenter, de manière un tant soit peu ordonnée, les questions qui tournent en boucle dans mon esprit après avoir déjà un peu réfléchi au sujet. Je vais donc tenter une synthèse nouvelle à partir d'études personnelles qui me paraissent complémentaires⁵ et de nouvelles lectures.

3 Dans la religion romaine, le flamine est le prêtre attaché à une divinité particulière : les trois flamines majeurs sont attachés au culte de Jupiter, Mars et Romulus et les douze flamines mineurs sont attachés chacun au culte d'un autre dieu. Chacun jugera si l'informaticien est un flamine majeur ou mineur.

4 La citation complète de Montaigne est : « Le monde n'est qu'une branloire pérenne. Toutes choses y branlent sans cesse : la terre, les rochers du Caucase, les pyramides d'Égypte, et du branle public et du leur. La constance même n'est autre chose qu'un branle plus languissant. » Le mot « branle » est utilisé par Montaigne dans le premier sens qu'en donne Littré : « Mouvement d'un corps qui va tantôt d'un côté, tantôt de l'autre. »

5 « Antiquité et territoires connectés », article issu d'une conférence prononcée dans le cadre de la rencontre des Académies de l'Ouest (La Rochelle, 15 juin 2019 - <https://www.youtube.com/watch?v=eG75-sbb4fU>), publiée dans les *Annales de l'Académie*, t. 22 (nouv. série), 2020, p. 59-76 ; « La pensée de gouvernement », article né de la conférence inaugurale du colloque des psychiatres d'exercice privé (Biarritz, 26 septembre 2019), in *Psychiatries*, « Autorité. Perte et reconnaissance », n° 173, mars 2021, p. 6-23 ; « L'*egomet*. Réflexion sur la dimension juridique de l'homme libre », *Historia e jus*, n° 20, 2021, p. 147 (http://www.historiaetius.eu/uploads/5/9/4/8/5948821/bouineau_20.pdf), article dédié à mes cinq derniers élèves avant mon entrée dans l'éméritat ; « Citoyen et *res publica* 2.0 », in Jacques BOUINEAU

Peut-on tout d'abord partir d'une définition de l'IA ? Non, d'après Jean-Gabriel Ganascia dans l'entretien qu'il accorde à *Alternatives économiques*⁶. Donc « l'œuvre de toute machine qui apparaît intelligente peut relever du champ de l'IA ». Toutefois, il faut distinguer d'emblée entre deux niveaux : celui de l'IA « étroite »⁷, celle que l'on connaît avec les instruments connectés dont on dispose aujourd'hui et celui de l'IAG (intelligence artificielle générale⁸), qui n'est pour l'heure que de la science-fiction⁹ et où tout pouvoir appartiendrait à des robots omniscients et omnipotents¹⁰. Pour que cela devînt réalité, il faudrait des apprentissages nouveaux et surtout que les robots fussent émotionnellement intelligents. Or rien ne prouve que les progrès fulgurants accomplis ces dernières années vont continuer à ce rythme : au cours des deux dernières décennies, « aucun changement aussi radical que le *deep learning* n'est survenu... des décennies, peut-être des siècles, nous séparent encore de ce tournant majeur. Il pourrait même ne jamais advenir »¹¹.

(dir.), *Domination et Antiquité. Aspects sociaux et économiques*, Paris, L'Harmattan, 2021, p. 187-222.

6 « C'est une discipline scientifique née dans les années 1950 qui a pour objectif de simuler les différentes fonctions cognitives pour reproduire une forme d'intelligence. » *Alternatives économiques*, n° 416, octobre 2021, p. 22.

7 Ou « faible ».

8 Ou « forte ».

9 « À ce jour, l'algorithme ou le procédé de fabrication qui nous conduira à l'intelligence artificielle générale n'existe pas... l'avènement de l'IAG nécessiterait une série de découvertes scientifiques majeures, comparables au *deep learning* – *a minima*. » LEE, 2019 (trad.), p. 224.

10 Même si l'autonomie des machines existe déjà : « Depuis 2014, des "réseaux antagonistes génératifs" permettent... à des logiciels de produire de fausses vidéos qui remplacent un visage ou modifient les propos d'une personnalité (ou *deepfakes*). Or, à ces réseaux s'opposent des algorithmes chargés de les détruire... » Personne n'ayant écrit les codes pour produire ces *deepfakes*, « c'est un combat entre machines », d'après Liam Newcombe, cité par PITRON, 2021, p. 233.

11 LEE, 2019 (trad.), p. 225-226.

À cela, il convient d'ajouter le fait que le domaine n'est pas d'approche aisée. Les mots sont de moins en moins traduits et l'on se trouve en face d'un sabir qui rend compte de réalités difficiles à saisir. Dès lors, « beaucoup de gens se méfient d'une technologie qui reste difficile à comprendre et qui, souvent, ne permet pas de savoir comment elle en arrive à ses résultats »¹².

Je ne sais donc pas de quoi je vais parler, ni où nous en sommes collectivement d'un processus qui enflamme les imaginations. J'envisagerai donc l'IA comme le monde dans lequel, « grâce à la 5G... [des] milliards d'objets et autres machines [sont] appelés à cohabiter avec, mais aussi indépendamment des humains »¹³.

Le mouvement général de ma réflexion va consister à envisager simultanément la façon dont le numérique, et plus spécifiquement ce qu'on a coutume d'appeler l'IA, percute les structures en place. Pourquoi utiliser un verbe aussi fort (« percuter ») qui semble signer, d'entrée de jeu, la crispation du vieux bourgeois à qui le monde nouveau échappe, mais qui, possédant encore l'arme redoutable des mots, cherche avec malice à dissoudre dans l'acide de sa rancœur un futur qui l'a déjà conduit sur les bords de l'Achéron¹⁴ ? Parce que le dépit – ou l'enthousiasme, du reste – personnels ne sont pas de mise à l'heure où tout peut basculer.

D'abord, la numérisation de la société coïncide avec un programme politique : la dérégulation, à propos de laquelle on me fera fort justement remarquer qu'il ne s'agit pas en la matière de politique, mais d'économie. À ceci près que le référent philosophique de notre monde est devenu l'économie. Cette dérégulation se double d'une déconstruction des statuts

12 Entretien avec Brent Seales, *Sciences et avenir*, n° 897, novembre 2021, p.53.

13 PITRON, 2021, p. 231.

14 Branche aérienne du Styx – qui est une rivière souterraine –, l'Achéron est le fleuve qui mène aux Enfers dans la mythologie grecque, et désigne simplement la mort ou l'enfer.

et donc des hommes, largement définis par ceux-ci depuis que les sociétés sont établies. Le transhumanisme qui se dessine risque donc de succéder à une humanité aussi écorchée que Marsyas¹⁵ et donc vouée à une mort certaine.

Comment l'IA – et plus généralement le numérique – percute-t-elle nos structures ? La cause en est à chercher dans une réification globale : les choses sont ce qu'elles sont, mais les hommes sont désormais traités comme des choses qui possèdent un logiciel, qui peuvent se trouver en surchauffe, avec comme objectif d'atteindre des performances. L'humanité est donc constituée par des Pinocchio automatés en lutte perpétuelle les uns contre les autres. Pour éviter ce *Walhalla*¹⁶, seul le Léviathan¹⁷ informatisé peut écarter l'*hybris*¹⁸. Et c'est là le deuxième niveau de la réflexion, qui permet de comprendre qu'avec le numérique, la société passe de la verticalité à l'horizontalité : les statuts sont supprimés, comme cela a été noté, mais le pouvoir lui-même s'est transformé. À proprement parler, il ne gouverne plus : il contraint, ou il maîtrise, s'il le peut. Il est donc nécessairement perçu comme d'une autre nature et la démocratie n'a plus droit

15 Affronté à Apollon pour savoir quel instrument de la lyre (attribué à Apollon) et de la flûte (instrument de Marsyas) était supérieur à l'autre, Marsyas perdit devant ses juges et Apollon put, aux termes de l'accord préalable au combat, en disposer à sa guise : il pendit donc Marsyas à un arbre et l'écorcha vif. Et le seul juge (Midas, roi de Phrygie) qui avait soutenu le parti de Marsyas se vit pousser des oreilles d'âne.

16 Nom du Paradis des Scandinaves avant leur christianisation ; il se caractérisait par la guerre éternelle... vertu suprême aux yeux de ses sectateurs.

17 On sait que Thomas Hobbes a choisi comme titre de l'ouvrage dans lequel il dépeint un système politique où les hommes sont entièrement asservis au pouvoir un nom emprunté à l'Ancien Testament – comme on le dit souvent – et qui est en fait le nom d'un monstre marin dans la culture d'Ougarit.

18 Antithèse de l'*eunomia* (« les bonnes lois »), l'*hybris* désigne chez Solon – et plus généralement dans la culture politique des Grecs – l'arrogance face aux dieux et l'excès face aux hommes, et d'après Bailly : « tout ce qui dépasse la mesure, excès ».

de cité. Ce qui ne veut évidemment pas dire que l'on soit en dictature ; c'est autre chose, un innommé auquel on cherche encore un nom. Dès lors, et c'est le troisième temps, s'il n'y a plus de *polis* vraiment, l'homme ne fait plus société, or on sait qu'un groupe qui n'est pas un peuple est une foule¹⁹. Une foule constituée d'atomes indépendants, incapables de se structurer en molécules, constamment séparés au contraire par les messages publicitaires « personnalisés » envoyés par le *big data* sur une masse décharnée, dénervée, démusclée, dépulpée « Que le trait de la mort sans pardon a frappé » pour lui attribuer ce pastiche de Ronsard²⁰.

Ensuite, si l'on considère ensemble les cadres (c'est-à-dire le pouvoir) et la société à l'aune du numérique, on note une première déchirure dans le fait que, malgré son accroissement numérique, l'homme au pouvoir demeure limité, sauf pour lui à phagocyter « son » peuple, ce qui se pratique plus ou moins ouvertement un peu partout. Mais, de surcroît, cette limitation engendre de la frustration chez celui qui se rêve en *exemplum*²¹, comme la *Légende dorée*²² en avait vulgarisé les archétypes. Et par ailleurs, la machine a des ratés, elle *bugge*, se plante, pour la plus grande sauvegarde de ce qui reste des

19 Chez les Grecs, la foule (*ochlos*) est un amas sociologique, si j'ose utiliser une explication contemporaine, tandis que le peuple (*demos*) est une réalité politique – et les Romains diront : juridique – dans laquelle les individus possèdent des droits, ce qui est la caractéristique du citoyen. Pour qu'il y ait une cité (*polis*), il ne suffit pas qu'il y ait un amas d'hommes, il faut encore que celui-ci soit ordonné.

20 « Je n'ai plus que les os », un des derniers poèmes de Ronsard, publié avec quelques autres après sa mort par ses amis, sous le titre : *Derniers vers de Pierre de Ronsard*.

21 En latin, et dans le sens abstrait, le mot est très fort ; il désigne un modèle moral offert à l'imitation de ceux à qui il est destiné.

22 Écrit par Jacques de Voragine – de son vrai Jacques de Varage – entre 1261 et 1266, l'ouvrage reprend la technique des *Vies des hommes illustres* de Plutarque en substituant des modèles chrétiens aux archétypes antiques. Rédigé dans un latin assez quelconque, l'ouvrage a eu un succès considérable à la fin du Moyen Âge et a marqué profondément la culture chrétienne.

libertés humaines ; mais elle s'améliore et l'IA vise à la perfection.

Par ailleurs ce lieu clos qui oppose l'homme de pouvoir et Marsyas, fort éloigné de l'espace « qui est entre les hommes et qui crée du lien », comme le rêvait Hannah Arendt²³, ce lieu clos donc se pense comme « moderne », *absolutus* au sens étymologique, c'est-à-dire « délié²⁴ », depuis qu'Adam Smith a imaginé la « main invisible²⁵ » qui allait tout régler, que les physiocrates²⁶ ont été convaincus de découvrir la logique du fonctionnement du monde, préparant le lit tant à Saint-Simon, Say ou Bastiat²⁷, qu'aux *Chicago boys*²⁸, aux shootés du *smartphone* et aux accros de la 5G.

Or « selon des scénarios pessimistes, l'IA pourrait accaparer la moitié de la production mondiale d'électricité en 2040²⁹ ». Car une chose est certaine : si on ne sait pas vraiment définir l'IA, on sait que son fonctionnement n'a rien

23 Dans *Qu'est-ce que la politique ?*, Hannah Arendt introduit cette idée que la politique n'est pas un rapport agonistique entre dominants et dominés, mais un espace de rencontre dans lequel doit éclore une décision commune.

24 Ce n'est que par une évolution sémantique, qui a accompagné les excès et dérives de la monarchie absolue, que le mot en est venu – en France – à désigner un abus dans l'exercice du pouvoir. Lors de son élaboration, les juristes y voient l'indépendance du souverain par rapport à l'Église et aux seigneurs.

25 Pour Adam Smith, les hommes sont naturellement faits pour échanger et donc le marché s'équilibre de lui-même. Mais il ne faut jamais perdre de vue qu'Adam Smith considère que cet équilibre est en fait l'œuvre de Dieu.

26 Ayant vécu comme Smith au XVIII^e siècle, ils pensent avoir découvert la vérité du monde dans la nature (*physis*, en grec, d'où leur nom), ils veulent être modernes et pensent, comme nous dirions aujourd'hui, le « monde d'après » : ils rejettent la culture classique et refusent de chercher leurs modèles dans la Tradition.

27 Pères du libéralisme français, ils rompent avec la tradition française et veulent une société avec le moins d'État possible.

28 Adeptes de la quasi-disparition de l'État, ils ont largement soutenu les dictatures d'extrême droite en Amérique latine au XX^e siècle.

29 D'après Emmanuel Assié, fondateur et président de Webaxys, Data Center World, 2019, cité par PITRON, 2021, p. 251.

d'artificiel, ni d'immatériel : le numérique n'est rendu possible que grâce à des machines, faites de pièces et de morceaux réels, et grandes consommatrices d'énergie, qui ont été nourries de données « expliquées » par des cerveaux humains soucieux d'efficacité, pénétrés d'une certaine vision du monde et agencés dans une langue.

Du coup, comme disent les jeunes, l'IA est évidemment en pole position pour se substituer à tous les dieux morts ou à ceux dont on ne veut pas. Toutefois, le passage n'est pas encore fait que les nuages déjà s'accumulent : dérèglement climatique, surpopulation, virus en embuscade... Or le numérique a besoin de se projeter sans cesse, non seulement parce que l'innovation est considérée aujourd'hui comme une vertu, mais encore parce qu'il ne peut tenir que s'il peut faire rêver. Comme tous les dogmes ou toutes les idéologies. Le grand soir se limite souvent à la sortie du portable dernier cri, frappant d'obsolescence le précédent, pas si vieux que cela, et surtout les terres rares qui le composaient en partie, qui toutes ne seront pas récupérées et qui, comme leur nom l'indique, sont rares. L'horizon raisonnable de l'IA paraît donc être de manière certaine le combat généralisé fondé sur la rencontre agonistique du désir et de la rareté, comme déjà Cro-Magnon pouvait l'éprouver face à ses obsessions à lui. *Nihil novi sub sole*³⁰. Et de manière plus lointaine, le *mundus informaticus* touche à un horizon aussi embrumé que celui des promesses annoncées depuis des millénaires par les religions et plus brièvement, mais de manière plus proche de nous, par les adeptes des lendemains radieux de la libération du prolétariat.

Alors comment vous présenter les choses ? La triade « science-raison-cause », que mon maître Jean-Louis Martres³¹ opposait volontiers à la Trinité, cède avec l'informatique la place à un troisième âge, celui de l'*hybris* par

30 « Rien de nouveau sous le soleil. »

31 *Les Grilles de la pensée politique*, Nice, Libre-échange, 2017, vol. 2, p. 173.

décomposition maquillée par le métavers, qui ne peut plus déboucher sur une *eunomia* – laquelle a besoin d’êtres de raison pour exister – mais simplement sur un *Walhalla* de type nouveau, nouvelle version de la domination féodale : comme jadis, nous évoluons dans un contexte d’anarchie tempérée par le pouvoir personnel³² au sein d’une « agonistique d[’...]appétits privés »³³.

HYBRIS PAR DÉCOMPOSITION

L’IA n’est pas responsable de la décomposition de nos sociétés, mais elle en constitue un des symptômes – que je nomme *hybris konektikè*³⁴, né de l’*hybris* en embuscade dans toute construction humaine et relayé par ses thuriféraires, les *homines conectici*³⁵.

32 Pour parodier cette fois Fustel de Coulanges, qui décrivait le pouvoir des rois mérovingiens comme « une tyrannie tempérée par l’assassinat ».

33 J. BOUINEAU, « La pensée de gouvernement », *op. cit.*, 2021, p. 13.

34 Facétie à partir d’*hybris* (dont le sens a été rappelé plus haut) ; et de *konektikos*, néologisme hardi pour « connecté ».

35 Il existe bien *conexus* (lié) en latin, mais je lui ai préféré le néologisme *conectus*, plus immédiatement compréhensible pour les non-latinistes et plus facétieux pour les latinistes. Et pourquoi utiliser du pseudo-grec dans un cas et du pseudo-latin dans l’autre ? Parce, sous la Révolution française, l’utilisation de la référence à l’Antiquité revêt un sens politique et que les emprunts ne sont pas similaires : les références à la Grèce ont d’abord été majoritairement (schématiquement sous la Constituante et sous la Législative) des références à Athènes où l’on citait davantage les civilisations que les hommes – BOUINEAU, 2013, p. 4 – avant de devenir (sous la Convention) des références à Sparte où l’on citait plus volontiers les hommes que les civilisations – *ibid.*, p. 7 ; l’*hybris konektikè* correspond donc à la première approche d’un phénomène complexe, correspondant aux deux premiers temps (expression et explication) de l’idéologie chez J.-L. Martres, alors que les *homines conectici* renvoient à la troisième phase de l’idéologie (celle de la conquête du pouvoir) – cf. LAGRAULA, 2018, p. 92-104. Cette double référence à la Grèce et à Rome est donc le signe à mes yeux de la révolution conceptuelle que l’IA est en train de perpétrer.

Hybris konektikè

Frénésie. Tel semble être le mot le plus congruent pour désigner l'*hybris* dans laquelle nous vibronnons : en 2020, « chaque minute, 1,3 million de personnes se connectent sur Facebook, 4,1 millions de recherches sont effectuées sur Google, 4,7 millions de vidéos sont consultées sur YouTube et 1,1 million de dollars sont dépensés sur des sites de vente en ligne »³⁶. Ces quelques chiffres livrent deux enseignements : d'une part l'activité sur *Internet* occupe environ une personne sur 500 par minute, soit une sur quatre par jour à l'échelle planétaire, et d'autre part *Internet* est un gigantesque marché. Si l'on traduit cela en termes juridiques : une personne sur quatre par jour se soumet à une licence d'utilisation – d'un logiciel auquel elle ne peut pas toucher – pour acquérir une propriété. Nouvelle déclinaison de la propriété éminente et de la propriété utile³⁷ ?

Un des effets sociaux les plus calamiteux du monde « dématérialisé » tient d'une part dans l'impossibilité d'exercer une défense des statuts³⁸ – stigmatisés de toute manière au nom de l'adaptabilité des hommes dans un environnement liquide –, d'autre part dans l'obsession du profit, qui prime sur tout, voire qui exclut ce qui n'est pas lui³⁹.

36 PITRON, 2021, p. 141.

37 En droit féodal le seigneur possède un domaine éminent, c'est-à-dire une réalité incorporelle – exception faite de la réserve, c'est-à-dire des terres qu'il se réserve et qu'il fait exploiter pour son compte – qui donne naissance à l'exercice de droits (essentiellement, pour être simple, des prélèvements en argent ou en nature quand ce droit s'exerce envers des paysans, ou à des prestations armées quand ce droit s'exerce envers ses vassaux), lesquels s'appesantissent sur un domaine utile ; par ex. le duc d'Aquitaine, en vertu de son domaine éminent retire des profits du domaine utile des vigneron – et d'autres – et commande une armée de vassaux du Sud-Ouest, tous ces tenanciers utiles vivant au sein du domaine éminent du duc.

38 « Par quels moyens bloquera-t-on l'accès à un espace de travail qui a migré en ligne ? », s'interroge PITRON, 2021, p. 186.

39 « Les fonds algorithmiques sont paramétrés pour chercher le profit avant tout, et non pour prévenir la fonte des glaces. » PITRON, 2021, p. 243.

Tout cela est encore aggravé par l'apparition du « *design* de l'attention », encore appelé « captologie », « dont l'objectif est de rendre les consommateurs de plus en plus dépendants des outils connectés »⁴⁰ par des déclencheurs qui incitent à consulter le téléphone connecté et à s'y laisser piéger⁴¹, par les sonneries, mais aussi par les couleurs⁴².

Or, si l'individu est à ce point nié, si la logique même du marché informatisé soumet tout à la loi du profit avant tout, la réalité politique chinoise pousse le pays à protéger militairement ses infrastructures et, par la diffusion universelle de ses technologies de surveillance, à faire accroire au monde entier que son modèle politique est le meilleur⁴³, puisque la France elle-même pirate les câbles sous-marins pour y capturer l'information⁴⁴. Mais une seconde conséquence éclate si l'on considère la négation des individus⁴⁵ : l'ensemble du système de l'IA repose sur la notion de « dématérialisation », alors que le fonctionnement du système informatisé n'est qu'un vaste assemblage de tubes, de câbles et autres pièces usinées⁴⁶, vorace en énergie électrique⁴⁷, produite au tiers à l'échelle mondiale à partir d'usines thermiques, pour l'alimentation desquelles le *mountaintop removal* (littéralement le scalp des montagnes) est aujourd'hui un sport fort prisé⁴⁸. D'où la

40 *Ibid.*, p. 222.

41 Le phénomène est analysé chez EYAL, 2018 (trad.), *passim*.

42 Il paraît que le bleu est la couleur la plus prisée de la captologie, alors que le gris laisserait une bien plus grande liberté de choix au consommateur – cf PITRON, 2021, p. 228.

43 *Ibid.*, p. 294.

44 « Comment la France écoute (aussi) le monde », *L'Obs*, 25 juin 2015, cité par PITRON, 2021, p. 301.

45 La perte de sens (le métier donne un statut) mène à toutes les dérives. Sur ce point, v. LEE, 2019 (trad.), p. 266 *sq.*

46¹ Dont les antennes relais, que l'arrivée de la 6G va multiplier de manière exponentielle.

47 Pour certains auteurs, s'il s'agissait d'un État, ce serait le troisième plus grand consommateur d'électricité au monde – cf. PITRON, 2021, p. 44.

48 *Ibid.*, p. 159.

nécessité de ne pas réfléchir la tête dans le *cloud*⁴⁹, mais en ayant présent à l'esprit l'apport matériel par unité de service (MIPS⁵⁰), « c'est-à-dire la quantité de ressources nécessaires à la fabrication d'un produit ou d'un service »⁵¹ ; car si, en usant de techniques de marketing adaptées, l'IA sera bien à même de faire baisser nos consommations énergivores, comme celle de la viande par exemple⁵², il n'en ira évidemment pas de même de l'usage de l'informatique en soi⁵³.

Toutefois dans le même temps, et peut-être de manière apotropaïque, l'homme de l'anthropocène développe une formidable envie de croire⁵⁴, ce qui constitue une disposition d'esprit des plus favorables sans doute à une adhésion

49 Le « nuage » (*cloud* en américain) constitue l'ensemble des câbles et des armoires (les centres de données ou *data centers*) dans lesquels circulent ou sont stockées les informations qui apparaissent quand on « clique » sur son écran à partir de son clavier.

50 *Material input per service unit*, en américain. L'exemple extrême est celui de la puce électronique : « 32 kilos de matière pour un circuit intégré de 2 grammes », PITRON, 2021, p. 90.

51 *Ibid.*, p. 86.

52 *Ibid.*, p. 250.

53 En 2018, les experts estimaient que « la part d'émissions de gaz à effet de serre (GES) attribuable au Numérique passerait ainsi de 2,5 % en 2013 à 4 % en 2020... le Numérique devrait émettre en 2020 autant de CO₂ que l'Inde en 2015. » « LEAN ICT – Pour une sobriété numérique. Rapport du groupe de travail dirigé par Hugues Ferrebœuf pour le *think tank* The Shift Project

– octobre 2018 », <https://theshiftproject.org/wpcontent/uploads/2018/11/Rapport-final-v8-WEB.pdf> (consulté le 23 novembre 2021). Je n'ai pas trouvé la réalité du chiffre qui était avancé en 2018 pour 2020 ; il est possible qu'il soit inférieur, dans la mesure où la pandémie a fait diminuer l'émission des GES (v. par exemple le cas de l'Europe : Arthur Olivier (mis à jour le 28 X 2021 par Raphaël Chamoulard), « Infographies : les émissions de gaz à effet de serre dans l'Union européenne », <https://www.toutteleurope.eu/environnement/les-emissions-de-gaz-a-effet-de-serre-dans-lunion-europeenne/> (consulté le 27 décembre 2021).

54 « Nous aimons croire, après l'amputation de notre part divine, en l'existence d'un secret que nous demandons à des Illuminés de révéler. » MARTRES, 2017, p. 186.

inconditionnelle à l'IA, et *de facto* le terrain le plus propice au développement de l'*hybris konektikè*. Car l'homme du XXI^e siècle croit moins en son dieu qu'à une sorte de mission dont il serait porteur⁵⁵. Certains voient même là la marque de fabrique du *geek*⁵⁶ de la Silicon Valley. De manière moins lyrique, on peut certainement s'accorder sur le fait que l'IA conduit chacun à se percevoir comme le centre du monde grâce aux offres personnalisées, surtout si elles sont présentées comme un « privilège »⁵⁷. Côté chinois : la culture encourage le conformisme à la tradition, la soumission aux ordres et la nécessité de prendre une revanche sur la pauvreté et l'humiliation⁵⁸. Dès lors, très méfiants dans un premier temps pour ce qui est inconnu, les Chinois s'y lancent avec détermination dès qu'ils ont reçu le feu vert des autorités, a

55 C'est au fond la manifestation adaptée au XXI^e siècle de cette même conscience qui, au XIX^e siècle, avait guidé les pionniers du « Nouveau Monde » dans leur conquête du territoire nord-américain et conduit à la quasi-éradication des peuples autochtones.

56 Le « dingo de l'informatique » ou l'« accro 2.0 » en traduction libre ; pour les femmes, c'est *geekettes*. Le mot est d'origine germanique et apparaît à la fin du XIX^e siècle, par fusion semble-t-il du *geck* (« sot », en langue populaire) anglais et du *gek* (« fou, sot ») hollandais. À soi seule l'étymologie nous indique que nous entrons chez Ubu roi. Plus sérieusement, la notion d'*hybris* est bien contenue dans le mot même.

57 Il faut se souvenir de l'étymologie du mot (*privata lex*) pour en goûter tout le sel dans l'environnement actuel de privatisation de la puissance publique : le monde des privilèges est historiquement celui de la féodalité, sous l'empire de laquelle aucune loi générale ne subsiste, sauf à l'état embryonnaire au sein de l'Église – du moins avant la réforme grégorienne (fin XI^e siècle) – ou dans l'idée du royaume, qui ne commence à être vraiment repris en main en France qu'au XIII^e siècle. Les privilèges d'Ancien Régime ne constituent qu'une survivance du Moyen Âge et une des manifestations de l'échec de la monarchie dans sa reconstruction de la *res publica* et non pas son désir de favoriser certains aux dépens d'autres ; v. BOUINEAU, 2009, p. 163.

58¹ K.-F. LEE, 2019 (trad.), p. 59. L'épisode où il rapporte la première copie effectuée par les Chinois du savoir-faire occidental (les pendules au XVII^e siècle) mérite d'être lu attentivement, p. 62-63.

fortiori politiques⁵⁹. Qu'il s'agisse du modèle américain ou du modèle chinois, l'asymptote est donc la même : la soumission entière au Léviathan⁶⁰, même si l'antagonisme entre les deux actuels leaders en la matière n'est peut-être pas aussi éternel que Kai-Fu Lee le pense⁶¹, et même si l'arrivée dans le groupe de tête de nouveaux partenaires (comme l'Inde et ceux qui se pressent derrière elle) risque de ne pas changer grand-chose à cette réalité.

Comme au temps de la Renaissance, ceux qui sont fascinés par l'IA vivent donc une véritable euphorie à la perspective des « innovations » qu'elle promet ; mais l'époque actuelle crée un nouveau monde merveilleux par la déréalisation à laquelle conduit cette technique, contrairement à la Renaissance où la raison traquait radicalement les superstitions. Autre différence : à la Renaissance, le rire subversif tenait une grande part ; l'homme de l'informatique ne rit plus : il innove, acquiert des compétences et se doit d'être adaptable⁶². Il évolue au centre d'un monde qui l'a transformé en balle de flipper, et cela procède de plusieurs facteurs.

Le premier tient à mon sens à une substitution du temps à l'espace. Dans les sociétés traditionnelles un espace donné est soumis à deux temps différents : le temps du mythe, qui fonde les religions, les espoirs et se présente comme un réservoir de baume pour les maux présents⁶³ et le temps des hommes, où les souffrances sont allégées par l'explication mythologique de

59 *Ibid.*, p. 115.

60 Qualifié d'« informatique » (BOUINEAU, « Antiquité... », *op. cit.*, 2020, p. 64), ou d'« électronique » (PITRON, 2021, p. 22).

61 LEE, 2019 (trad.), p. 48, par exemple. Il lui manque à mon avis de s'inscrire dans une perspective dynamique : il projette dans le futur ce qui s'esquisse aujourd'hui, sans prendre suffisamment en compte le phénomène de délitement des structures.

62 Et il doit donc se détacher de son statut pour s'identifier à ses compétences, comme cela a été noté en introduction.

63 Le Paradis des chrétiens ou des musulmans, ou bien les temps à venir avec la société sans classes et sans État des communistes.

l'existence⁶⁴ ; ou l'ivresse de maîtriser son destin par l'exercice de la raison⁶⁵. L'actuelle pandémie vient même « nous rappeler la force de l'*anakyklôsis* »⁶⁶, or, avec l'IA, il n'existe plus qu'un espace (le monde) noyé dans un temps en mouvement perpétuel, qui ainsi n'appartient plus à quiconque : ni aux dieux, ni aux hommes, ni à un chef quelconque, mais à une aventure humaine à la « coconstruction »⁶⁷ de laquelle chacun est invité à participer, les yeux rivés vers un *far-tomorrow*⁶⁸ enchanté. Or cet environnement nouveau implique soit une *concordia discordantium canonum*⁶⁹, soit l'asservissement à une vérité nouvelle. La première branche de l'alternative supposerait que l'on puisse concilier des contraires, mais l'expérience vécue montre qu'il n'en est rien et que c'est une vérité nouvelle qui s'impose, caractérisée par la perte de la dimension verticale⁷⁰

64 Dictée par la volonté du dieu de l'espace concerné.

65 Car « La ligne brisée prévaut sur la droite » (MARTRES, 2017, p. 178), et alternent ou coexistent des périodes ou des lieux dominés par la raison ou par le mythe.

66 Que Bailly traduit par « révolution politique » et qui est perçu souvent comme l'éternel retour est en fait la théorie platonicienne de l'évolution et des transformations des régimes politiques.

67 Emblématique du *new public management*, le mot s'applique de manière indifférenciée à tous les secteurs d'activité.

68 « Demain lointain » (au double sens d'éloigné dans le temps réel et dans le temps imaginaire). Jeux de mots encore, formé en référence au *far-west*, qui constituait un mythe de découvertes, de conquêtes, de « monde nouveau », sans un regard pour ceux qui ont payé le prix de cette *frontier* des WASP – *White Anglo Saxon Protestant* – : les peuples autochtones. Aujourd'hui, les *geeks* et les *geekettes* sont les nouveaux *cow-boys* qui asservissent ou massacrent ceux qui ne sont pas déjà convertis.

69 Littéralement « concorde des canons discordants », sous-titre donné au Décret de Gratien, qui a réuni les textes épars et souvent contradictoires qui régissaient alors – nous sommes au milieu du XII^e siècle – le droit de l'Église, de manière à en faire un corps de doctrine, qui a constitué le socle du droit canonique.

70 « Pour acheter, et pour acheter vite, beaucoup et n'importe quoi, il faut se laisser conduire par le “coup de cœur”, comme on dit de nos jours, qui

– comme cela a été relevé dès l’introduction –, qui implique l’abandon de la notion aristotélicienne de « nature des choses » et une réification de l’être⁷¹, qui possède moins une conscience qu’un « logiciel », géré par des ressources humaines, le tout subsumé dans le contractualisme⁷². Toutefois, dès lors que le Marché est mondialisé, il n’y a plus d’extérieur, mais seulement des dissidences, que l’informatique peut fort à propos éradiquer. Il faut donc faire miroiter le transhumanisme comme libération de tous : les gouvernants étant sûrs d’y être obéis, les gouvernés, sûrs d’y être protégés. Et l’IA devient naturellement le Graal de nos sociétés. De même, l’esprit de croisade avait jadis enflammé les cœurs au lendemain de la prédication d’Urbain II⁷³. Même logique, même fonctionnement mental, même résultat.

Quant à l’espace, il est pulvérisé. Non seulement parce que dès lors qu’on considère que 250 millisecondes pour communiquer entre l’Europe et l’Asie via le canal de Suez, c’est beaucoup trop long, la communication par l’océan glacial

n’est en fait que l’émotionnel, lui-même traduction de l’*hybris* que les Grecs en général et Platon en particulier stigmatisaient car ils savaient qu’il menait à la violence. » BOUINEAU, 2020, p. 65.

71 Cf. *supra*, introduction également.

72 Qu’Alain Supiot définit comme : « l’idée selon laquelle le lien contractuel serait la forme la plus achevée du lien social et aurait vocation à se substituer partout aux impératifs unilatéraux de la loi, est une composante de l’idéologie économique, qui conçoit la société comme un amas d’individus mus par la seule vertu du calcul d’intérêts. » SUPIOT, 2005, p. 142-143 ; et un peu plus loin (p. 172) comme un « Symptôme de l’hybridation de la loi et du contrat ».

73 En 1095, le pape Urbain II a prêché la croisade pour la première fois au Concile de Clermont, avant d’effectuer un voyage dans le royaume de Francie (en fait les documents portent *Francia*, en latin ; le titre de « roi de France » apparaît pour la première fois en 1204 pour l’entrée du Capétien dans Rouen, même si G. Duby le place environ un siècle plus tôt ; quoi qu’il en soit, l’usage devient une règle définitive en 1254, et en 1095, le roi n’est roi que de Francie et pas de France, comme les *geeks* et les *geekettes* évoluent aujourd’hui dans un espace qu’ils nomment, mais qui existe surtout dans leur imaginaire).

arctique (qui abaisserait la durée à 150 millisecondes⁷⁴) apparaît évidemment comme un horizon enviable. C'est-à-dire que l'on confond tout à la fois le temps et la durée d'une part, et d'autre part que l'on asservit des zones, non seulement protégées mais hostiles, à des caprices d'enfants gâtés, incapables de se situer dans le temps et l'espace.

Ces différentes corrélations, nées du fonctionnement de l'IA, renvoient à trois questions de fond : quels sont la nature et le juste prix – notions aristotéliennes⁷⁵ – des choses ? Qu'est-ce qui fait société ? Qu'est-ce que l'être qui vit dans une société ? Sur le premier point, la lumière commence à poindre, non pas en partant d'Aristote, mais en se posant une question simple, qui apparaît de plus en plus comme la mère de toutes les autres : comment établir le « juste prix » d'*Internet* pour éviter un désastre écologique encore plus grand, induit par un usage immodéré et non régulé de l'IA, ce qui peut conduire aux pires excès, parce que l'*hybris*⁷⁶ y règne en maîtresse ? La deuxième question en induit en fait une autre : la relation individu/machine constitue-t-elle une mayonnaise (émulsion d'éléments disparates) ou une *krasis* (fusion de deux réalités autonomes) ? À coup sûr une

74 PITRON, 2021, p. 290.

75 Si l'on résume à très grands traits la pensée économique d'Aristote, on constate qu'il considère que la valeur des choses dépend de leur utilité. En tout état de cause, il condamne le désir effréné qui consiste à n'être jamais rassasié de biens matériels par le simple fait que les biens susceptibles d'être acquis sont en nombre limité. Il résume cette conviction dans la notion de « chrématistique » – qui renvoie aux biens que l'on peut acquérir par le commerce – alors que l'« économie » a trait aux biens qui sont produits par la nature et qui font l'objet d'acquisition de la part des hommes.

76 « L'intelligence artificielle va jeter de l'huile sur tous les brasiers socio-économiques du monde. » LEE, 2019 (trad.), p. 230. Certes on ne peut pas chiffrer les disparitions d'emplois à venir, mais il faut évidemment s'attendre à de très gros dommages et surtout à l'alignement des hommes sur les fluctuations du marché de l'emploi.

mayonnaise, car « on passe de la *krasis* au *chôrismos*⁷⁷, de l'union à la séparation »⁷⁸. *Homo conecticus*, que nous allons voir bientôt, est bien plus l'héritier de cet homme qu'Adam Smith voyait comme naturellement destiné à échanger, et qui évolue seul au milieu de l'espace⁷⁹, que celui que je désigne sous le nom d'*egomet*⁸⁰. Dit autrement : *homo conecticus* a été façonné par des Geppetto revus par Walt Disney, car la présence de l'homme sur terre possède une finalité, qu'il faut « libérer » – de préférence sur un mode ludique, comparable à celui des dessins animés – en rompant les amarres avec les modèles anciens, pour placer *homo conecticus* seul face au Marché. On peut bien sûr imaginer des « quotas de connexion, ou des limitations au stockage des données essentielles »⁸¹. Mais d'une part il faudrait définir une instance de régulation, et d'autre part « l'alternative est en effet impossible, parce qu'elle nierait le présent. *There is no alternative* (TINA) »⁸². Dans ces conditions, la réponse à la troisième question va de soi : L'IA mène à la déréalisation. On y quitte le terrain de la raison, c'est-à-dire de la maîtrise de soi, pour devenir un produit (on passe de l'être à la chose) captif pour la publicité qu'Elon Musk envisage d'envoyer sur terre depuis l'espace. La *krasis* qui avait donné lieu à l'*egomet* cède la place à un *chôrismos*⁸³ dans lequel l'homme est perdant d'avance car

77 Pour l'opposition de ces deux notions, v. BOUINEAU, « La pensée de gouvernement », 2021, p. 6-23.

78 *Ibid.*, « *L'egomet...* », *op. cit.*, p. 16.

79 Il faut se souvenir des mots de Margaret Thatcher (que je cite de mémoire et donc sans guillemets) : quand je sors, je ne vois pas de société, mais des hommes.

80 C'est-à-dire l'être authentique, détaché de ses attributs sociaux et familiaux, et gouverné par la raison sensible. V. BOUINEAU, *L'egomet*, 2021, *op. cit.*

81 Celles qui présentent un « caractère médical, militaire, financier » et qui sont « essentielles au bien commun », PITRON, 2021, p. 314.

82 BOUINEAU, « La pensée de gouvernement », 2021, p. 18.

83 « Dans l'arène du Web chinois, il faut tuer ou être tué. » K.-F. LEE, 2019 (trad.), p. 84. La sentence est sans appel.

séparé de son essence d'être pour être réifié. Dès lors, les destinataires de l'action politique sont bien toujours des *actores*⁸⁴, mais il n'y a plus d'*auctores*⁸⁵ ; seulement des *domini*⁸⁶, eux-mêmes soumis de plus en plus à la tyrannie du Marché que seul le *big data* régit.

La parade à tous ces dangers potentiels et déjà bien présents est simple : l'éducation. C'est par une éducation civique en phase avec la technologie⁸⁷, que l'on protégera le citoyen et que l'on pourra le rendre vraiment libre⁸⁸ non pas en l'abrutissant de billevesées soi-disant virtuelles⁸⁹, maquillées en métavers.

84 Mais cela cesserait d'être le cas si le métavers transformait l'être humain en réceptacle uniquement passif de perceptions. Si cela devait advenir, les conséquences physiques sur la santé liées à une sédentarité excessive contraindraient la médecine à intervenir et cette hypothèse demeure, à l'heure où j'écris, simplement une des faces noires de la science-fiction. Cela étant dit, le fort taux d'abstention électoral, lui bien réel dans nos démocraties, constituerait-il la première manifestation de cette mutation d'un *actor* en coryphée au mieux, *spectator* au pire ?

85 Singulièrement par des politiques publiques adaptées, *Ibid.*, p. 50.

86 L'*actor* est, pour moi, celui qui est soumis au pouvoir d'un *auctor*. La source de cette typologie repose sur l'opposition de Duguit entre gouvernés et gouvernants ; elle s'en sépare parce qu'elle doit être analysée en tenant compte de l'*egomet*, que Duguit n'envisageait pas. Sur ces notions, se reporter à BOUINEAU, « La pensée de gouvernement », 2021, p. 6.

87 Je reviendrai sur ces notions ; disons pour faire simple que l'*auctor* exerce un pouvoir de droit public alors que le *dominus* exerce un pouvoir de droit privé.

88 « Pourquoi ne pas envisager la création à l'école, d'un enseignement spécifique sur le fonctionnement, l'utilisation des réseaux sociaux, à égalité avec les cours de français, mathématique et autres disciplines ? », Grand Collège des Rites écossais, GODF, *Construire demain ? Regards maçonniques sur les solidarités post-Covid*, p. 263, <https://lbapres.godf.org/app/uploads/2020/12/GCDRE-Contribution-au-Livre-blanc.pdf> (consulté le 29 novembre 2021).

89 « Un quart d'heure d'actualité numérique devait ainsi être instauré dans toutes les écoles, permettant ainsi d'éduquer à la pratique de l'esprit critique », *Ibid.*, p. 264.

Car l'objectif consiste à maîtriser⁹⁰ un phénomène qui, pour l'heure semble sans boussole.

Homines conectici

« L'intelligence artificielle va entraîner une aggravation spectaculaire des inégalités, au sein des États et entre eux... [car l'IA] est naturellement attirée vers les situations de monopole »⁹¹. On a vu, dans le point précédent, que l'espace de l'IA était régi par l'*hybris*, car si la valeur de référence rappelée en permanence est celle du Marché qui s'équilibre, la réalité est un peu différente, pour une raison très simple : si l'État disparaît, l'espace public cède le terrain à un lieu où s'affrontent non plus des citoyens, mais une foule constituée d'atomes indépendants, comme on l'a relevé plus haut. Or une foule n'est pas un peuple, car elle n'est pas régie par le droit. Dès lors *homo conecticus* se trouve connecté dans un lieu de nulle part, au sens propre : une *utopia*, dans laquelle l'individu est une chose inféodée à un tout dont elle ne constitue qu'un atome⁹², ou l'équivalent dans le monde réel d'un avatar dans le monde virtuel⁹³. Et le métavers ne fera qu'amplifier le phénomène. Vivons-nous dès lors une rupture ou une simple évolution ?

L'accumulation d'argent dans les mains des marchands a largement précédé la révolution industrielle, mais depuis le XIII^e siècle, l'État avait profité lui aussi du retour de la

90 D'après la Banque mondiale, en 2016 « dans les pays en développement, les ménages qui possèdent un téléphone mobile sont plus nombreux que ceux qui ont accès à l'électricité ou à de l'eau salubre. » *The Shift Project*, 2018, p. 42.

91 LEE, 2019 (trad.), p. 260.

92 C'est par exemple le cas chez Campanella ; v. BOUINEAU, 2015, p. 99 *sq.*

93 Ce qui pose au demeurant la question du statut juridique du métavers, comme le souligne LAVERDET, 2022, p. 81.

finance⁹⁴. Aujourd'hui, l'État est en situation de dépérissement en raison de la néo-féodalité qui caractérise le monde néolibéral. La cause n'est pas à chercher dans la technologie en général et l'IA en particulier, mais dans le mouvement de privatisation de l'espace public qui les accompagne. Nous nous trouvons dans un temps déjà vécu du III^e au VII^e siècle et du IX^e au XI^e siècle⁹⁵, à ceci près que la privatisation atteint non seulement l'espace public, mais aussi le canal de communication qui y met les hommes en relation : *Internet*⁹⁶. Par ailleurs, le contractualisme réduit le droit à une dimension purement utilitaire, ce qui était la conception des droits « barbares »⁹⁷ et qui demeure la vision anglo-saxonne⁹⁸ où le

94 Marx avait vu une issue à cette enflure : l'accès du prolétariat au pouvoir et le dépérissement de l'État, comme le rappelle MARTRES, 2017, p. 174.

95 La crise du III^e siècle dans l'Empire romain a bouleversé l'ensemble des structures, sonnait le glas de ce qu'on appelle le Haut-Empire ; à partir de là s'ouvre d'abord le temps du Bas-Empire (IV^e – V^e), puis – mais la date de départ est discutée et certains font commencer la période dès le IV^e siècle, au moment de la fondation de Constantinople (330) – jusqu'au VII^e siècle, celle de l'Antiquité tardive, qui enregistre un fort dépérissement de l'État ; après le décès de Charlemagne (814) et surtout dans la seconde moitié du IX^e siècle et jusqu'au début du XI^e (on date *l'incastellamentum*, c'est-à-dire le moment où le pouvoir politique est tombé entre les mains des châtelains, des années 1020), c'est à l'effondrement de l'empire carolingien que l'on assiste et à la privatisation encore plus grande de l'espace public. BOUINEAU, 2004, p. 97-381.

96 « On est en train d'assister à une privatisation d'une partie d'Internet au profit de quelques acteurs sans que cela n'émeuve personne. » PITRON, 2021, p. 307.

97 Si l'on suit l'appellation ancienne ; on parle aujourd'hui plutôt de droits romano-germaniques, pour atténuer ce que l'opposition au droit romain pouvait avoir de choquant. On se situe en tout cas durant l'Antiquité tardive.

98 Alain Supiot la fait découler de la « théorie de l'*efficient breach of contract* [qui] est justifiée par l'idée d'allocation optimale des ressources sur le marché... Cette théorie trouve un écho en France chez des juristes qui professent qu'il n'y a pas de différence entre exécuter ses engagements et réparer leur violation. La confiance étant une valeur incalculable, ne compte plus ici pour rien... », SUPIOT, 2005, p. 172.

droit n'est qu'« une pure technique, neutre en soi et vide de sens »⁹⁹.

L'une des conséquences de la néo-féodalité doit nous retenir ici : celle qui a trait à la violence. L'État n'étant plus là pour assurer la paix, les acteurs privés s'en chargent au sein de leur système de valeurs, en l'occurrence, le contractualisme qui porte en soi l'*hybris* puisque le calcul de l'intérêt¹⁰⁰ se substitue à la conscience de faire société. Dans le domaine plus spécifique de la recherche en IA, on a évidemment tout à redouter de l'intrusion des capitaux privés en matière environnementale¹⁰¹, mais pas seulement, et des conflits d'intérêts qui peuvent en résulter¹⁰². Car, on ne le dira jamais assez : l'IA n'est pas la manifestation immatérielle d'un esprit planant comme l'Ariel¹⁰³ de Shakespeare, c'est le résultat d'une programmation initiale, comme le rappelle le chercheur suisse Florian Colombo¹⁰⁴. La solution retenue ne dépend donc plus d'un *auctor*, mais d'un *dominus*, ce qui est la caractéristique de la féodalité, ou d'un État dirigiste, pour ne pas dire plus dans le cas de la Chine¹⁰⁵ ; or, qu'il s'agisse des États-Unis ou de la Chine, les grandes firmes ont pignon sur

99 BOUINEAU, « *L'egomet...* », 2021, p. 16.

100 Cf. *supra*.

101 « L'industrie a confisqué le discours sur les bénéfices environnementaux d'*Internet*. Elle a essayé par tous les moyens de vendre cette idée et elle continue. » Entretien avec Françoise Berthoud (ingénieure de recherche en informatique au CNRS), 2019 et 2020, cité par PITRON, 2021, p. 43.

102 Le risque couvre tous les secteurs, puisque le partenariat public/privé incite en fait à multiplier les financements privés ; pour le domaine qui nous occupe ici, v. LAROUSSE, 2021, p. 5.

103 L'esprit qui plane au-dessus de Prospero dans *La Tempête* de Shakespeare.

104 *Sciences et avenir, op. cit.*, p. 52.

105 « Les algorithmes du *deep Learning* ne savent optimiser que ce qu'ils parviennent à "voir" via les données. L'écosystème technologique chinois, enraciné dans la réalité physique, les dote de millions d'yeux supplémentaires grâce auxquels ils dissèquent nos vies quotidiennes. » LEE, 2019 (trad.), p. 102.

rué et tiennent souvent la dragée haute au pouvoir politique¹⁰⁶.

Privé d'*auctor* dans un espace en proie à l'*hybris*, *homo conecticus* y est donc exposé à l'IA, mais qui est-il ?

Un individu qui disparaît derrière sa marchandisation¹⁰⁷, mais plus encore derrière le système de valeurs qui sous-tend l'IA. À l'image de n'importe quel dogme religieux celui de l'universelle connectique fait croire à l'homme qu'il se libère en s'aliénant : « Je suis le chemin, la Vérité et la vie » ne peut en effet libérer que ceux qui croient à la lecture chrétienne de Jésus, tout comme la libération par la 5G ne peut séduire que les convaincus, avec peut-être plus de chances de succès que dans le cas du message chrétien, car l'IA ne demande pas d'ascèse ; elle flatte au contraire l'utilisateur et l'algorithme précède ses désirs même, lui permettant de devenir omphalopsyque¹⁰⁸ sans réserve. L'IA n'a au demeurant cure de la liberté de l'homme, qu'elle ne considère pas dans sa totalité, mais dans ses différentes composantes. L'individu n'est plus un *egomet*, mais un *actor* aux masques multiples¹⁰⁹. *Homo conecticus* est nécessairement un *fellow*¹¹⁰, et donc son épiphanie doit commencer dans les pays latins¹¹¹ par un lavage de cerveau et dans le reste du monde par une colonisation orchestrée en *globish*¹¹².

106 V. le dossier d'*Alternatives économiques*, 2021, spéc. p. 27-28.

107 BOUINEAU, « Citoyen et *res publica* 2.0 », 2021, p. 198.

108 L'omphalopsyque est à l'origine le moine hésychiaste du Mont Athos, qui se concentre sur son nombril dans la solitude pour se rapprocher de Dieu.

109 C'est un peu comme si l'on définissait la Nation par la somme de ses éléments constituants, alors que c'est un corps mystique.

110 C'est-à-dire le personnage qui évolue non pas au sein d'une *res publica* – ensemble régi par le droit et composant une personne morale de droit public – mais au sein d'un *commonwealth* – ensemble d'hommes régi par le droit et constituant une communauté humaine. Sur la différence entre *egomet* et *fellow*, v. BOUINEAU, « *L'egomet...* », 2021, p. 18-24.

111 C'est-à-dire ceux qui sont de culture romano-canonique, et où l'espace public peut s'analyser comme une *res publica*.

112 BOUINEAU, *Ibid.*, p. 27.

Formaté par l'IA, *homo conecticus* ne voit qu'à travers le prisme dématérialisé¹¹³, ignorant l'aliénation qui le guette¹¹⁴ et les déchets qu'il rejette¹¹⁵. Aveugle à la réalité¹¹⁶, il offre sur un plateau à ceux qui l'entourent de leurs rets toutes les données qu'ils souhaitent et au-delà¹¹⁷, puisque même la vie la plus intime est bradée sans retenue et décomposée en autant d'éléments venant perfectionner le *big data* en quantité¹¹⁸ et en qualité¹¹⁹. En Chine, il offre également son temps¹²⁰, et l'Occident connaît de plus en plus cette absence d'alternance entre jour et nuit, travail et repos, et en fait public et privé.

113 Paré de toutes les vertus chez Mark ZUCKERBERG quand il parle de métavers, v. REDAUD, 2021, p. 4.

114 L'un des fantasmes qui entoure l'IA se tapit justement là : jusqu'à quel point l'homme va-t-il subsister ? Le « robot » ne va-t-il pas supplanter l'homme ? C'est déjà partiellement fait, puisque dans un des secteurs les plus sensibles, celui de la banque, « à côté des fonds dits "actifs" (où les décisions d'investissement incombent encore majoritairement aux humains) se multiplie un nombre croissant de "fonds passifs", où les opérations de finance sont de plus en plus placées en pilotage automatique. » PITRON, 2021, p. 240 ; v. aussi LEE, 2019 (trad.), p. 22 sq.

115 « *Sapiens* est [...] devenu un *Homo détrit* qui produit, chaque année, l'équivalent de cinq mille tours Eiffel de déchets électroniques... » PITRON, 2021, p. 75-76.

116 L'IA repose sur la collecte massive des données, généreusement fournies par les utilisateurs des réseaux dits sociaux. « 98 % des revenus de Facebook proviennent de la publicité », d'après BOULIER, *Science et Vie*, 2022, p. 87.

117 « Les compagnies bien établies [Google, Baidu, Alibaba, YouTube]... s'enrichissent à chacun de nos clics » K.-F. LEE, 2019 (trad.), p. 174.

118 Les chiffres sont partout, mais un résumé limpide est fourni dans *Alternatives économiques*, 2021, p. 27.

119 On a l'impression que le Web s'améliore, puisqu'il nous envoie de plus en plus de choses qui nous flattent (et uniquement ce qui nous flatte). Sauf que, dans le même temps, les banques utilisent déjà « une multitude de paramètres "mineurs" » (rapidité à taper la date de naissance, pourcentage de batterie du *smartphone*, etc.), qui permettent à l'algorithme de prédire la solvabilité du client...LEE, 2019 (trad.), p. 182.

120 On peut y recevoir un ordre à n'importe quelle heure, n'importe quel jour ; *Ibid.*, p. 84.

Donc, si l'humanoïde n'est pas déjà là¹²¹, les dangers sont partout et les garde-fous semblent bien résider dans l'État ou quelque chose qui lui ressemble, pour mettre en œuvre des politiques publiques¹²².

ANARCHIE TEMPÉRÉE PAR LE POUVOIR PERSONNEL

L'IA est peut-être un peu plus responsable de l'évolution de nos sociétés, car elle renforce deux traits : le passage de la verticalité à l'horizontalité, né de la négation de l'instance tierce et accentué par l'action du *deep learning*, et l'infantilisation renforcée des *homines conectici*.

Deep Learning et instance tierce

En français « apprentissage en profondeur », le *deep learning* signifie la reconnaissance « des schémas récurrents profondément enfouis »¹²³. Quant à l'instance tierce, il s'agit d'un concept défini par Pierre Legendre¹²⁴ qui désigne la référence abstraite par rapport à laquelle les différenciations concrètes sont établies ; s'agissant du pouvoir, il s'identifie au totem par rapport auquel il s'articule : en *res publica* une idée, en *commonwealth* « c'est le contrat qui peut conduire jusqu'à ce qu'Alain Supiot nomme "l'homme décomposé"¹²⁵ »¹²⁶. Une logique verticale dans le premier cas, horizontale dans le

121 Et il n'existera peut-être jamais, v. *Ibid.*, p. 226.

122 *Ibid.*, p. 307 sq.

123 LEE, 2019 (trad.), p. 31. Rappelons que c'est à partir de l'utilisation de la mémoire des corrélations que la décision est prise. C'est ce que l'on nomme « l'"intelligence artificielle étroite" », par opposition à « l'"intelligence artificielle générale" [IAG], cette technologie universelle qui pourrait un jour effectuer toutes les tâches qu'accomplissent les êtres humains », déjà évoquée plus haut, mais dont il faut bien se souvenir ici.

124 « L'instance tierce est le lieu logique de la Loi en tant que référence pure à l'absolu du pouvoir qui se confond avec un principe logique de différenciation. » LEGENDRE, 1985, p. 145.

125 SUPIOT, 2005, p. 68 sq.

126 Cité par BOUINEAU, « *L'egomet...* », 2021, p. 36.

second¹²⁷. Or la puissance de l'IA se développe dans un contexte d'anarchie tempérée par des pouvoirs personnels (MM. Xi Jinping, Poutine et autres) qui opposent leurs détenteurs (*domini* et non *auctores*) à « ceux qui, souillés par la crasse du monde » – pour reprendre cette fois les mots d'Adalbéron¹²⁸ – n'ont plus comme échappatoire que l'eau lustrale de la consommation et du métavers. L'instance tierce y est impossible, car « l'une des caractéristiques du cyberspace est son caractère originellement non institutionnel »¹²⁹. En fait, le cyberspace et sa projection terrestre dans laquelle nous évoluons – que j'ai qualifiée plus haut de *mundus informaticus* – rappellent la Hanse¹³⁰ par bien des traits, et singulièrement par le fait que l'instance tierce s'y confond avec le *dominus*, ici Lübeck, là les GAFAM. Et s'il est facile de « démystifier la notion de réseau, qui n'est rien d'autre qu'une accumulation d'ordinateurs connectés les uns aux autres »¹³¹, il est aussi difficile d'altérer le totem engendré

127 Hannah Arendt a fort bien démontré, dans son analyse des groupes d'étudiants américains des années 1950, toute la force d'une autorité acceptée émanant d'un groupe.

128 Évêque de Laon au XI^e siècle, on lui doit la formule de ce que les historiens ont appelé la « trifonctionnalité » médiévale : « *Sunt qui orant, sunt qui pugnant, sunt qui laborant* » [il y a ceux qui prient, ceux qui combattent et ceux qui travaillent]. Il appartenait aux groupes aristocratiques de Lotharingie et en partageait les idées politiques.

129 Grand Collège..., 2020, p. 261.

130 Objet juridique mal identifié, la Hanse est une organisation commerciale hiérarchisée (à sa tête : Lübeck, en dessous des comptoirs – Bergen, Novgorod, Londres et Bruges – dont dépendent des factoreries, comme celle de La Rochelle, relevant du comptoir de Bruges), qui pratique une politique indépendante (Les villes hanséatiques refusent que l'empereur du saint empire romain germanique, dans le territoire duquel se trouve Lübeck, en fasse partie, mais elles y intègrent les communautés paysannes des Dithmarschen), mène sa propre politique internationale (des guerres sont déclarées par la Hanse) et détermine – en fait le corps de ville de Lübeck – la liste des villes affiliées... que l'on ne peut jamais connaître avec précision et qui change sans cesse. Sauf Lübeck.

131 *Ibid.*, p. 263.

par le métavers qu'à Platon de détourner les hommes de l'adoration des ombres au fond de la caverne. Mais ce n'est pas impossible.

Il suffit d'abord de partir du constat qu'il n'y a rien d'immatériel dans l'IA : ni son fonctionnement, comme cela a été remarqué plus haut, ni son existence même, puisque « L'IA reproduit le monde tel qu'on le lui présente »¹³² et qu'elle ne matérialise évidemment en rien quelque divinité. Tout d'abord, ce que l'on nomme IA dépend de milliards¹³³ d'informations stockées dans des armoires rangées au sein de « centres de données »¹³⁴ (*data centers* en américain), situés volontiers dans les pays froids¹³⁵ car le dégagement de chaleur induit par leur fonctionnement est considérable et que les refroidir consomme énormément d'électricité, donc ouvrir la fenêtre dans un pays froid présente un avantage sérieux. Ensuite, il s'agit d'une technique parvenue à son deuxième stade¹³⁶ de développement, et face à laquelle les plus intrépides des *geeks* ont en ligne de mire un troisième stade, qu'ils

132 DELÉPINE, 2021, p. 24.

133 Quarante-sept zettaoctets produits chaque année (PITRON, 2021, p. 117) : un zettaoctet vaut 1 000 exaoctets (toutes les informations produites dans le monde jusqu'en 2003 représentent cinq exaoctets ; v. l'annexe 5 dans *Ibid.*, p. 333).

134 À l'heure où j'écris, il en existe environ trois millions dans le monde, dont les plus petits font moins de 500 m² et le plus grand 600 000 m² – soit l'équivalent de 110 terrains de football (*Ibid.*, p. 131) – à Langfang en Chine. L'afflux des données est tel que la nécessaire multiplication des *data centers* pour les stocker devient un véritable problème : ne pouvant pas bétonner à l'infini des espaces de plus en plus vastes pour les héberger, on commence à envisager « des mini *data centers* de proximité [qui] pourraient ainsi fleurir dans chaque quartier », LEROUX, 2022, p. 82.

135 Rovaniemi est ainsi aujourd'hui, non seulement la ville du Père Noël mais aussi celle de l'IA.

136 Au premier stade, des entreprises – comme IBM – avaient réalisé la connexion des ordinateurs entre eux ; avec la seconde, ce sont des moteurs de recherche ou des réseaux dits sociaux dépendant de sociétés – comme Google ou Facebook – qui connectent les internautes entre eux.

qualifient « l’*Internet* de tout »¹³⁷, c’est-à-dire la possibilité de connecter tout ce qui pourra recevoir un capteur : objets, hommes et éléments naturels (arbres, pierres, etc.). Cette technique porte donc en soi un programme philosophique de réification, par assimilation des hommes aux choses, par mutation des perceptions¹³⁸ et par explosion de l’homme en composantes¹³⁹, comme cela a été relevé plus haut. Par ailleurs, sans tomber dans l’angoisse millénariste d’une humanité asservie à la toute-puissance des robots, comme l’IA ne calcule qu’en fonction de ce dont on l’a nourrie, il convient de prêter attention à son régime alimentaire, et au mois d’avril 2021 « la Commission européenne a proposé de nouvelles règles pour l’utilisation d’une IA “digne de confiance” »¹⁴⁰. La faille du système réside en cela qu’il n’y a pas d’instance tierce au niveau européen¹⁴¹ : ce sont les GAFAM qui régulent leurs activités et en cas de contentieux, les tribunaux compétents sont ceux des pays dans lesquels les GAFAM sont implantés et jusqu’en 2021 l’Irlande, éden des GAFAM, n’avait pris qu’une sanction¹⁴² en la matière, et elle était la seule, même si tout récemment un frémissement dans l’autre sens semble s’esquisser. Nous sommes donc bien pour l’heure en présence

137 Qu’ils nomment *The Internet of Everything*, parce qu’ils parlent américain.

138 Le phénomène existe déjà, grâce aux « accessoires dits de “retour haptique” » – LEROUX, 2022, p. 75 – c’est-à-dire les gants, vestes, combinaisons qui permettent à celui qui s’en équipe de percevoir des sensations extérieures de chaud, froid, pluie, etc.

139 Le *Guide pour une communication inclusive* présenté par Helena Dalli à la Commission européenne fin octobre 2021 et révélé le 28 novembre par *Il Giornale* est à ce titre très instructif : il déconseille l’emploi du mot « citoyen » (c’est-à-dire ce qui désigne l’homme abstrait, sujet de droit) pour inciter à considérer des foules de types humains variés.

140 FOULON, 2021., p. 30.

141 Or l’Europe possède le niveau de protection le plus élevé au monde. Même si « la pollution numérique semble être le cadet des soucis des institutions bruxelloises », PITRON, 2021, p. 50.

142 *Alternatives économiques*, 2021, p. 32.

d'une anarchie tempérée par le pouvoir personnel, faute d'instance tierce.

Bien au contraire, née dans un monde en déséquilibre majeur, l'IA va enkyster les disparités, par le fait, d'abord, que la capacité à maîtriser les interfaces numériques suppose en amont de savoir lire et écrire¹⁴³, ce qui repose la question de la priorité de l'enseignement. De plus, il va bien falloir résoudre une alternative simple : ou bien l'IA est faite pour les hommes et elle doit donc leur faciliter la vie et, autant que faire se peut, être adaptée à ce qu'ils sont dans leur diversité¹⁴⁴, ou bien elle constitue un instrument de domination au profit d'une oligarchie se comportant en *dominus* et non pas en *auctor*, décidément, qui commence par imposer ses valeurs, sa vision du monde et naturellement sa langue, dans le temps même où les progrès techniques mettent la traduction simultanée à portée d'oreille pour tout un chacun. Mais comme tout dogme, celui de la programmation des machines qui permettent l'IA obéit à des valeurs : l'absence d'instance tierce et l'alignement de tous sur le lit de Procuste¹⁴⁵ de la consommation ; car l'IA possède un objectif : celui de servir un Marché et non pas des hommes. La suppression des statuts¹⁴⁶ y est donc un prérequis pour qu'advienne *homo conecticus*, comme cela a été vu. Or si d'apparence on pourrait faire le rapprochement avec les Lumières, la réalité est tout autre : les philosophes du XVIII^e siècle ont désacralisé une lettre (*i.e.* un ensemble de dogmes), mais pas un esprit (*i.e.* la cohérence de l'être

143 Évidence rappelée à fort juste titre régulièrement ; v. par exemple *The Shift Project*, 2018, p. 44.

144 C'est pourquoi on parle volontiers d'« un écosystème numérique local », prévoyant le « développement de contenus locaux (informations et langues) sur les sites, hébergeurs nationaux ou régionaux, etc. » *Ibid.*, p. 58.

145 Ayant établi son QG à un carrefour de l'Attique, il voulait réduire les hommes à la même taille en les allongeant sur un lit de fer : il y étirait les plus petits et coupait ce qui dépassait du lit aux plus grands. Tous étaient frappés et ils en mouraient tous...

146 Ou plus exactement la qualité des personnes liée au statut.

humain). En revanche, si l'on n'y prend garde, le *deep learning* désarticule l'être humain en le soumettant à un *dominus*. Or, si l'on en croit Aristote, l'homme est un animal politique. J'ajouterai : l'homme d'une *res publica* et tous ceux qui relèvent de civilisations traditionnelles. L'homme occidental en *res publica* forme une *ecclesia*¹⁴⁷, avec des repères communs et partagés ; c'est un citoyen¹⁴⁸. Lui ôter cela revient à priver le chat de ses vibrisses. Mais le *geek* voit-il que le chat a des vibrisses ?

Car bien entendu, l'IA n'est qu'une magnifique technique à la base. En revanche si elle se substitue aux valeurs constitutives des individus, elle fait d'*homo conecticus* un colonisé¹⁴⁹. Et si l'environnement planétaire doit se modeler sur celui de Masdar¹⁵⁰, il conviendra au préalable de résoudre deux difficultés : la première celle qui consiste à faire accepter un modèle de vie uniforme¹⁵¹ et la seconde, peut-être encore plus difficile à atteindre, transformer l'existant. Parce que derrière la problématique de l'IA se tapit une question que je n'ai jamais vu abordée dans mes lectures : celle de la possibilité d'un monde futur pluriel, envisagé autrement que sur le mode bigarré au sein d'un système de valeurs occidental, version anglo-saxonne, et festonné d'équipements

147 Ce qui veut dire « assemblée » en grec, d'où l'Église tirera à la fois son nom et son programme d'action : assembler les hommes.

148 Il est intéressant de voir le *Guide...* européen évoqué plus haut proscrire l'emploi de ce terme, pour ne pas exclure les migrants. On passe ainsi d'une vision juridique à une approche sociologique, caractéristique des analyses anglo-saxonnes. Mais l'autrice de *Guide* est une Maltaise.

149 Ou « simplifié », pour reprendre le titre de l'ouvrage de BESNIER, *L'Homme simplifié : le syndrome de la touche étroite*, Paris, Fayard, 2012, cité par PENEL, 2020, p. 341, n. 8.

150 Tout à la fois la Jérusalem, La Mecque et la Rome des *geeks* et des *geekettes*, qui y voient le modèle de la cité idéale, où tout est connecté et informatisé au maximum des possibilités actuellement maîtrisées. Elle est construite dans un bout du désert d'Abou Dabi.

151 « Masdar promet d'être un modèle régional et même mondial de développement urbain intelligent. » G. PITRON, 2021, p. 26.

électroniques. Or autant, en effet, il semble nécessaire que la formidable puissance de calcul de l'algorithme aide à conjurer les défis de la surpopulation, du climat¹⁵², de la pollution, bref de toutes les difficultés, disons même de tous les dangers qui s'amplifient au fil des ans, autant ce serait folie que de vouloir asservir l'ensemble du monde à un seul mode de vie. Sans aller chercher d'exemples aux antipodes, la « mise aux normes » de certains immeubles historiques à Paris sous la volonté d'une ministre emportée par une sensibilité opiniâtrement écologique suffit pour faire comprendre l'impasse qui nous menace. Et pour ne rien dire des processus de rejet violent et armé qui sont certes sous-tendus par bien des motifs, mais qui se rejoignent dans le refus de se soumettre à la colonisation d'une technologie martelée en *globish*.

Peut-être est-il temps de relire Montaigne. Son incertitude et l'intelligence de son regard. Je reste pour ma part atterré devant certaines initiatives, comme celle qui consiste à utiliser les services d'un robot pour écarter « les faiblesses émotionnelles et psychologiques qui encombrant le raisonnement humain »¹⁵³. À cela, il convient d'ajouter deux éléments qui peuvent transformer l'IA en menace : l'effet rebond et le principe de plaisir. L'effet rebond consiste à induire une utilisation plus importante à mesure que les contraintes s'abaissent et que les possibilités offertes se multiplient¹⁵⁴ ; en la matière la 5G constitue un cas d'école : les prouesses technologiques induites sont telles que les utilisateurs l'utiliseront par ce que son utilisation fera naître de nouveaux désirs¹⁵⁵. Le principe de plaisir découle de la

152 À supposer que la « ville intelligente » soit la vraie solution, ce que critiquent certains chercheurs, qui y voient une aggravation des difficultés environnementales ; *Ibid.*, p. 31.

153 *Ibid.*, p. 245.

154 Alors même que l'effet rebond atténue le bénéfice des progrès, dans la mesure où il induit une utilisation accrue dans un même secteur ; v. *The Shift Project*, 2018, p. 11 et 61.

155 PITRON, 2021, p. 213.

destruction de l'instance tierce, des attaques contre la *res publica* – là où elle existe – et d'un comportement omphalopsyque généralisé, accru par une angoisse obsidionale qui fait littéralement implorer *homo conecticus* derrière ses lunettes de métavers.

La catastrophe globale est-elle inévitable¹⁵⁶ ? Outre qu'on ne peut jamais l'assurer tant qu'elle n'est pas là, quelques indicateurs peuvent encourager à garder espoir : un *Internet* globalisé semble désormais impossible et certaines grandes puissances et certaines grandes firmes risquent à l'avenir de dominer des réseaux « régionaux »¹⁵⁷ et par ailleurs, au rythme où croît le stockage des données, pourra-t-on vraiment tout conserver, ou tout conserver sur le même plan ? À ces deux cas de figure, une seule et même solution : une régulation, qui ne sera vraisemblablement pas une instance tierce en raison de la philosophie qui anime l'IA, mais du moins la recherche d'un équilibre ; une *politeia*¹⁵⁸ ?

IA et infantilisation

Et si le sursaut venait de l'homme lui-même ? L'ordinateur qui rend possible l'IA n'est pas un outil, c'est une machine : l'outil prolonge la main, la machine est « un instrument propre à communiquer un mouvement » et son étymologie¹⁵⁹ insiste sur la notion de création. L'ordinateur est bien créé par l'homme et l'IA qui en est issue est une création adaptée non pas à la main, mais à l'imagination. Si dématérialisation il y a, elle se trouve à ce niveau et à ce niveau seulement, car pour la

156 « Nous ne comprenons pas que nous enfantons un monde où les divertissements généreront des tensions, voire des conflits, parce qu'ils ne nous seront offerts qu'au prix d'un impact spatial, matériel, dont nous ne pourrons jamais nous soustraire. » *Ibid.*, p. 305.

157 Grand Collège..., 2020, p. 264.

158 Concept « conclusif » à l'*anakyklôsis* chez Aristote, et titre donné à son ouvrage, traduit généralement par : « La politique ». Réduit à une idée, on peut le résumer à un moment d'équilibre.

159 *Mèchanè*, « invention ingénieuse », selon Littré.

mettre en œuvre, il y a besoin d'objets et d'outils, mais on ne cesse de parer la machine d'une force propre : c'est une épiphanie¹⁶⁰. Toutefois, aligner l'homme sur la machine conduit à une déshumanisation¹⁶¹, une réification comme je l'ai déjà écrit, et donc « à lui faire cesser d'être la mesure des choses¹⁶² pour en faire un demiurge »¹⁶³. Ou un enfant, qui s'imagine pouvoir refaire le monde à sa mesure.

Les modèles nés de l'IA sont en effet des parcelles de *fellows* recomposées en *exempla*¹⁶⁴ : le monde virtuel fait – au niveau philosophique – passer de la 3D à la 2D, car il n'existe plus ni hauteur, ni verticalité, à l'image des constructions produites par les enfants, qui sont des manifestations de leur imaginaire et de leur manière de voir le monde, sans intervention d'instance tierce. Or, si les savants ont remplacé les prêtres à l'époque des Lumières, de nos jours l'informaticien, nouveau flamme, chasse le savant et de nouveaux fidèles le servent : les *geeks*. Mais si « le développement des sciences fit apparaître des savants sûrs d'eux, jaloux des philosophes et des prêtres, sans pour autant leur apporter le pouvoir politique »¹⁶⁵, l'IA place ses sicaires hors d'atteinte du vulgaire, par l'ésotérisme de son langage, la puissance de ses soutiens et surtout l'assimilation que l'on fait entre elle et la Vérité, dématérialisée comme une divinité¹⁶⁶.

160 *Epifaneia*, « action de se montrer, d'où apparition », précise Littré comme premier sens induit.

161 V. les travaux de BERROU et DUPONT, 2012, *passim* ; et écouter BERROU : « L'intelligence artificielle devient réalité » (<https://www.canalacademie.com/ida10135-L-intelligence-artificielle-devient-realite.html>) (consulté le 11 juillet 2019).

162 Ce qui est la véritable formule de Protagoras, même si le confort de l'habitude fait répéter l'erreur commune : l'homme est la mesure de toute chose.

163 BOUINEAU, 2020, p. 69.

164 *IDEM.*, « *L'egomet...* », 2021, p. 36.

165 MARTRES, 2017, p. 175.

166 « Nous avons coutume de considérer le numérique tel un messie venu parmi les hommes pour les sauver. » G. PITRON, 2021, p. 315. L'esthétique

Ainsi naît l'« utopie de l'affranchissement de la matière »¹⁶⁷ : comme dans le *Phédon* Platon enseigne que le corps, par sa matérialité, entrave l'âme, le *geek* ne vit plus dans un monde d'objets, mais de services : Deezer ne vend pas des disques, mais offre de la musique, et comme dans le monde enchanté des enfants, tout est gratuit, hormis le fait que nous livrons nos données en échange, c'est-à-dire que nous nous livrons de manière volontaire à l'Une¹⁶⁸ – qui peut nous écorcher comme Marsyas –, en rêvant pour certains d'une « transmission de données sans transmission » grâce à la technologie quantique¹⁶⁹. Car l'IA se présente comme une divinité aimable, qui flatte chacun, puisque le but de l'algorithme consiste à trouver quel produit vendre, et donc à séduire le client potentiel par des tentations qui lui parlent d'autant mieux qu'elles lui sont faites *intuitu personae* à partir des données qu'il a lui-même fournies. Mais c'est une divinité à laquelle il est impossible d'échapper, car l'*adiaphoria*¹⁷⁰ est devenue inopérante au sein d'une pseudo-*ecclesia* qui n'est plus institutionnelle, ni même une communauté d'hommes¹⁷¹.

De plus, dans le nouvel espace de l'IA les comportements comme les discours conduisent à une infantilisation, par prévalence du principe de plaisir, du monde merveilleux et

des objets sous-tend cette dimension et renforce le sentiment de déréalisation : Apple se réfère ainsi explicitement au bouddhisme zen pour dessiner ses appareils – v. *Ibid.*, p. 70 – alors que l'éclatement du collectif en atomes individuels engendre *ipso facto* l'agonistique.

167 *Ibid.*, p. 67.

168 L'ami de Montaigne qualifiait le tyran « l'Un », mais là où La Boétie mettait en garde contre la servitude volontaire, l'IA se bâtit sur les chaînes que nous lui apportons pour nous entraver.

169 PITRON, 2021, p. 282.

170 « Indifférence », en grec. Le terme a été utilisé par les stoïciens pour désigner ce qui, n'étant ni bien ni mal, n'avait pas d'importance à leurs yeux, mais aussi dans les conflits religieux où les groupes persécutés laissent croire qu'ils se soumettent à la religion majoritaire, alors qu'il n'y a là qu'apparence (l'exemple le plus connu est sans doute celui des adiaphoristes, luthériens du XVI^e siècle).

171 BOUINEAU, « *L'egomet...* », 2021, p. 32.

d'un langage souvent simpliste¹⁷². Et par un changement dans la conscience du temps : « Nous sommes passés de la logique du présent à celle de l'instant »¹⁷³. Dès lors, le temps n'a plus de durée et le futur est vécu comme un présent qui tarde, ce qui explique les délires des plus enthousiastes des *geeks* ou les terreurs des plus timorés des inquiets¹⁷⁴. Au demeurant, même les plus sérieux proposent parfois des perspectives bien naïves, comme l'« allocation d'investissement social » pour ceux qui accompagnent, aident, éduquent ou le « nouveau pacte social » autour des « activités socialement bénéfiques »¹⁷⁵. Il manque à ce raisonnement, à mon avis, deux choses : d'une part la prise en considération du fait que les hommes sont viscéralement violents et que sans principe régulateur, les équilibres ne naîtront pas plus que dans l'espace actuel où le Marché fait la loi ; d'autre part – mais c'est sans doute là une marque de sa culture – le fait que sans instance tierce, c'est-à-dire sans totem extérieur auquel adhérer, la perte du sens collectif ne pourra que s'accroître. Dès lors, peut-on conclure, comme le fait d'une part Kai-Fu Lee, que l'IA n'est qu'une étape pour la découverte de la nature profonde de l'homme et d'autre part que « la clé de l'“IA” n'est pas le développement cognitif »¹⁷⁶ ?

172 Ce qui vient confirmer l'hypothèse que j'émis voici peu : « Ce serait donc par immaturité, en raison de l'absence d'un passage au stade adulte dans lequel, en principe, l'individu est capable d'évoluer entre des bornes convenues, que Hobbes appelle de ses vœux un Léviathan. » BOUINEAU, *La pensée de gouvernement*, 2021, p. 16.

173 PITRON, 2021, p. 145.

174 Pour une présentation synthétique des écrits les plus emblématiques des deux camps, il faut se reporter au passage déjà cité de LEE, 2019 (trad.), p. 222 sq.

175 *Ibid.*, p. 331.

176 PENEL, 2020, p. 344.

BIBLIOGRAPHIE

BERROU, Claude, « L'intelligence artificielle devient réalité », *Canal Académies*, 2013. (https://www.canalacademie.com/id_a10135-L-intelligence-artificielle-devient-realite.html)

BERROU, Claude et DUPONT, Vincent, *Petite mathématique du cerveau. Une théorie à la formation mentale*, Paris, Odile Jacob, 2012.

BOUINEAU, Jacques, « Antiquité et territoires connectés », *Annales de l'académie des Belles-Lettres, Sciences et Arts de La Rochelle*, t. 22 (nouvelle série), 2020, p. 59-76.

BOUINEAU, Jacques, « Citoyen et *res publica* 2.0 », *Domination et Antiquité. Aspects sociaux et économiques*, Jacques BOUINEAU (dir.), Paris, L'Harmattan, 2021, p. 187-222.

BOUINEAU, Jacques, « L'*egomet*. Réflexion sur la dimension juridique de l'homme libre », *Historia et jus*, n° 20, 2021, p. 1-47.

(http://www.historiaetius.eu/uploads/5/9/4/8/5948821/bouineau_20.pdf)

BOUINEAU, Jacques, « La pensée de gouvernement », *Psychiatries*, « Autorité. Perte et reconnaissance », n° 173, mars 2021, p. 6-23.

BOUINEAU, Jacques, « La solidarité chez Campanella », in Burt KASPARIAN (dir.), *Les espaces de solidarité. La famille, l'État, l'Europe et le monde*, Rennes, PUR, 2015, p. 99 sq.

BOUINEAU, Jacques, *Traité d'histoire européenne des institutions (I^{er}-XV^e siècle)*, Paris, Litec, 2004.

BOUINEAU, Jacques, *Traité d'histoire européenne des institutions (XVI^e-XX^e siècle)*, Paris, Litec, 2009.

DELÉPINE, Justin, « L'intelligence artificielle en trois questions », *Alternatives économiques*, n° 416, octobre 2021.

EYAL, Nir, *Hooked Comment créer un produit ou un service qui ancre des habitudes*, Paris, Eyrolles, 2018 (trad.).

FOULON, Sandrine, « Des algorithmes encore très humains », *Alternatives économiques*, n° 416, octobre 2021, p.

Grand Collège des Rites Écossais, GODF, *Construire demain ? Regards maçonniques sur les solidarités post-Covid*, <https://lbapres.godf.org/app/uploads/2020/12/GCDRE-Contribution-au-Livre-blanc>.

LAROUSSERIE, David, « L'intelligence artificielle, génie de la biologie moléculaire », *Le Monde « Science et médecine »* n° 23 882, 20 octobre 2021.

LAVERDET, Caroline, entretien dans *Science et Vie* n° 1252, janvier 2022.

« LEAN ICT – Pour une sobriété numérique », Rapport du groupe de travail dirigé par Hugues FERREBŒUF pour le think tank The Shift Project – octobre 2018. <https://theshiftproject.org/wp-content/uploads/2018/11/Rapport-final-v8-WEB.pdf>

LEE, Kai-Fu, I.A. *La plus grande mutation de l'histoire*, Paris, Les Arènes, 2019 (trad.).

LEGENDRE, Pierre, *L'inestimable objet de la transmission*, Paris, Fayard, 1985.

LEROUX, Hugo, « La face cachée du métavers. Quel impact sur le monde réel ? », *Science et Vie*, n° 1252, janvier 2022.

MARTRES, Jean-Louis, *Les Grilles de la pensée politique*, Nice, Libre-échange, 2017, vol. 2.

OLIVIER, Arthur (mis à jour le 28 X 2021 par Raphaël CHAMOULARD), « Infographies : les émissions de gaz à effet de

serre dans l'Union européenne », [https://www.touteleurope.eu/environnement/les-emissions-de-gaz-a-effet-de-serre-dans-lunion-europeenne/"s-emissions-de-gaz-a-effet-de-serre-dans-lunion-europeenne/](https://www.touteleurope.eu/environnement/les-emissions-de-gaz-a-effet-de-serre-dans-lunion-europeenne/)

PENEL, Patrick, « Ruée vers l'intelligence artificielle et les technologies qui y sont liées : réalités et fantasmes ? », *Revue de l'académie du Var*, année 2019, Nouvelle série, t. XX.

PITRON, Guillaume, *L'enfer numérique*, Paris, Éditions Les liens qui libèrent, 2021.

REDAUD, Lorraine, « Réalité virtuelle Zuckerberg veut tous nous transformer en Mélanchon », *Charlie Hebdo*, n° 1516 du

11 août 2021.

SEALES, Brent, *Sciences et avenir*, n° 897, novembre 2021.

SUPIOT, Alain, *Homo juridicus Essai sur la fonction anthropologique du droit*, Paris, Seuil, 2005.

LA POLITIQUE DE LA COMMISSION ET DE L'UNION EN MATIÈRE D'INTELLIGENCE ARTIFICIELLE ET DE SOUVERAINETÉ NUMÉRIQUE

*par M. Thierry Breton
Commissaire européen au Marché intérieur*

La crise du Covid-19 puis l'invasion russe en Ukraine ont mis tragiquement en évidence l'importance pour l'Europe de construire une véritable autonomie stratégique – en matière de chaînes d'approvisionnement, de sécurité énergétique, de capacités de défense et dans l'espace numérique. La souveraineté numérique repose sur plusieurs piliers indissociables, y compris la puissance de calcul, la connectivité sécurisée et la maîtrise des données. Dans un monde qui génère chaque jour davantage de données, l'intelligence artificielle (IA) est une technologie-clef pour les gérer, les interpréter et les transformer en applications concrètes. À cette fin, il nous faut impliquer toute une filière industrielle.

LES SEMI-CONDUCTEURS

Commençons avec la dimension matérielle, les semi-conducteurs. À cet égard, l'Europe a de très bonnes cartes en main. Nous sommes le centre mondial en matière de recherche sur les semi-conducteurs. L'ensemble de mes interlocuteurs publics comme privés le reconnaissent : sans l'Europe et ses centres de recherche de premier plan, il n'y aurait pas d'avancée technologique possible. L'Europe est au cœur à la fois de la technologie FinFet nécessaire pour graver des puces sous les 5 nanomètres et de la technologie FDSOI cruciale pour la réduction de la consommation énergétique. Ces technologies sont donc vitales pour la transition écologique et numérique.

Cependant, je ne crois pas que l'Europe doive se concentrer uniquement sur la recherche et la conception, laissant la production aux seuls acteurs asiatiques et américains. Une autonomie stratégique sans capacité de production est vouée à l'échec. C'est ce que nous enseigne la pénurie actuelle. Il ne s'agit évidemment pas de tout faire ou de tout relocaliser en Europe, mais nous devons anticiper les risques, les analyser, prendre les mesures adéquates et rééquilibrer les chaînes d'approvisionnement.

Notre ambition est de passer de 10 % de la production mondiale à 20 % d'ici à 10 ans. Or, comme le marché des semi-conducteurs va doubler en 10 ans (passant de 500 milliards à 1 000 milliards d'euros du fait de l'essor de la numérisation industrielle, de l'internet des objets ou de l'IA), il s'agira en fait de quadrupler la capacité de production actuelle en Europe.

C'est la raison pour laquelle la Commission a adopté au mois de février 2021 un paquet législatif sur les semi-conducteurs. Grâce à ce paquet législatif, nous renforcerons notre excellence en matière de recherche et faciliterons le passage du laboratoire à l'usine. Nous mobiliserons, pour renforcer nos capacités de production, des fonds publics considérables qui attirent déjà d'importants investissements privés. Nous mettrons, enfin, tout en place pour sécuriser l'ensemble de la chaîne d'approvisionnement et éviter de futurs chocs pour notre économie, comme ceux auxquels nous avons été confrontés avec la pénurie actuelle.

Car ce sont précisément les améliorations de performance des puces électroniques qui ont rendu possible l'essor récent de l'IA. Des puces de plus en plus puissantes sont indispensables pour gérer les quantités importantes de données utilisées par l'IA. Par exemple, les GPUs (*Graphic Processing Units*) sont passées de simples puces à vocation purement graphique à des éléments centraux des modèles d'apprentissage automatique parce qu'ils peuvent gérer de multiples calculs de façon simultanée.

LES APPLICATIONS DE L'IA

Passons donc aux applications de l'IA. Ici, l'approche de la Commission est double. D'un côté il s'agit d'investir dans l'excellence européenne, tandis que de l'autre côté il est impératif de s'assurer que l'IA en Europe est développée selon nos règles et nos normes.

En ce qui concerne *l'excellence européenne*, la Commission investira à partir de ses programmes Digital Europe et Horizon Europe aux alentours d'un milliard d'euros par an. Cet investissement public européen pourra générer des investissements privés complémentaires qui nous permettront de propulser l'Europe à la pointe de la technologie. Notre objectif est d'atteindre un volume d'investissement annuel collectif par l'UE et les États membres de 20 milliards d'euros par an au cours de cette décennie, grâce aussi à une partie des 140 milliards d'euros qui financeront les investissements numériques du plan de relance NextGenerationEU. L'objectif est ambitieux, certes, mais – à titre de comparaison – aux États-Unis le seul secteur privé a investi presque 15 milliards de dollars l'année dernière.

À quoi serviront ces investissements concrètement ? Par exemple :

- à des **pôles d'innovation numérique** (« Digital Innovation Hubs ») spécialisés dans l'IA : dans chaque État membre des guichets uniques donneront accès à l'expertise technique et à l'expérimentation pour permettre aux PME de « tester avant d'investir ».
- à des **installations d'essai et d'expérimentation** qui permettront aux PME d'effectuer gratuitement des tests sur leurs systèmes d'IA, sans avoir à construire elles-mêmes une nouvelle installation dans des environnements réels. Nous prévoyons également la possibilité d'introduire des bacs-à-sable réglementaires (*regulatory sandboxes*) afin de soutenir nos PME et nos jeunes pousses.

- aux **compétences numériques avancées**. Plus de 80 % des entreprises mentionnent une pénurie des compétences comme obstacle à l'adoption de l'IA. Une meilleure analyse des compétences demandées ainsi qu'un soutien aux stages dans les filières numériques ne sont que deux exemples de comment nous pouvons assister les États membres dans ce domaine.

La **dimension écologique** doit aussi être prise en compte. Il est important de souligner que l'IA peut avoir des bénéfices énergétiques considérables, souvent ignorés dans les discussions. Par exemple, les voitures autonomes ne réduiront pas seulement la pollution grâce à la réduction des embouteillages, elles éviteront aussi la plupart des accidents, ainsi que la consommation d'énergie nécessaire aux réparations.

Dans le domaine de la **santé**, très prometteur, traiter une maladie diagnostiquée plus tôt et mieux, pourra aider à réduire les séjours à l'hôpital, mieux anticiper les interventions chirurgicales et cibler l'utilisation de médicaments.

Avec potentiellement des centaines de milliers d'accidents de voitures évités et un nombre similaire d'interventions médicales évitées, rien que ces deux cas d'utilisation de l'IA peuvent avoir des bénéfices importants pour la société. D'autant que l'IA n'est pas concentrée dans quelques secteurs mais a vocation à être mise en œuvre dans tous les pans de l'économie. Ainsi c'est l'accumulation de nombreux – souvent petits – bénéfices environnementaux, qui donnera de l'envergure sociétale à l'intelligence artificielle.

Afin d'augmenter ces bénéfices, la Commission soutient la recherche et le développement d'une IA plus écologique. Par exemple, elle finance des projets d'analyse de la gestion intelligente de l'énergie, basée sur l'IA, qui améliore la recherche de l'équilibre entre l'offre et la demande. Citons aussi le secteur agro-alimentaire, où existent déjà des systèmes intégrés d'agriculture de précision qui permettent d'améliorer la santé des plantes tout en utilisant moins de pesticides. Là

aussi, la Commission finance des projets très avancés.

Il faut avoir, cependant, conscience du *revers de la médaille*. L'IA consomme énormément d'énergie. Un algorithme extrêmement perfectionné basé sur 1.5 milliard de paramètres produit également des milliers de tonnes de CO₂. Il est, par exemple, estimé que GPT-3, un modèle puissant de langage naturel développé par OpenAI, a consommé autant d'énergie dans sa phase d'entraînement qu'une voiture ayant fait un trajet aller-retour de la terre à la lune.

C'est pour cela qu'une de nos priorités est de financer des recherches pour rendre l'IA elle-même plus écologique. Des solutions envisagées incluent, notamment, des puces à basse consommation d'énergie, l'optimisation des architectures et modèles d'IA permettant de travailler avec des ensembles de données plus restreints, et de soutenir d'autres approches qui montrent une réduction substantielle de la consommation d'énergie et de données. Mais pour analyser ces potentiels plus en détail, nous devons également améliorer nos méthodes pour mesurer la consommation. Ainsi, nous visons à créer plus de transparence concernant l'impact des services numériques, et notamment de l'IA.

Toutes ces actions, la Commission ne peut naturellement les mener seule. Les États Membres sont en train de mettre à jour leurs stratégies nationales – telles que la révision de la stratégie française l'année dernière, à la fois ambitieuse et réaliste – que nous appuyons et qui soutiennent nos actions. Notre démarche s'inscrit donc dans une approche globale conjointe avec les autorités nationales, détaillée dans le Plan Coordonné pour l'IA révisé que nous avons adopté en avril 2021 en même temps que la proposition de règlement sur l'IA.

LE CADRE RÉGLEMENTAIRE SUR L'IA

La proposition de *création d'un cadre juridique* présentée par la Commission s'insère dans la double approche décrite plus haut, à savoir développer le potentiel de l'IA tout en

s'attaquant aux problèmes que ces technologies peuvent causer. Elle n'a pas du tout été conçue dans une vision négative de l'IA. Nous désirons créer un écosystème d'excellence et de confiance. Ainsi, la proposition se fonde sur les valeurs de l'UE et les droits fondamentaux, et vise à donner aux personnes et aux autres utilisateurs la confiance nécessaire pour adopter des solutions fondées sur l'IA, tout en encourageant les entreprises à développer ces solutions.

La proposition présente une approche réglementaire horizontale qui se limite aux exigences minimales nécessaires pour répondre aux risques et aux problèmes liés à l'IA, sans restreindre ou freiner indûment le développement technologique ni augmenter de manière disproportionnée les coûts de mise sur le marché de solutions d'IA. Elle établit un cadre conçu pour résister à l'épreuve du temps, notamment au travers d'exigences fondées sur des principes auxquels les systèmes d'IA devraient se conformer. Par ailleurs, le cadre juridique prévoit des mécanismes qui permettent de l'adapter de manière dynamique à l'évolution de la technologie et aux nouvelles situations potentiellement préoccupantes.

La proposition de la Commission est fondée sur une approche liée aux risques, puisqu'elle se concentre sur les cas d'utilisation de l'IA dits « à haut risque ». Sont considérés à haut risque des systèmes tels que ceux utilisés pour évaluer la solvabilité des personnes physiques, pour établir leur note de crédit, ou encore pour le recrutement ou le filtrage des candidatures. Ceci reflète notamment l'impact que ces systèmes peuvent avoir sur les personnes affectées. Seuls ces cas doivent respecter un ensemble d'exigences spécifiquement conçues pour renforcer la transparence et minimiser les risques pour la sécurité et les droits fondamentaux.

Des règles proportionnées et flexibles sont donc proposées pour répondre aux *risques spécifiques posés par les systèmes d'IA à haut risque dans certains contextes*. Parmi les exigences spécifiques, citons une gouvernance des données de qualité, une documentation technique adéquate, un

enregistrement automatique systématique, de la transparence et la fourniture d'informations aux utilisateurs ou encore un contrôle humain lorsque les circonstances l'exigent. Notons enfin que des restrictions et des garanties spécifiques sont proposées en ce qui concerne certaines utilisations de systèmes d'identification biométrique à distance à des fins répressives. D'ailleurs, parmi les systèmes d'IA autonomes à haut risque, seuls les systèmes d'identification biométrique à distance devraient être soumis à une évaluation de la conformité réalisée par un tiers.

De plus, certaines pratiques – peu nombreuses – d'IA particulièrement néfastes sont interdites en raison de leur caractère contraire aux valeurs de l'Union. La proposition interdit notamment la notation sociale fondée sur l'IA effectuée à des fins générales par les autorités publiques. Les interdictions portent également sur les pratiques qui présentent un risque important de manipuler des personnes par des techniques subliminales agissant sur leur inconscient, ou susceptibles d'exploiter les vulnérabilités de groupes spécifiques tels que les enfants ou les personnes handicapées afin d'altérer sensiblement leur comportement d'une manière susceptible de causer un préjudice psychologique ou physique à la personne concernée ou à une autre personne.

Enfin, pour certains systèmes d'IA spécifiques, seules des obligations minimales en matière de transparence sont proposées, en particulier lorsque des dialogueurs (*chatbots*) ou des trucs vidéo ultra-réalistes sont utilisés.

Ce cadre réglementaire unifié renforcera le fonctionnement du marché unique. Il soutiendra l'innovation à travers davantage de sécurité juridique et la mise en place d'un environnement prévisible pour les développeurs, les innovateurs et les entreprises. Il augmentera la confiance et, par conséquent, l'utilisation de l'IA. Il contribuera ainsi à l'objectif de l'Union Européenne de devenir un centre névralgique du développement et de l'utilisation d'une IA digne de confiance. Car, ne nous trompons pas, la souveraineté

numérique passe par l'excellence technologique. Ce doit être leur performance, rendue possible par le cadre que nous aurons créé, qui permettra aux entreprises européennes de réussir face à la concurrence internationale, et non pas une protection particulière.

Notre proposition de réglementation constitue la première approche au monde qui couvre les risques tous azimuts engendrés par l'IA. Il y a énormément d'intérêt à l'international pour notre approche, car le monde entier est confronté à des problèmes similaires. Notre cadre européen peut donc devenir la référence internationale de la réglementation de l'IA.

Le contrôle de l'application des règles proposées sera assuré au moyen d'un système de gouvernance au niveau des États membres reposant sur des structures déjà existantes ainsi que d'un mécanisme de coopération au niveau de l'Union accompagnant la création d'un Comité européen de l'IA.

LE DÉVERROUILLAGE DES DONNÉES

Troisième et dernier axe enfin, les *données* sont essentielles. La disponibilité de données est nécessaire pour la formation de systèmes d'IA, qui en ont besoin en grande quantité et variété, à condition que leur niveau de qualité soit adéquat. Les données sont vitales, au-delà de l'IA, pour le développement économique : elles constituent la base de nombreux produits et services nouveaux à l'origine de gains de productivité et d'efficacité dans l'utilisation des ressources dans tous les secteurs de l'économie, permettant de proposer des produits et des services plus personnalisés, d'améliorer l'élaboration des politiques et de moderniser les services publics. Elles représentent une ressource essentielle pour les *start-ups* et les PME. Or, actuellement, une grande partie des données du monde entier est aux mains d'un petit nombre d'entreprises, souvent qualifiées de « *Big Tech* ». Cette concentration pourrait réduire considérablement les incitations

à l'émergence, la croissance et l'innovation d'entreprises fondées sur les données dans l'UE aujourd'hui, mais des possibilités considérables se profilent à l'horizon.

À cet égard, il convient de souligner l'adoption récente de deux propositions par la Commission. D'une part, *l'acte sur la gouvernance des données* améliorera le partage des données et le développement des espaces communs européens de données. Il augmentera la confiance en établissant des règles pour les intermédiaires traitant des données, ce qui fera croître le volume des données disponibles et échangées dans l'UE par-delà les secteurs et les frontières. À cette fin, il propose aussi de nouvelles règles quant à l'accès et à la réutilisation des données publiques, donne aux individus plus de contrôle sur leurs données, et crée un environnement sécurisé pour tous ceux qui veulent s'engager dans l'altruisme en matière de données.

D'autre part, *l'acte sur les données*, proposé en début d'année, vise à rendre la distribution de la valeur dérivée des données plus équitable. La proposition porte notamment sur l'accès et l'utilisation des données en provenance des objets connectés et l'utilisation des données gérées par les acteurs privés pour certains objectifs exceptionnels d'intérêt public. En même temps, elle s'attelle à la protection légale des bases de données tout en préservant les incitations économiques aux investissements dans la génération des données.

Le triptyque exposé, centré sur les puces, les applications de l'IA et les données, s'inscrit dans une stratégie plus générale, très ambitieuse, en matière de numérique. La connectivité, les compétences numériques en général et en matière d'IA en particulier, le soutien à la recherche ou encore les mesures en faveur de la cybersécurité, élaborées dans la stratégie de l'Union Européenne pour la Décennie Numérique, sont autant d'éléments nécessaires à la souveraineté numérique européenne.

Agir sur tous ces leviers demande indubitablement la mobilisation de tous, mais l'Europe dispose de tous les atouts nécessaires.

INTELLIGENCE ARTIFICIELLE, UN BRIN DE FOLIE

*par M. Jacques Boucharlat
de Académie Delphinale*

INTRODUCTION

Le terme de folie est désuet, car polysémique. Il a longtemps désigné les troubles de l'esprit, les maladies mentales, mais il disparaît progressivement au profit du terme de troubles mentaux. Il signifie, actuellement, le goût excessif, la passion échappant à tout contrôle et non la maladie mentale. Érasme en faisait déjà l'éloge en 1511. Giraudoux écrit *La folle de Chaillot* pour évoquer l'extrême originalité des années folles, la folle du logis, la vacance de la raison, la démesure, « *l'hubris* ».

La folie est aussi le reflet de l'au-delà, selon Tobie Nathan. Pour Shakespeare, dans sa pièce *Comme il vous plaira*, la nature, la maladie sont des êtres vivants, liés au *sacré*. « Il y a des voix qui parlent dans les arbres, dans les ruisseaux qui coulent ». Baudelaire reprendra ce thème dans ses *Correspondances*.

La notion d'intelligence artificielle va apparaître en 1950 : elle est « l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence ». Elle permet d'obtenir des progrès spectaculaires en médecine et en chirurgie, voire en psychiatrie. Elle rencontre un « succès fou », tout le monde en parle. Mais elle contient aussi le mot *artifice*, apparence, ruse en vue de tromper, car elle peut cacher une volonté de puissance en véhiculant des *concepts* métaphysiques, *cherchant à expliquer les secrets de l'Univers afin de*

conquérir l'immortalité. Elle peut engendrer la soif de Pouvoir d'un tyran, celle du docteur Folamour par exemple.

L'INTELLIGENCE ARTIFICIELLE

Comme l'intelligence naturelle, l'intelligence artificielle permet de comprendre la pensée et ses représentations abstraites que sont les idées. L'apport des neurosciences est considérable : elles ont permis la connaissance du cerveau doté d'un cortex préfrontal, de neurones miroirs (1990), donnant l'image de soi-même, prenant en compte l'image renvoyée par les autres, les milliards de synapses relayant les informations. L'intelligence artificielle, avec ses algorithmes, nous plonge dans ce monde numérique. Dès 1938, l'électroencéphalographie et l'électrochoc permettaient le repérage des signaux neuronaux et le traitement éventuel des anomalies. Actuellement, les progrès sont chargés d'espoir.

En effet, les neurosciences permettent de déceler les mystères de la transmission des idées pour effectuer des tâches, des actions, grâce à l'interface cerveau/machine, ordinateur, comme le révèle l'article de l'Inserm, « Agir par la pensée », paru en 2017, faisant référence aux travaux de François Berger, directeur du laboratoire de technologie Translationnelle Clinatéc (CEA/CHU de Grenoble / Inserm / Université Joseph Fourier, Grenoble) et Jérémie Mattout, chercheur dans l'équipe Dycog du Centre de recherche en neurosciences de Lyon (Inserm/ CNRS /Université Claude Bernard, Lyon).

Elle fait partie du monde rationnel, mais aussi irrationnel.

La pensée rationnelle, scientifique, objective

Elle regroupe l'ensemble des connaissances et études d'une valeur universelle ; elle est caractérisée par un objet et une méthode fondée sur des observations objectives *vérifiables*. Le réel est ce qui existe ou a existé véritablement, ce qui peut être

démontré par des raisonnements rigoureux et des expériences. C'est le monde de la Raison, des idées logiques. L'intelligence artificielle, par ses algorithmes est facteur de progrès.

Elle fait partie également de la pensée irrationnelle

C'est le domaine affectif

Le monde du ressenti, du subjectif, des sentiments, des émotions, des passions, des croyances qui sont déclenchées par la perception que nous avons des situations et événements qui nous entourent. Le symbolisme exprime à travers contes, légendes, mythe, archétype, ce monde de la pensée affective

C'est le monde de l'Imaginaire

L'Imaginaire est le système organisateur d'images donnant la capacité à un groupe ou à un individu de percevoir et de se représenter le monde, et de le réinventer, de façon merveilleuse parfois prodigieuse, fantastique. Il privilégie l'analogie, le rapport de ressemblance, la métaphore. Il est de l'ordre du sentiment et non de la raison. Il s'exprime dans les beaux-arts. Mais la fonction d'imagination est pourtant maîtresse d'erreur et de fausseté, car la raison y est souvent absente ou dominée par des croyances.

C'est le monde de l'inconscient

Il se manifeste dans nos rêves, dans nos complexes, dans les psychoses, nos fantasmes, nos désirs refoulés. Il est accessible par la psychanalyse fondée par Freud. Mais l'influence de Jung est évidente. Pour Jung, l'inconscient est collectif. Il correspond au dépôt des expériences ancestrales, quels que soient le lieu et l'époque. Il influence l'imaginaire et le comportement humain. Il est fondé sur les archétypes ce « fonds d'images anciennes qui appartient au trésor commun de l'humanité », qui sert de modèle de comportement et que l'on retrouve dans les contes et légendes, dans les mythologies, les rêves, la Mère et le pays natal mais aussi les délires (mythe

du héros Soleil, de l'or, de Sirius, mythe du Dragon). L'intelligence artificielle en est aussi influencée.

Les pensées rationnelles et irrationnelles doivent s'équilibrer.

Le réel, l'imaginaire et le symbolique doivent être intriqués. L'imagination a sa place, mais sans verser dans des chimères, la Raison ne doit pas écraser l'imaginaire, nier l'inconscient individuel et surtout collectif. L'intelligence artificielle peut être tentée d'en faciliter la coexistence et de participer aux moyens thérapeutiques en enregistrant les signaux idéiques et en tentant de les redresser.

Entre raison et l'imaginaire irrationnel.

La personnalité est l'organisation dynamique construite sur le capital neurobiologique et adaptée aux exigences socio-culturelles ; elle s'exprime à travers les comportements et la relation, mais les normes restent sujettes à caution. Elle assure l'équilibre psychique de ce couple d'opposés. Elle doit être harmonieuse et pour cela, elle exige la conscience de soi, le sentiment de sa propre existence et de celle d'autrui, qui débouche sur la pensée éthique, voire le sacré, ce qui dépasse l'homme, et donne un sens à sa vie. L'intelligence artificielle peut y contribuer.

Le comportement

Il désigne classiquement ce qui est observable, indépendamment des pensées. Mais il doit être compris.

Ainsi, l'intelligence artificielle peut permettre à l'intelligence normale de mieux s'adapter à des situations ou des problèmes nouveaux, favorisant le Progrès.

DISTORSIONS

Origine des distorsions

Les anomalies des idées et du comportement sont dues à des distorsions qui surviennent au cours de l'enfance en raison de l'apparition de schèmes cognitifs et affectifs issus de l'environnement. Les schèmes sont des structures d'ensemble d'un processus qui débouche sur un projet : le développement psychomoteur de l'enfant, puis de l'adulte. Les schèmes sensori-moteurs sont déclenchés par des postures inadéquates, entraînant une assimilation d'expériences douloureusement vécues. Les schèmes émotionnels sont créés par la répétition de remarques désobligeantes, des échecs au cours de l'enfance, une surprotection parentale, qui fragilisent la chaîne associative et la déforment si l'accommodation s'avère déficiente. Ces schèmes sont conservés dans la Mémoire. Ils demeurent dans les soubassements de la personnalité, comme ces plans des temples de l'antiquité, retrouvés sous les pieds des colonnes, pour servir d'arc-boutant à l'édifice de la personnalité, la « casa » de Freud, parcourue par les idées, les sentiments, les actions et susceptible d'être fragilisée, déformée, dépersonnalisée.

Conséquences

Le sujet devenu adolescent et face aux événements vitaux, peut devenir dépressif, inférieurisé, ou au contraire tout-puissant, peut voir la frontière avec la réalité et le monde extérieur s'altérer. La conscience de soi, l'introspection deviennent des sources de souffrance. Ces déformations d'origine affective, cognitive, déterminent

Des symptômes névrotiques, angoisse, phobie, obsession

L'angoisse : sentiment pénible d'attente d'un danger mal définissable. La phobie est une crainte angoissante déclenchée

par un objet ou une situation qui symbolise la punition d'une pulsion refoulée. L'apprentissage et le conditionnement peuvent jouer un rôle. L'obsession est l'irruption dans la pensée d'une idée, d'un sentiment, d'une tendance impulsive, en désaccord avec le Moi conscient, assiégeant le sujet, sont par leur chronicité, des facteurs de souffrance passant inaperçues pour l'entourage. La phobie est la crainte angoissante déclenchée par un objet ou une situation objectivement peu dangereuse.

Les troubles de l'humeur

La dépression et surtout la mélancolie, avec ses idées de culpabilité, d'indignité, de deuil et de ruine, de possession.

Des troubles psychotiques

La croyance délirante est une conviction inébranlable, inaccessible à la critique et au raisonnement. Les mécanismes idéiques sont l'hallucination, la fabulation, l'intuition, l'interprétation, la projection. La nosologie décrit alors :

La schizophrénie : c'est un délire chronique à thème de persécution, d'influence, de transformation corporelle, de possession. L'absence de structure rend le délire flou, le discours désorganisé, la pensée discordante, le comportement bizarre, coupé de la réalité.

La paranoïa : sur une personnalité marquée par l'orgueil, la psychorigidité, la fausseté du jugement, la méfiance, l'apparition de troubles passionnels (érotomanie, jalousie, revendication) ou d'interprétation délirante à thème de persécution signent la psychose. *Freud* a souligné l'inflation des coïncidences et du narcissisme, liés à l'absence d'objet d'amour. Les pulsions agressives seraient projetées sur des persécuteurs substitués de l'image maternelle.

La paraphrénie est un délire d'imagination, les idées de grandeur et de filiation trouvant dans des constructions

imaginaires une compensation à des frustrations infantiles.

Les psychoses réactionnelles brèves (bouffées délirantes), les troubles induits par des substances psychoactives, entraînant des troubles de la conscience, de la vigilance et de la synthèse mentale.

Des troubles de la personnalité

Déséquilibre psychique, avec comportement antisocial, perturbateur, agressif, ses passages à l'acte, paranoïaque avec orgueil, fausseté du jugement, méfiance, source de comportement conflictuel.

Les démences, les troubles cognitifs

Caractérisés par l'altération de la mémoire, l'aphasie, l'apraxie, l'agnosie, mais aussi par la fabulation, la persécution, les troubles de l'identité de l'Alzheimer, entraînent la dissolution des fonctions psychiques supérieures, comme les fonctions du réel, les fonctions d'intégration et de synthèse de la conscience, libérant les forces pulsionnelles sous-jacentes.

L'INTELLIGENCE ARTIFICIELLE PEUT PARTICIPER AUX TRAITEMENTS

1- au niveau des thérapies comportementales et cognitives : les dysfonctions comportementales peuvent être observées soit directement, soit indirectement par la verbalisation de la pensée, des sentiments, le but étant de modifier les *apprentissages* responsables des comportements déviants.

2- la lutte contre le passage à la chronicité : le maintien des automatismes, des habitudes de la vie sociale, la lutte contre les croyances irrationnelles, peut utiliser l'aide fournis par les robots humanoïdes, capables de contrôler l'économie de jetons.

3- au niveau de la recherche en neuroscience : des progrès pourront être obtenus grâce à l'apport de l'intelligence artificielle qui, permettra de détecter les signaux cérébraux pathologiques, témoins des distorsions idéiques ?

Des tentatives de « redressement » des idées déviantes, hallucinations, fabulations, interprétations seront peut-être possibles un jour, après enregistrement de signaux pathologiques et correction par stimulation de zones repérées ou correction des chaînes d'ADN. Pour l'instant, il s'agit de science-fiction, mais les recherches avancent.

CONSÉQUENCE AMBIVALENCE DE L'IA

L'apparition de l'intelligence artificielle accentue la puissance de la Science, mais fait aussi « rêver ». Elle peut apporter des solutions, mais aussi peut accentuer les risques de déstabilisation, par son ambivalence.

Dans son aspect technique : l'IA est essentiellement rationnelle

L'IA correspond à un ensemble de concepts et de technologies plus qu'à une discipline autonome. Elle est rationnelle, objective, augmente le potentiel humain, compense les handicaps. Souvent classée dans le groupe des sciences cognitives, l'intelligence artificielle fait appel à la « neurobiologie computationnelle ». Elle recherche des méthodes de résolution de problèmes grâce à l'algorithme. Ces techniques peuvent s'appliquer à l'apprentissage, participer au cognitivisme et sont facteurs de progrès :

Dans le domaine médical

L'intelligence artificielle anime les robots chirurgicaux, les techniques d'imagerie, la recherche.

Le risque de dérive

Les dérives peuvent survenir en raison de l'autonomisation éventuelle des robots. Les robots humanoïdes, devenus de véritables aides – soignants dans certains services de psychogériatrie peuvent prendre notre place relationnelle. En effet, certains chercheurs estiment que ces robots pourront acquérir la conscience de soi, apprendre à désirer et à envier, ce cancer de la relation humaine.

Dans le domaine psychiatrique, l'enregistrement des idées, ce qui devient possible en laboratoire, peut amener les psychiatres à des tentatives de correction de ces idées fixes, déviantes, voire délirantes soit par des procédés d'apprentissage issus des thérapies cognitive et comportementale, soit par stimulation d'électrodes intra ou extra cérébrale qui cherche à modifier le signal pathologique.

Le risque de dérive n'est pas négligeable, d'autant que la psychiatrie est parcourue par des courants divergents. L'antipsychiatrie conteste le terme de maladie mentale, attribuant l'origine des troubles à la société. Elle se méfie des contraintes que fait peser l'intelligence artificielle sur la formation des soignants et sur la prise en charge des futurs patients des institutions psychiatriques ou gériatriques.

Là aussi, une surveillance éthique s'avère nécessaire.

Dans son aspect philosophique

Les dérives peuvent survenir en raison de l'utilisation de l'intelligence artificielle par certains courants philosophiques. Devant les avancées de celle-ci, la tentation de percer les secrets de l'univers et du sacré est grande. Les désirs de toute-puissance peuvent surgir, notamment centrés sur la recherche de l'immortalité, la vie perpétuelle disposant d'un corps resté jeune. Deux courants peuvent être concernés :

1- Certains mouvements humanistes apparus vers 1920, proposent une réflexion sur les moyens pouvant améliorer la

condition humaine par augmentation des capacités de l'homme. Ils estiment que la science et les techniques sont capables de lutter contre la maladie, le handicap, la malnutrition et la pauvreté. Pour eux, créer la vie, échapper au vieillissement et à la mort est une fin essentielle. Leur but est louable, mais nécessite un encadrement. Car l'augmentation du nombre des personnes âgées risque de déséquilibrer la démographie.

2- Le mouvement « New Age » cherche à réaliser des progrès en médecine, notamment concernant l'autisme ; surtout, il propose un syncrétisme des religions. Et cherche à résoudre les problèmes écologiques face à la croissance industrielle qui menace l'avenir de notre planète Gaïa. L'intelligence artificielle peut les y aider. Mais la fuite dans les espaces interstellaires est-elle la solution ?

CONCLUSION

L'Intelligence artificielle va être écartelée entre les deux positions, rationnelle et irrationnelle. Ceci exige l'équilibre.

L'intelligence artificielle n'est pas « une folie ». Elle doit être un « gai savoir », selon l'expression de Stendhal, reprise par Nietzsche, et non une métaphysique, encore moins un fanatisme. L'Intelligence Artificielle doit rester technique, créatrice et refuser ce versant métaphysique, et sa volonté de puissance. Son rôle dans le diagnostic et la thérapeutique des troubles mentaux est intéressant. Elle est chargée d'espoir. Son apparition est un bienfait.

Mais elle ne doit pas être source d'entrave vis-à-vis de la liberté psychique, ni l'enjeu d'une philosophie messianique et utopique.

Les interrogations fondamentales sur la mort, l'âme, la métaphysique, sur notre destinée, cette puissance souveraine

qui mène le cours de l'histoire, persistent de nos jours. Les défenses habituelles restent la fuite en avant dans le divertissement selon Pascal ou Giono. Certes, notre société est menacée par les guerres, la pollution, la mondialisation ; le progrès scientifique peut paraître menaçant. Mais l'imaginaire collectif revient toujours en ressac et rétablit l'équilibre. Les droits de l'homme doivent être rappelés.

En fait, le danger vient de la superstition, cette croyance irraisonnée et du fanatisme, la passion du croyant, dénoncé par Voltaire. Le désir de remplacer Dieu est un thème retrouvé dans les délires. L'intelligence artificielle peut y contribuer, donnant l'illusion que l'Homme peut tout résoudre. Par contre, la conception d'une l'âme immortelle, immatérielle conçue par Platon n'intéresse que peu actuellement, car une partie de la population préfère privilégier l'immortalité du corps en ces périodes matérialistes plutôt qu'investir son avenir dans une âme invisible.

Le risque peut résider dans la personnalité pathologique du chercheur, ce qui est heureusement très rare. La mégalomanie de la personnalité paranoïaque, faite d'orgueil, de mépris, de méfiance, ce qui peut correspondre à des frustrations et des projections de désirs inconscients sur la réalité, peut se cristalliser et devenir conviction délirante, L'intelligence artificielle devient une arme entre ses mains. Une dérive sectaire reste toujours possible.

Ainsi, la communauté humaine doit rester vigilante en proposant des réserves éthiques devant les dangers de déshumanisation face à la robotisation et les risques d'escroquerie face aux promesses chimériques !

Il faut donc que l'intelligence artificielle se démarque de cette démarche métaphysique, d'autant que l'attrait actuel pour le paranormal, le surnaturel, l'occultisme, les fausses sciences

progresses malgré les avancées de la Science, témoignant de l'angoisse existentielle de nos contemporains.

Le rôle de nos Académies, qui regroupent des personnes d'orientation différente et complémentaire pourrait être de signaler d'éventuelles dérives ou de donner des avis au Comité consultatif national d'éthique, créé en 1984, lorsque des déviations dangereuses sont constatées, traduisant un déséquilibre entre le rationnel et l'irrationnel. Respectons la science, mais laissons un peu de place aux Muses !

BIBLIOGRAPHIE

BOUCHARLAT Jacques, « Folie et intelligence artificielle », *colloque de l'Académie Delphinale*, 2011, (à paraître).

CROCQ Marc-Antoine, GUELFY Julien-Daniel, BOYER Patrice, PULL Charles-Bernard, PULL Marie-Claire, *DSM-5 Manuel diagnostique et statistique des troubles mentaux*, Paris, Masson, 2015.

GOLDMAN H H., *Manuel de psychiatrie générale*, Piccin, 1996.

LEMPERIERE André, *Psychiatrie de l'adulte*, Masson, 2006.

IRM fonctionnel. Wikipédia.

L'ART AU CROISEMENT DES TECHNOLOGIES ET DE L'HUMANITÉ

*par M. Richard FORESTIER
de l'Académie des Sciences, Arts et Belles-Lettres de
Touraine*

INTRODUCTION

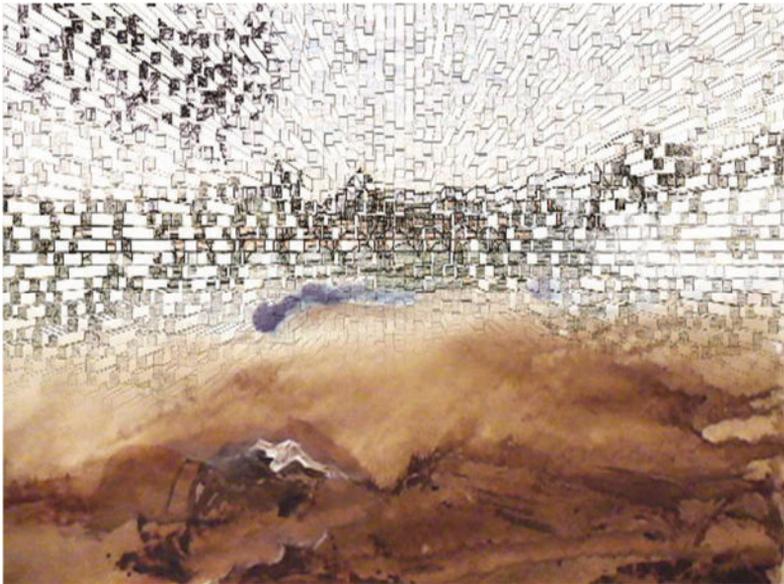
La bienveillance envers autrui est une détermination de l'humanité. Cette qualité humaine participe aux fondements de l'Humanisme. Par nature l'être humain vit. Pour vivre il doit animer ses capacités, facultés et potentialités. Soit qu'il les extraie de lui-même et les exploite de façon autonome comme pour l'Art, soit qu'il bénéficie d'aide pour les exploiter avec les technologies par exemple. L'Art favorise l'expression humaine et les technologies en assistent les manifestations.

En puisant dans les fondements de la nature humaine, l'Art met en exergue la complémentarité qui s'instaure entre la raison et l'émotion. La dynamique qui en résulte favorise l'animation humaniste.

On doit à Léonard de Vinci l'élan d'une science originale qui conjugue la part scientifique inhérente à l'Art avec l'originalité artistique dans l'épanouissement scientifique. Plus tardivement l'activité artistique est théorisée par Paul Valéry. L'épistémologie artistique s'élabore ensuite dans les sciences humaines et fait l'objet d'enseignements universitaires. « L'étude critique et analytique de la valeur, des concepts, des moyens et des fins de l'activité artistique »¹ définit l'épistémologie artistique. Science nouvelle, elle donne un cadre scientifique capable de circonscrire, déterminer, analyser et exploiter l'expérience artistique. Son originalité se fonde sur l'intimité des rapports qui s'instaurent entre l'Art et l'Humanité. De la caresse artistique superficielle au ressenti

1 FORESTIER, 2017, p. 161

existentiel des profondeurs humaines, il revient à la création d'initier l'expression humaniste que l'Art peut révéler. Un ensemble de nouveautés technologiques pénètre l'art contemporain. Mais toute innovation est-elle création ?



La Loire

Agnès FORESTIER, production artistique, technologie et peinture, 2021.

L'épistémologie artistique permet une approche nouvelle de l'Art. Une présentation de ce regard original s'impose d'abord dans mon propos pour comprendre l'expression de l'humanité qui caractérise l'Art². L'apport technologique est ensuite exposé.

² Remarque : nous distinguons l'art : savoir faire (a minuscule), de l'Art : domaine artistique (A majuscule)

L'ART EST UNE EXPRESSION DE L'HUMANITÉ

La technè au service de la beauté

De la *technè* dans la Grèce ancienne à la technologie du XXI^e siècle, l'art « savoir-faire » s'attache à la réalisation formelle des choses. Disposition susceptible de création, l'art ne se rapporte pas à l'action en tant que telle nous dit Aristote³. *Poïesis* et *praxis* amorcent leur distinction dans les activités humaines et le terme création recouvre aujourd'hui, autant l'acte producteur que le produit réalisé. L'artiste crée et la première au théâtre est une création. Dans un rapport où l'utile et l'agréable se confrontent, la beauté va s'orienter vers l'agréable. Avec l'évolution scientifique, philosophique et artistique, le Beau et le sentiment du Beau vont se différencier⁴.

Paul Valéry, scientifique, philosophe et artiste, élabore une théorisation de l'activité artistique. Les sensations initiales produisent d'autres sensations jusqu'à produire une sensation créatrice nous dit-il dans sa notion générale de l'art⁵. Les sensations conséquentes sont utiles. Elles ne sont que transitives et se transforment en représentations, décisions ou actes. L'œuvre d'Art en est une conséquence. Au regard des sensations initiales, à l'origine de cette chaîne sensorielle, leur nature uniquement existentielle détermine l'inutilité de l'œuvre d'Art. À la différence de l'art médical, de l'art équestre ou de l'art de se taire pour reprendre Paul Valéry⁶, l'Art, dont nous parlons ici se distingue de l'art « savoir-faire » assimilé à la technicité. On doit à l'esthétique la particularité de cette distinction. Paul Valéry hésite entre la science du Beau et la science des sensations pour aborder l'esthétique⁷. Il repère

3 ARISTOTE, 1965, p. 156-157.

4 DIDEROT, 1965, p. 417-419.

KANT, 1955.

5 VALÉRY, 1957, p. 14-12.

6 *Ibid* p. 1044.

7 VALÉRY, 1939, p. 236.

l'aspect agréable de la beauté et précise que le plaisir qu'elle procure est inexplicable⁸. En cela, le jugement du Beau a une valeur d'énigme⁹, ce qui n'est qu'une reprise du constat de Socrate sous la plume de Platon : « Je ne sais même pas ce qu'est le beau en lui-même »¹⁰.

La vie ne peut être connue que par le vivant. De même pour le Beau, réalité qui échappe à une définition véritable et définitive. En l'exprimant, son pouvoir devient opérant : « À force de construire [...] je me suis construit moi-même »¹¹. C'est à l'effort et l'organisation de l'activité productrice d'œuvre d'Art que Paul Valéry attribue la dénomination de « poïétique »¹², dont l'épistémologie artistique saura s'inspirer.

Les incidences du rapport de l'Art avec l'univers

D'une sensibilité créatrice au ressenti individuel et personnalisé, Paul Valéry juxtapose, voire conjugue l'inutilité de l'œuvre d'Art avec l'arbitraire de l'artiste¹³. En écho à l'intérêt que Paul Valéry porte à la création, les connaissances actuelles de l'astrophysique entraînent l'Art dans l'aventure universelle. L'œuvre d'Art s'inscrit dans un vide initial. Le vide, unité première de l'Univers, est à l'origine de l'évolution du monde. « L'histoire du monde débute par la déchirure du vide » nous rappelle l'astrophysicien Michel Cassé¹⁴. Le monde comprend l'ensemble des éléments qui le composent : par exemple les œuvres d'Art, les êtres humains ou leurs éléments contextuels. La forme de l'œuvre d'Art est soumise aux éléments extérieurs auxquels elle est confrontée, tel le soleil qui impacte la couleur d'une peinture ou la mauvaise

8 *Ibid* p. 263.

9 *Ibid* p. 240.

10 PLATON, (428/427-348/347 av. J.C.) 1967, p. 395-396.

11 VALÉRY, 1957, p. 86.

12 VALÉRY, 1939, p. 263.

13 VALÉRY, 1957, p. 1405.

14 CASSÉ, 1993, p. 169.

acoustique d'un auditorium qui pénalise un concertiste. De la même façon que des aléas extérieurs à l'œuvre peuvent avoir une incidence sur sa formation et sa présentation artistique, les qualités des matériaux qui la composent et l'organisent (l'usure d'un tissu ou la fragilité du verre) ont aussi une incidence sur le maintien et la préservation de la forme artistique. « La somme d'interaction des parties élémentaires » agit sur la bonne tenue de l'œuvre¹⁵ (exemples : le marbre de la sculpture n'a pas la même résistance que le bois, la constitution physique du danseur influe sur ses mouvements).

Le XXI^e siècle complète la poïétique de Paul Valéry par la considération en eux-mêmes de chacun des éléments qui composent le phénomène artistique. Avec l'épistémologie artistique, un apport scientifico-artistique nouveau amène à reconsidérer les repères conventionnels des relations qui s'instaurent entre les œuvres d'Art et leur public.

De la beauté à l'esthétique

Considérée en elle-même et pour elle-même, seule la présence de l'œuvre justifie son existence. En complément de l'inutilité précisée par Paul Valéry, l'œuvre d'Art est ainsi un « en soi » désintéressé. Ce désintérêt détonne dans l'ensemble des productions humaines. Il suscite par sa singularité, l'intérêt des personnes sensibles à cet état. Ce pouvoir spécifique de l'Art provoque des sensations particulières ou ressentis esthétiques. Déjà repérée au XVIII^e siècle, la relation entre l'esthésie en physiologie et ce type de sensations au contact de l'Art permet à Baumgarten l'invention en 1735 du concept d'esthétique¹⁶. L'extension conceptuelle relative à l'épistémologie artistique amène à concevoir le fait que c'est par son désintérêt que l'œuvre d'Art est intéressante¹⁷.

15 *Ibid* p. 33.

16 BAUMGARTEN, 1988.

17 FORESTIER, 2021, p. 23, section 7.

Le ressenti esthétique est muet écrit Paul Valéry¹⁸. Il est hors verbal, complète l'épistémologie artistique¹⁹. Comme pour le vide qui est plein (d'énergie par exemple), ce ressenti hors verbal est animé par une émotion. La sensibilité créatrice de la poïétique mêlée au vide de l'astrophysique est le terreau de la productivité artistique. En associant l'individu qui caractérise la détermination de l'espèce humaine à la personne qui singularise l'arbitraire, comprendre l'Art de la sorte est une façon de parvenir à ses fondements et d'accéder à l'humanité qu'il exprime.

DE L'ORIGINE ARTISTIQUE AUX TECHNOLOGIES CONTEMPORAINES

La créativité artistique s'extrait du néant

L'œuvre d'Art est le témoignage d'une empreinte humaine dans l'Univers. Extraction du néant, c'est un « en plus » dans l'Univers. En affirmant que le vide est non assimilable au néant²⁰, l'astrophysique contemporaine reprend et développe les idées de Léonard de Vinci telles que l'on peut les trouver dans ses carnets²¹. L'Art qui organise le chaos initial dans la pensée artistique de l'Antiquité grecque est un tremplin intellectuel pour une détermination du néant : l'élan scientifico-artistique y trouve son origine et sa particularité. En reprenant les idées de Léonard, le contact qui unit la surface de l'eau d'un lac à la surface de l'atmosphère qui la surplombe existe bien, tout en n'étant aucun des deux, ni leur mélange. Ce contact existe par son inexistence. Il en est de même pour l'instant qui n'est ni le passé, ni le futur, mais un présent fugace. Le point qui nous mène jusqu'à l'infiniment petit devient le passage de l'invisible au visible²². Ce creuset dans

18 VALÉRY, 1939, p. 15.

19 FORESTIER, 2021, p. 7, section 5.

20 CASSÉ, 1993, p. 162.

21 VINCI, 1987, p. 68.

22 *Ibid* p. 80-81, 390-391.

lequel l'existant s'extrait de l'inexistant détermine le néant. Cette extraction du néant définit la création. La créativité artistique en est la manifestation. Par exemple pour produire un son sur une guitare il faut établir un contact entre un doigt et une corde. Or le contact n'est ni le doigt, ni la corde, ni les deux. De ce contact, un élément inexistant devient un son, puis des sons qu'une musique organisera.



*La création musicale*²³

Sans répondre à la question initiale de l'origine du néant, la création devient un concept rassurant et apaisant pour l'esprit humain confronté à la quête existentielle de « l'origine de l'origine ».

L'Art manifeste la permanence de l'humanité

Le XIX^e siècle, dégagé de la mystique de la Renaissance, porte un intérêt scientifique et humaniste aux premières œuvres d'Art connues. Le rapport à la mort des êtres humains de la préhistoire révèle un questionnement existentiel relatif à l'aléatoire de la vie et à la permanence de l'humanité, ce que leurs productions artistiques semblent attester. Les maternités

23 FORESTIER, *Art et Humanité, Introduction à l'épistémologie artistique*, Tours, http://academie-de-touraine.com/wp-content/uploads/2021/03/Forestier_Art-et-Humanité.pdf, 2021, p. 25.

des vénus Préhistoriques de Montpazier en témoignent²⁴. Du symbolisme de Leroi-Gourhan, en passant par le mysticisme de l'abbé Breuil, c'est certainement à Sautuola, l'un des premiers préhistoriens qui constate la lointaine origine de la beauté des œuvres préhistoriques, que l'on doit la reconnaissance de l'émergence d'une expression de l'humanité dans les œuvres de cette période²⁵. Ensuite le passage de la nécropole au cimetière tend à promouvoir le respect de la personne pour une vie meilleure après la mort. Cela est entretenu par un symbolisme apotropaïque attaché au défunt. Il est possible de constater cela sur les sarcophages en plâtre du VII^e siècle de Saint-Denis. Avec l'intérêt porté à la santé, la vie de l'humain se distingue de l'existence de son être. Bien vivre et Bien être doivent s'accorder.

La technologie pousse l'Art dans ses retranchements

L'Art qui n'a jamais rien guéri exploite toute la puissance de son pouvoir dans le bien-être. Il peut donner l'envie de guérir. L'Art au service de la médecine révèle le rapport bien vivre / bien être. Les prothèses, les moyens technologiques peuvent remplacer, modifier, aider des parties défaillantes du corps humain. Mais la copie, le même, le substitut, le clone ne seront jamais l'original. L'Antiquité grecque l'avait déjà remarqué²⁶. Il est possible de remplacer un membre du corps humain, mais il est impossible de remplacer l'humanité d'une personne.

L'art, savoir-faire bénéficie de ces modalités technologiques nouvelles. Dans ce cadre, les techniques et technologies orientées vers l'Art peuvent faire évoluer les formes et les moyens artistiques. L'Humanisme qui s'exprime bénéficie des progrès technologiques tout en préservant le fond naturel de l'espèce humaine. Avec le développement des

24 A. C. WELTE, 2015, p. 33.

25 M. S. DE SAUTUOLA, 1880.

26 PLATON, 1967, p. 460.

connaissances scientifiques jumelées aux expériences et productions artistiques, l'esthétique trouve une envergure plus large et plus fondamentale autant dans les fondements artistiques que dans une participation à l'énergie humaniste des personnes. Les idées de l'esthétique traditionnelle sont repensées et l'exclusivité de la beauté conventionnelle qui la caractérise, est remise en cause dans ce domaine.



Imbroglia

RO TSAER Friedrich, production artistique, document personnel, 2021

L'INCIDENCE TECHNOLOGIQUE SUR L'ART

La technologie déstabilise l'Art conventionnel

De l'ordinateur quantique dans les œuvres de Libby Heaney qui mêlent les qubits aux particules, ou l'exposition immersive du centre d'Art numérique parisien qui plonge le spectateur dans les réalisations artistiques, jusqu'à l'innovation incrémentale du « corps transparent » d'Emmanuelle Lainé qui représente et organise dans une installation filmique les activités cérébrales des corps humains lors d'activités quotidiennes, l'apport nouveau de la technologie enrichit les modalités expressives artistiques. L'Art conventionnel en est bousculé.

L'œuvre d'Art numérique, totalement virtuelle *Everydays : the first 5 000 days* de l'Américain Beeple, n'existe que sous une forme « NFT »²⁷. Les œuvres immatérielles, uniquement disponibles en ligne valident leur nature artistique par un certificat d'authenticité infalsifiable d'appartenance unique et exclusive du produit original. Si la propriété et le placement financier animent souvent le désir des acheteurs²⁸, il n'en reste pas moins que l'unicité et l'exclusivité revendiquées de la propriété, renvoient à la singularité de la personne dans la communauté humaine. L'intérêt pour l'œuvre dépasse l'œuvre. L'implication dominante des activités mentales et psychosociales humaines dans la détermination artistique de l'œuvre deviennent une évidence aux dépens de la forme matérialisée.

Sous couvert d'originalité et parfois de provocation pour intéresser un public, une déclinaison de l'innovation technologique est exploitée par certains artistes. Nouveauté et création artistique tendent à se confondre. Quelques exemples illustrent cela : la prestation chorégraphique de Kaori Hito, danseuse à l'écoute d'un robot vocal qui lui impose une gestuelle, les « Bruyantes mécaniques » de Basanta, Keene, Bissière et le studio bruyant sont des objets interactifs et aléatoires qui expriment la conscience des machines. Un dernier exemple relevé à l'exposition du Palais de Tokyo en 2018, *L'ennemi de mon ennemi* de Neil Beloufa, est une association de reproductions, d'images et d'artefacts par des formes instables que des robots animent en permanence en respectant une logique déterminée par des algorithmes. Les objectifs de ces œuvres, énoncés par leurs auteurs, sont des interprétations libres, souvent intellectualisées, d'un ressenti à connotation artistique. Avec le désir d'une recherche technique innovante pour revendiquer l'origine créative d'une émotion,

27 Non Fongible Token.

28 Prix de vente aux enchères de l'œuvre de Beeple : soixante-neuf millions de dollars, 69 000 000 de dollars.

d'un style ou d'un courant, le risque d'une technocratie artistique se pose. Il semble nécessaire de distinguer l'Art de l'exploitation de son prestige. L'émerveillement que provoquent les technologies peut détourner l'activité de sa finalité artistique.

L'éthique est concernée par l'évolution artistique.

Inscrit dans une époque, l'Art ne peut pas s'extraire d'influences sociales, politiques ou culturelles. La seule subjectivité relative aux ressentis esthétiques ou la dénomination socio-politique de courants artistiques tel que « l'Art dégénéré » de triste mémoire, n'ont en aucune façon l'autorité pour la détermination artistique des productions. Toujours dans l'esprit de l'épistémologie artistique, il n'y a pas d'humain plus ou moins humain que d'autres, de même il n'y a pas d'œuvre plus ou moins artistique que d'autres. La qualité de l'œuvre, aussi relative soit elle, ne l'exclut pas du domaine artistique dans lequel elle s'inscrit.

Le caractère bienveillant de l'Humanisme imprègne l'Art. Libre à chacun de ne pas apprécier une œuvre. Cela peut influencer ou concerner la réputation de l'artiste mais ne pénalise en rien son humanité.

L'Art est toujours contemporain de son époque. L'apport technologique doit s'y soumettre.

L'épistémologie artistique précise et organise l'activité artistique

Qu'il nous vienne à l'idée de nier la nature artistique d'œuvres aussi surprenantes que celles proposées dans l'Art contemporain, rappelons-nous la remarque désobligeante en matière de dessin de Michel Ange à propos de Titien²⁹ ou celle de Platon qui affirme que Homère « le père de la tragédie » est un menteur³⁰. À la fluctuation passagère des avis humains

29 VASARI, 1986, p. 31.

30 PLATON, 1966, p. 127-363.

répond la permanence de l'humanité. L'existence et l'essence des éléments impliqués dans l'Art trouvent un point de convergence avec l'esthétique. Si l'Art a un pouvoir naturel, les moyens pour l'exprimer n'ont que des effets.

De la pratique « amateur » avec le peintre du dimanche aux concepts les plus abstraits comme le silence musical 4'33 de John Cage, l'épistémologie artistique préconise un principe télescopique qui anime le tout et ses parties. Il est instauré « par la cohérence entre la créativité primordiale relative à l'Univers et la créativité artistique relative à l'œuvre d'Art [...] les extrêmes sont reliés et ne peuvent se départir l'un de l'autre »³¹. L'épanouissement humain qui caractérise l'Humanisme trouve un cadre capable de circonscrire sa dynamique évolutive.

La poétique moderne concerne l'insertion de l'Art dans les activités humaines. Elle détermine la praxie artistique qu'une opération spécifique synthétise. Un tableau simplifié présente les principales caractéristiques et modalités humaines impliquées dans et par l'expression artistique.

Tableau 1 – « Implications artistiques humaines : Généralités »

		<u>Art</u>	<u>Implication sanitaire</u>	<u>Modes et Sites d'actions</u>	<u>Implication existentielle</u>
Être humain Être Humain (existentiel) (sanitaire)	Plait	Beau	Goût	Sensibilité Affirmation	Sympathie
	Fait	Bien	Style	Structure corporelle Confiance	Espoir
	Veut (heureux)	Bon	Engagement	Tonus existentiel Considération Estime Amitié (amour) (quantitatif) (qualitatif)	Fierté

FORESTIER Richard, *Art et Humanité, Introduction à l'épistémologie artistique*, Tours,
http://academie-de-touraine.com/wp-content/uploads/2021/03/Forestier_Art-et-Humanité.pdf, 2021, p.17

La science avec les technologies qu'elle génère, cherche un résultat exact, définitif si possible, et reconnu par la communauté scientifique, cela avec les meilleurs moyens pour y parvenir. L'objectivité est de mise. L'Art et les œuvres qui en émanent sont les meilleurs résultats du moment par des moyens justes et personnalisés, que certaines personnes ressentent et apprécient. Le caractère subjectif qui détermine les ressentis esthétiques s'avère être l'un des aspects objectifs de l'Art. Objectivité et subjectivité vont de pair en matière d'Art.

ART ET TECHNOLOGIE : PROSPECTIVES ET DÉRIVES

La technologie transforme l'activité artistique

Les œuvres du répertoire artistique classique sont complétées progressivement par des productions nouvelles. L'amplification et la saturation sonores informatisées en musique actuelle ou la radiographie numérisée travaillée pour exprimer la beauté intérieure d'un corps humain en arts plastiques sont des exemples d'apports nouveaux dans les formes artistiques.

Par son rayonnement, l'œuvre d'Art agit sur les émotions humaines. Manifestations de l'animation existentielle, les émotions bénéficient des incidences relatives au contact de ces nouvelles productions. L'animation en est stimulée.

Dans les rapports qui s'instaurent entre l'œuvre d'Art et l'émotion, les productions technologiques sont capables d'influer sur l'enchaînement des mécanismes humains impliqués dans ce genre d'activités, jusqu'à élargir le champ émotionnel et modifier l'œuvre d'Art classique avec le rayonnement qui la caractérise.

L'œuvre change de nature. Le public pénètre, intègre, crée l'œuvre. La proposition de l'artiste est une incitation à dégager l'Art de l'exclusivité de la forme sensible contemplée. Les frontières tombent. Les rapports entre l'artiste créateur, l'interprète, le public, l'œuvre et le contexte déterminent de nouvelles modalités artistiques. Les Subjectivités se mêlent pour fonder une créativité collective. L'esthétique y exprime toute son originalité dynamique et l'Art en assure l'impulsion humaine qui l'anime. L'Humanisme artistique est conforté. La grande fertilité de l'inventivité et de l'imagination humaines en donne les limites. L'inconnu de ces limitations est une stimulation à la créativité artistique. À cet inconnu répond le connu proposé par l'œuvre d'Art. La vie qui est à l'origine de l'œuvre inclut, de fait, ce connu. Or elle renvoie à son fondement existentiel que l'inconnu caractérise (l'origine de la création). Les extrêmes se rejoignent et s'inspirent.

L'incitation créative pour sortir du cadre, aller hors limites n'est pas toujours sans danger. Mêler l'œuvre avec son créateur est un risque de confusion entre l'empathie fusionnelle et la sympathie coalescente. « Les cent visages » du duo Scenocosme composé de Grégory Lasserre et Anaïs Met den Ancxt est une installation interactive visuelle/*morphing* interactif. Il s'agit d'une installation qui propose au public de toucher des images de visages, de les imbriquer les unes dans les autres, pour produire un nouveau visage aux traits évolutifs. Cette œuvre est contemplée par le public qui l'a créée. Tendre à la confusion de l'œuvre avec l'artiste, peut annuler la distinction fondamentale qui existe entre le créateur et ce qu'il a créé. Le double sens du mot création ne doit pas éliminer l'évolution de la réflexion humaniste, rappelée par l'expression courante dans les milieux artistiques : « Je fais ce que je suis, mais je ne suis pas ce que j'ai fait ».

Considérant l'étendue des possibilités technologiques, une mise en garde est nécessaire. Bien qu'au fil du temps, l'être qui existe se soit accordé à l'humain espèce vivante,

aujourd'hui l'être humain est une entité réelle conceptuelle et concrète. Désaccorder le bien-être et le bien-vivre rend possible et réaliste le fait d'être en bonne santé et malheureux. L'Art peut y remédier. Les pratiques artistiques en milieux de soins en témoignent.



L'Art auprès d'une personne ayant un grand retard mental,

Lucien, Document personnel, 1986.

Du dessin à l'histoire naturelle et à l'anatomie avec Cuvier, jusqu'au robot humanoïde de Jacques Vaucanson, l'essor de l'Art vers la technologie médicale se profile au XVIII^e siècle. Conjointement, les dessins du corps humain par les anatomistes sont parfois d'une beauté surprenante. Art et science se rejoignent. Ce n'est pas Jacques Lordat, médecin pionnier de la neuropsychologie, auteur d'un essai sur les rapports qui existent entre le dessin et la médecine en 1833, qui contredirait cette association³².

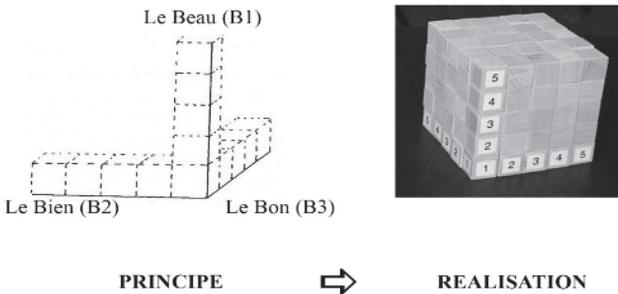
Une exploitation pratique de la technologie en Art

Les pratiques artistiques en milieux de soins se développent au XX^e siècle. Suite aux exigences médicales pour valider les effets thérapeutiques, le cube harmonique, principe évaluatif des capacités expressives de l'émotion artistique d'un patient, est inventé dans les années 1980 dans le cadre du Centre

32 LORDAT, 1833, p. 195-196-288.

d'études supérieures de l'Art en médecine en collaboration avec la Faculté de médecine de Tours. Premier outil du genre, le cube harmonique propose un paramétrage des trois aspects caractéristiques de l'Art : le Beau (plaisir), le Bien (fait), le Bon (moment heureux). Au contact d'une stimulation artistique, la personne soumise à cette évaluation signifie et réalise matériellement la représentation de ses cotations.

La forme cubique est un idéal théorique qui n'est pas un objectif à atteindre. Si l'émotion artistique est le substrat, la capacité expressive en est la manifestation et le seul intérêt de cette évaluation.

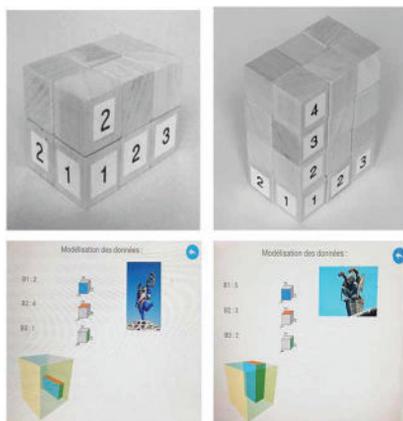


Cube harmonique : fondement théorique³³

Afin de rendre le cube harmonique attractif et concret, initialement une organisation des cotations par des petits cubes en bois compose une forme géométrique synthétique. Les cotations personnalisées s'adaptent aux décisions de la personne concernée. Généralement la variation des cotations est un signe salutaire.

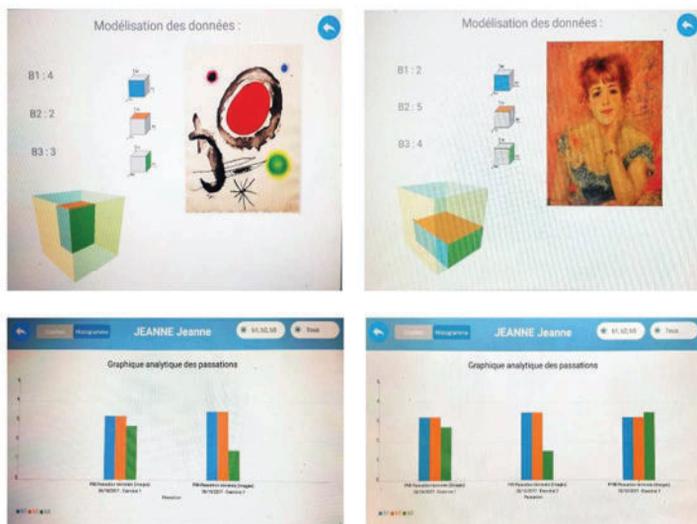
³³ FORESTIER Richard, *Art et Humanité, Introduction à l'épistémologie artistique*, Tours, p. 46.

http://academie-de-touraine.com/wp-content/uploads/2021/03/Forestier_Art-et-Humanité.pdf, 2021.



Application personnalisée du Cube harmonique

La difficulté pratique relative à la quantité et à la manipulation des cubes en bois s'avère être un frein à la bonne utilisation du cube harmonique. L'apport technologique d'une tablette numérique donne toute la pertinence à l'outil.



Cube harmonique tablette

FORESTIER Richard, document personnel, 2021.

Support unique, complet et général, association de fonctions (mémoires, historique, suivis, graphiques, statistiques, fiche thérapeutique, informations diverses...), la tablette est facilement accessible, manipulable, discrète et ludique.

L'ambiguïté technologique

Suite à de nombreuses passations, un fait qui intervient parfois retient l'attention. De l'objectif évaluatif, le cube harmonique peut glisser vers l'acte artistique tendu vers la réalisation d'une belle forme géométrique colorée que les cotations provoquent. L'outil provoque son propre détournement. L'expression de la personnalité s'estompe au profit des possibilités de l'outil.

De manière plus générale, du synthétiseur le plus sophistiqué jusqu'au téléphone portable de base, il est possible d'envisager la créativité et la réalisation artistiques. Cependant l'immense étendue des possibilités techniques, leur exploitation moins fastidieuse que les techniques traditionnelles, l'apport de nouveautés et leur caractère parfois ludique, font que la manipulation de l'outil informatique s'avère être souvent enivrante. L'ingénierie adaptée qui ne cesse de faire progresser les technologies est, parfois, une stimulation à une émancipation inventive propice à la réalisation de nouvelles formes. La pénalisation de l'autonomie des personnes au profit d'une dépendance aux moyens, est un risque tacite qu'il ne faut pas mésestimer. L'éventuelle servitude des personnes à de telles possibilités rappelle des états de dépendance technologique sous couvert de créativité artistique, que les milieux artistiques ou médicaux connaissent bien³⁴.

Un leurre créatif est susceptible d'envahir la personne, l'addiction au jeu informatique est quelquefois évoquée. Un questionnement s'impose : l'insertion technologique dans

34 HCSP, 2021.

l'Art peut-elle être source d'ambiguïtés, voire de tromperies dans l'implication et la responsabilité humaines ? Cela irait à l'encontre de l'éthique prônée par l'épistémologie artistique.

CONCLUSION

Avec l'évolution des savoirs, des progrès de la science et la singularité des formes artistiques contemporaines, la technologie pousse l'Art dans ses retranchements. Issus d'un creuset créatif commun, Art et science se distinguent dans leur finalité. Leurs moyens s'entrecroisent, la forme les rassemble, le fond les distingue. Dans un rapport qui harmonise le fond esthétique à la forme artistique, l'œuvre d'Art s'impose comme une manifestation privilégiée de l'Humanisme dans les activités humaines.

Confrontée à la variété, à l'abondance et à l'attractivité des modalités technologiques, la créativité artistique préserve naturellement l'autorité de l'essentiel sur l'artificiel. En atteste cette anecdote puisée dans le *Pantagruel*³⁵ de Rabelais d'un artiste qui meurt de rire devant l'une de ses œuvres. L'humanité de l'artiste prévaut dans son rapport à l'Art.

Enfin, après cette brève allusion à ma Touraine natale grâce à cet écrivain humaniste, le père de la « dive bouteille », rappelons que l'Humanisme qui bénéficie du respect pour valoriser la dignité humaine, peut trouver avec et par l'Art, un vecteur privilégié pour s'exprimer qui ne demande qu'à s'épanouir. Il peut changer les formes sans altérer le fond. L'œuvre d'Art en témoigne.

BIBLIOGRAPHIE

ARISTOTE, *Éthique à Nicomaque* VI, Paris, Garnier Flammarion, 1965.

BAUMGARTEN Alexander Gottlieb, *Esthétique*, Paris, De l'Herne, 1988.

35 RABELAIS, 1973, p. 631

- CASSÉ Michel, *Du vide et de la création*, Paris, O. Jacob, 1993.
- DE VINCI Léonard, *Les carnets*, T. 1, Paris, Tel Gallimard, 1987.
- DIDEROT Denis, « Recherches philosophiques sur l'origine et la nature du Beau » in *Encyclopédie* – article génie, œuvres esthétiques, Paris, Garnier, 1965.
- FORESTIER Richard, *Art et Humanité, Introduction à l'épistémologie artistique*, Tours, http://academie-de-touraine.com/wp-content/uploads/2021/03/Forestier_Art-et-Humanité.pdf, 2021.
- FORESTIER Richard, *Dictionnaire raisonné de l'Art en Médecine* CESAM, Lausanne, Favre, 2017.
- HCSF, *Analyse des données scientifiques. Effets de l'exposition des enfants et des jeunes aux écrans*. Rapport du Haut Conseil de la Santé Publique, Paris <https://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=1074>, 2021.
- KANT Emmanuel, *Le jugement esthétique*, Textes choisis, Paris, PUF, 1955.
- LORDAT Jacques, *Essai sur l'iconologie médicale ou sur les rapports d'utilité qui existent entre l'art du dessin et l'étude de la médecine*, Montpellier, Picot, 1833.
- PLATON, *Cratyle* 432d, Paris, Garnier Flammarion, 1967.
- PLATON, *Hippias Majeur* 306d, Paris, Garnier Flammarion, 1967.
- PLATON, *République*, t. X, 598e, Paris, Garnier Flammarion, 1966.
- RABELAIS François, *Pantagruel quart livre*, Paris, Seuil, 1973.
- SAUTUOLA Marcelino Sanz de, *Breves apuntes sobre alguna objetos prehistoricos de la provincia de Santander*, Santander, C. de la Real Academia de la Historia, Imp. Y lit de T. Martinez, Blanca 40, 1880.
- VALÉRY Paul, *Discours sur l'esthétique* in Variété IV, Paris, NRF Gallimard, 1939.
- VALÉRY Paul, *Eupalinos*, Paris, Bibliothèque de la Pléiade Gallimard, 1957.

VALÉRY Paul, *Notion générale de l'art*, Paris, Bibliothèque de la Pléiade Gallimard, 1957.

VASARI Giorgio, *Les vies des meilleurs peintres, sculpteurs et architectes*, t. X, Paris, Berger Levrault, 1986.

WELTE Anne Christine, *Le corps de l'homme* in Conférence nationale des académies, Paris, Akademos, 2015.

SYNESTHÉSIE

par M. Bernard DEVAUCHELLE,
de l'Académie d'Amiens

« Tu désires voir, écoute : l'audition est un degré vers la vision »

Bernard de Clairvaux

Confusion des sens chez le jeune enfant avant que ne se différencient certaines connexions synaptiques, substitution des sens chez l'adulte, la synesthésie est un phénomène dans lequel une stimulation unimodale conduit à une perception dans une autre modalité. Le poème « Voyelles » que Rimbaud publia en 1871, bien qu'il n'en était pas lui-même affecté, précède de peu le travail que Galton publie dans *Nature* en 1880¹.

Confusion des sens ou juxtaposition des sens ?

Illusion ou simple pouvoir métaphorique ?

Vertu ou handicap ?

La fréquence des synesthètes est évaluée entre 1/20 à 1/2 000 personnes, les excluant d'emblée de ce que l'OMS (Organisation mondiale santé) appelle « maladies » rares². Mathématiquement, il y aurait 120 formes cliniques différentes (factorielle 5^o). Selon Day³ (2004), la plus commune est la synesthésie graphème-couleur (prévalence de 68,8 %) qui, paradoxalement, ne lie pas deux sens différents mais superpose et confond forme et couleur dans une déclinaison différente. Des cinq sens, vision et audition sont le

1 GALTON, 1880.

2 Est considérée comme maladie rare toute pathologie malformative touchant moins de 1 personne sur 2000 dans la population générale (OMS).

3 DAY, 2004.

plus souvent associées. Mais toucher, goût et odorat offrent autant de combinaisons que chacun, sans nécessairement en avoir conscience, a pu éprouver.

« Plus qu'un autre sens, écrit Jean-Luc Nancy, la vision s'échappe au dehors et s'écarte de son lieu d'exercice alors que l'ouïe, le toucher, le goût et l'odorat appellent leur "dehors" à leur dedans et s'exercent entièrement dans un chiasme et une résonance de l'un et de l'autre »⁴.

Invitation à porter le regard sur un sujet dont la littérature est riche d'exemples, le propos ici tenu, après un détour obligé dans le monde de l'art, rapportera les liens spécifiques entre optique et haptique qui fondent le métier de chirurgien.

Neurophysiologistes, psychologues, généticiens, forts des progrès de la technologie, continuent d'explorer les voies impénétrables de cette « plasticité » cérébrale, cette « artificielle intelligence ». Invitation alors à nous tourner vers le philosophe. Privilège d'inconnaissance.

STOMA⁵

Détour apéritif dans le monde de la grande cuisine, Guy Savoye propose son menu : « couleurs, textures et saveurs ». Trois sens mélangés : vision, tact et goût dont la juxtaposition ne constitue pas à proprement parler une synesthésie. Certes, à se bander les yeux en mangeant, l'exacerbation compensatrice des autres sens donnera naissance à une image, mais cette compensation n'est pas nécessaire confusion.

Le Chef va cependant au-delà, en offrant une recette qui allie les cinq sens :

- a. Prendre une tranche de pain grillé ;
- b. La recouvrir d'une épaisse couche de beurre frais ;
- c. Déposer dessus un filet de sardine à l'huile ;

4 NANCY, 2014.

5 Stoma (στόμα) : la bouche, le trou (cf « Le visage organe » par B.Devauchelle in *Ruminatio* à paraître).

d. Saler et citronner légèrement.

L'expérience est savoureuse et concluante dès lors qu'on se prête au jeu, car la synesthésie s'éduque.

Dégustation multi-sensorielle. Les chercheurs de l'Institut Pasteur ont cheminé sur les voies de la gastronomie afin

« d'aider chacun à comprendre le fonctionnement du cerveau et des perceptions. [...] Un mets parle à la fois à nos yeux, à notre nez, à nos sens gustatifs et tactiles en bouche, aux sensations trigéminales et même à nos oreilles. Odeur et arôme différent par leur contexte sensoriel associé : une odeur qui arrive par le nez sera perçue en même temps que parviennent des informations visuelles, alors qu'un arôme naît en bouche, entouré de saveurs et de textures⁶ ».

Ces bases expérimentales posées, suit alors une longue description des dédales anatomiques des voies aérodigestives, puis des récepteurs gustatifs et olfactifs (400 récepteurs étalés sur quelques centimètres carrés de muqueuse nasale), avant que les fibres nerveuses issues des bulbes olfactifs se perdent dans les mailles inextricables des réseaux neuronaux et d'atteindre les centres dédiés (Gyrus, noyaux...). Constellation d'étoiles s'éclairant dans un espace-temps indéterminé. Prêtant à chacun des récepteurs sensoriels une identité génétique, suit alors en retour, une fois digéré par d'autres filtres, tout un vocabulaire descriptif alignant adjectifs métaphoriques relevant d'autres planètes (fugace ou long en bouche, lisse ou charpenté, maigre ou charnu, chatoyant ou mat).

Dans le même esprit, et poursuivant l'exercice d'Arthur Rimbaud, Françoise Héritier⁷ teste le goût des mots et leur musique :

- A) à la saveur amère du chocolat noir,
- E) la douceur sucrée de la vanille,

6 LEPOUSEZ, 2022.

7 HÉRITIER, 2013.

- I) a le feu du piment et du poivre,
- O) la rondeur de la cannelle,
- U) a le piquant des herbes fraîches parfumées.

Ailleurs :

- A) est grave et sombre comme une cloche d'airain,
- E) a le son limpide et fluide de la harpe
- I) le son aigu des pipeaux et crécelles,
- O) a le son plein et vibrant des cordes du violon, de l'alto et du violoncelle,
- U) a le déroulé ondoyant du piano.

La synesthésie se déclinerait donc dans tous les sens. Pour Arthur Rimbaud, « il s'agit d'arriver à l'inconnu par le dérèglement de tous les sens ». Retour au plus commun d'entre eux, la synesthésie graphème-couleur a fait l'objet d'un certain nombre de travaux en IRM fonctionnelle, prouvant avec réserve un modèle d'activation croisée des aires corticales. Mais cette proximité des aires cérébrales dédiées n'est pas applicable pour tous les sens. Et si pour Rimbaud, le « A » est noir, il revêt chez d'autres une couleur différente.

« L'ARC-EN-CIEL » DE NABOKOV

Dans un article paru en 1999 dans *Cerveau et Psycho*⁸, la synesthésie de cet écrivain est évoquée à travers son autobiographie :

« La sensation de couleur paraît être déterminée, chez moi, par l'acte même de former avec la bouche une lettre donnée tout en me représentant le tracé écrit. Le « a » de l'alphabet anglais a pour moi la nuance du bois sec, mais un « a » français évoque l'ébène poli. [...] Du fait qu'il existe une subtile interaction entre le son et la forme, je vois « Q » comme plus brun que « K », cependant que « S » n'est pas

⁸ DIEGUEZ, 1999.

le bleu clair de « C », mais un curieux mélange d'azur et de nacre [...]. Dans le groupe vert, il y a « F » feuille d'aulne, la pomme sure de « P », et « T » pistache [...]. Les jaunes comprennent différents « E » et « I », « D » crémeux, « Y » jaune d'or éclatant et « U » dont je ne peux exprimer la valeur alphabétique que par le ton cuivre jaune avec un reflet olive. Dans le groupe marron...⁹ »

Un autre article¹⁰ sur l'auteur peut une nouvelle fois poser la question de la différence de la nuance entre ce qui relève de simples métaphores et cette translittération littérale des sens que l'on retrouvera dans nombre d'écrits (Virginia Woolf, James Joyce, Charles Baudelaire...).

Lisons enfin l'orgue à bouche de Des Esseintes décrit par Joris-Karl Huysmans dans *À rebours*¹¹ :

« L'orgue se trouvait alors ouvert, les tiroirs étiquetés flûte, cor ou voix céleste étaient tirés, prêts à la manœuvre. Des Esseintes buvait une goutte, ici, là, se jouait des symphonies intérieures, arrivait à se procurer dans le gosier des sensations analogues à celles que la musique verse à l'oreille. Du reste, chaque liqueur correspondait, selon lui, comme goût, au son d'un instrument : le curaçao sec, par exemple, à la clarinette dont le chant est aigret et velouté ; le kummel au hautbois dont le timbre sonore nasille ; la menthe et l'anisette à la flûte tout à la fois sucrée et poivrée, piaulante et douce ; tandis que pour compléter l'orchestre, le kirsch sonne furieusement de la trompette ; le gin et le whisky emportent le palais avec les stridents éclats de pistons et de trombones ; l'eau de vie de marc fulmine avec les assourdissants vacarmes des tubas, pendant que roulent les coups de tonnerre de la cymbale et de la caisse frappés à tour de bras, dans la peau de la bouche, par les rakis de Chio et les mastics ! »

9 NABOKOV, 2021.

10 POPOVA, 2018.

11 HUYSMANS, 1884.

Substituant au monde de la littérature, celui de la musique et de la peinture (mais sont-ils si éloignés qu'ils émanent de la même griffe, de la même inscription sur la toile ?), explorons la relation amicale entre Arnold Schönberg et Kandinsky (lequel était synesthésique ?) pour mettre en avant la magnifique peinture de « l'âme », la *Gesamkunstwerk*, l'œuvre qui synthétise tous les arts.

Attirance réciproque précédant leur rencontre le 1^{er} janvier 1911, la veille du concert donné par le « Rosé Quartet » auquel Kandinsky assiste. Et quelques mois avant que ce dernier n'expose à Munich le *Cavalier bleu*, il invite le compositeur à publier dans sa chronique. Et c'est Schönberg lui-même qui apprendra à Kandinsky qu'il pratique la peinture. S'engageant l'un et l'autre pour la dissonance dans l'art, le peintre (également violoncelliste à ses heures) agrmente en compositeur musical quand le musicien apparaît en peintre expressionniste qui se laisse porter par son inconscient.

Ce croisement des regards, leur convergence suffisent-ils à appliquer le qualificatif synesthésique aux personnages quand l'un et l'autre évoquent la dimension spirituelle de leur art ? Correspondance, on renverrait alors au *Cri* d'Edward Munch qui fait porter spontanément les mains, non pas devant les yeux, mais contre les oreilles ou vers la peinture de Paul Klee qui maîtrisait à merveille le violon.

Reconnaissons enfin la contemporanéité (cette fin du XIX^e siècle, ce début du XX^e) de ce qui pourrait être lu comme une forme d'impressionnisme.

Déroulant alors le fil des sens, on déclinera à l'envi le travail de Michel Paysant, faisant de son œil, grâce à l'*eye-tacking*, le crayon qui dessine le portrait comme celui qui réécrit la partition, la main, le geste de Pierre Boulez dirigeant, dans ce qui pourrait être un sixième sens, son orchestre, ou Emmanuelle Laborit, sourde et muette, chantant son opéra en langue des signes.

PARÉIDOLIE

« La paréidolie est de ces vocations qui font entendre des voix dans le désert : elle assemble dans un furtif langage les fils éparpillés du réel¹² » écrit Stéphane Lambert.

Imaginer une anatomie particulière sous l'effet de *stimuli* tactiles et/ou auditifs, telle est cette forme de synesthésie qui envahirait notre monde médical et chirurgical.

L'exacerbation des sens fut longtemps le seul moyen dont disposait le médecin pour appréhender le corps malade. C'est de leur confrontation que naissait le diagnostic. Sans refaire l'histoire de la médecine, s'y sont substituées imagerie, biochimie, anatomopathologie. La cybernétique aidant, voilà le corps malade de plus en plus dématérialisé, éloigné du soignant, sujet devenu objet dans un double mouvement de déshumanisation que le vocabulaire trahit (malade → patient → usager) et d'atrophie progressive (lente) de nos sensorialités.

L'acte chirurgical fait-il exception ? Plus lentement, mais robotique, cœliochirurgie, téléchirurgie, requièrent certes une nouvelle gestique, mais risquent de remettre en cause moins un apprentissage qu'une sensorialité propre à son exécution.

Car la vertu du chirurgien, c'est cette capacité à unir formellement sensibilité tactile et vue, optique et haptique. Quelques extraits de *La fabrique du visage*¹³ :

- « Il faudrait donc revenir sur les liens qui se créent entre l'œil et la main du chirurgien dans l'acte opératoire et étudier dans quelle mesure leur union évidente ne repose pas sur un affrontement des sens. Car si la main opératoire avance sous la tutelle de l'œil, c'est pour trouver des solutions à jamais mystérieuses pour lui. Il pourrait être utile de stigmatiser

12 LAMBERT, 2021.

13 DELAPORTE, FOURNIER, DEVAUCHELLE, 2010.

cette distinction en réintroduisant un terme peu usité, « haptique », afin de désigner le toucher chirurgical d'une manière qui le rapproche et l'oppose à l'optique » (p. 257)

- « Affrontement des deux sens donc, mais aussi enchevêtrement dans ce qui serait une tactalisation de la vue et une visualisation du toucher, ce dialogue nécessaire entre optique et haptique définit l'acte chirurgical. Appliquée à la greffe du visage, cette inquiétante étrangeté, l'interrogation prend alors une dimension schizophrène, versant dans un vertige abyssal dont seule l'éthique, cette esthétique du dedans, évite l'engloutissement. » (p. 258)

- « Dans l'immédiateté de l'acte opératoire, la vue domine les autres sens, elle précède l'instrument et le guide dans les méandres de l'anatomie. Mais, rapidement, l'instrument, conscient de son pouvoir et de la suprématie de la vue sur la main chirurgicale, veut s'émanciper et tout à la fois en augmenter la potentialité et s'en affranchir. C'est ainsi qu'à la chirurgie immédiate, celle de la main prolongée de l'instrument, guidée à la vue et au su de l'opérateur, s'est substituée, au gré des progrès de la technologie, une chirurgie médiate qui interpose entre le sujet et l'opérateur un moyen (optique ou robotique) qui modifie la portée, la dimension de l'acte lui-même, et rompt ce lien somatognosique qui unissait jusqu'alors cette dyade. Un nouveau schéma corporel est à inventer qui se meut dans un nouvel espace dont le corps-même de l'opérateur et le corps de l'opéré sont absents,

matérialisés seulement par l'image de l'instrument et du champ auquel il s'applique. Qui n'a pas ressenti derrière le microscope cette perte de contrôle de la position des mains dans l'espace, il faut apprendre à être aveugle pour mieux voir. » (p. 262)

- « Tout ce qui précède relève de la vue, donc du visible. Demeure ce qui dans l'acte chirurgical procède du visuel, ce qui se voit au-delà de la vision, ce que Georges Didi Huberman appelle la visualité. Référence analogique faite à la Renaissance de Léonard de Vinci et du penseur Alberti, l'œil du chirurgien émettrait des rayons visuels qui s'entrecroisent avec ceux émis par les autres organismes et objets dans une sorte de vision suprême. C'est, dit autrement, une posture, plus qu'un regard absolu dont parle Foucault, cet œil qui parle, serviteur des choses et maître de la vérité et qui est vertu éminemment chirurgicale, indispensable à qui prétend reconstruire le visage, car il pose en filigrane la question du désir (le *Sehnsucht*), celui de la ressemblance et celui du devenir. Donner vision au regard que l'on porte (sur le malade, sur l'objet de l'intervention, sur le devenir de l'opération) constitue sans doute la vertu première du chirurgien. Elle ne lui est pas propre mais s'inscrivant d'une part dans l'instantanéité, d'autre part dans l'impossibilité du gommage du geste, enfin dans le caractère à chaque fois unique de l'acte opératoire, elle revêt une spécificité. Le détour par le philosophe et par l'artiste s'impose alors » (p. 263).

Un tel exemple de cette substitution des sens est illustré par la rhinoplastie quand, l'ostéotome glisse le long de l'orifice piriforme, guidé par le doigt à travers la peau, guidé aussi par le bruit ici mat, là plus creux quand il aborde les os propres, plus dur quand il arrive à l'os frontal. Les yeux fermés, on voit se dessiner le cheminement de l'instrument, glissant sous le périoste, évitant le canthus : l'œil du chirurgien n'est plus digital, aveugle en l'occurrence. Il est vision de l'invisible.

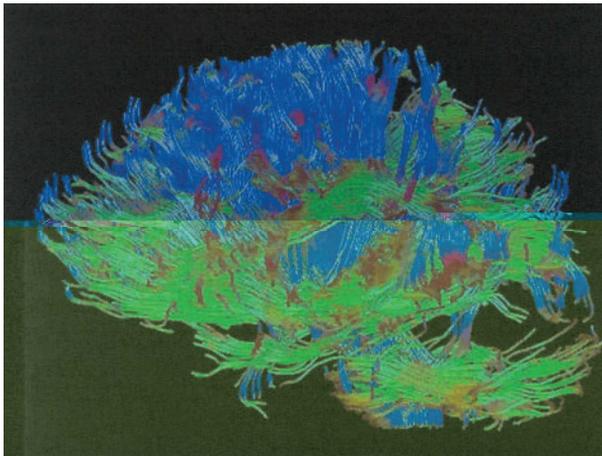
Synesthète de haut niveau ici, synesthète de bas niveau là, la synesthésie interroge neurophysiologistes et psychologues mais aussi généticiens dans leur recherche fondamentale. Faut-il voir la simple proximité d'aires cérébrales comme il a été dit quand vision et audition sont le plus souvent impliquées ? Faut-il évoquer chez les personnes qui prêtent couleur aux jours de la semaine un concept abstrait de séquence numérique qui suscite la couleur plutôt que l'apparence visuelle du signe ? La forme d'une ombre est reconnue dans le gyrus fusiforme, le message est relié vers le gyrus angulaire et le sillon intrapariétal. Le gyrus angulaire est impliqué dans le traitement des quantités et des séquences numériques et dans le raisonnement arithmétique abstrait. Peut-être que chez certains synesthètes, l'interférence se produit entre les deux régions de traitement de haut niveau de la couleur et du nombre dans le gyrus angulaire et non dans le gyrus fusiforme. Dans ce cas, ceci expliquerait pourquoi chez ces personnes la représentation abstraite du nombre ou l'idée de nombre évoquée par les jours de la semaine ou les mois de l'année, suscite fortement la couleur. En d'autres termes, il y aurait différentes formes de synesthésie selon les régions du cerveau dans lesquelles s'exprime le gène responsable. Cette hypothèse explique non seulement la connexion entre couleurs et lettres ou nombres mais aussi le rapport d'un synesthète qui dit ressentir un goût dans la bouche quand il entend des mots. Dans ce cas, cela serait dû au fait que le cortex auditif et le

cortex du goût sont adjacents respectivement dans le lobe temporal supérieur et dans l'insula.

De la même façon goûter des formes pourrait être dû à des connexions entre le cortex du goût dans l'insula et la représentation adjacente de la main dans le cortex somato-sensoriel. Dans le cas de la synesthésie, note de musique et couleur, les régions impliquées sont loin l'une de l'autre mais il existe des preuves que les connexions longue distance entre ces deux régions sont présentes à la naissance et si ces connexions demeurent en place, elles en expliqueraient persistance.

Science fondamentale, imagerie médicale, génétique se perdent donc en conjectures quand la synesthésie n'a aucun lien avec une pathologie. Métamorphose ou métaphore des sens, elle continuera d'interroger chacune et chacun d'entre nous, invitant à nous tourner à l'instar du philosophe¹⁴ vers les rapports entre le Je, le Moi et le Cerveau.

ILLUSTRATION



Visualisation en IRM de diffusion (DTI) des fibres axonales des réseaux neuronaux au sein des hémisphères cérébraux.

14 FOURNIER , 2012.

BIBLIOGRAPHIE :

DAY, Sean, « Some demographical and docio-cultural aspects of synesthesia » in *Synesthesia : Perspectives from Cognitive Neuroscience*, ed. D. Sagiv, New-York, NY Oxford University Press, 2004, p 11-33.

DELAPORTE, François, FOURNIER, Emmanuel, DEVAUCHELLE, Bernard, *La fabrique du visage*, Turnhout, Brepols, 2010.

DIEGUEZ Sébastien, « L'arc-en-ciel de Nabokov », *Cerveau et Psycho*, n° 20, mars-avril 2007, p. 90-93.

FOURNIER, Emmanuel, *Creuser la cervelle : variations sur l'idée du cerveau*, Paris, Puf, 2012.

GALTON, Francis, « Visualised Numerals », *Natrex 21*, 1880, p. 494-495.

HÉRITER, Françoise, *Le goût des mots*, Paris, Odile Jacob, 2013.

HUYSMANS, Joris-Karl, *À rebours*, Paris, G. Charpentier et Cie, 1884.

LAMBERT, Stéphane, *Pour Klee jusqu'au fond de l'avenir*, Paris, Arléo, 2021.

LEPOUSEZ, Gabriel, « Comprendre le cerveau grâce au vin », *Le Figaro* 19-20/02/2022, p. 29.

NABOKOV, Vladimir, *Œuvres romanesques complètes*, La Pléiade, Tome 3, Paris, NRF, 2021.

NANCY, Jean-Luc, *L'autre portrait*, Paris, Galilée, 2014.

POPOVA, Maria, « L'alphabet synesthésique de Nabokov : du bois patiné de A au nuage d'orage de Z », [Brainpickings.org/2018:05/15/nabokov-synesthesia](https://brainpickings.org/2018/05/15/nabokov-synesthesia)

.

ÉTHIQUE DE LA CONVERSION NUMÉRIQUE

*par M. Claude KIRCHNER
de l'Académie nationale de Bordeaux*

Notre monde et les humains qui le peuplent se convertissent au numérique. Comme l'explique Milad Doueïhi « une conversion exige un examen rétrospectif du passé (donc la réinterprétation et la réinsertion des anciens cadres et contenus dans les nouveaux) ainsi que de nouvelles explications des actes et des évènements ». Dans un tel contexte, notre représentation du monde change et induit des changements profonds. Les enjeux d'éthique de cette conversion sont les sujets de cet article.

NUMÉRIQUE ET ÉTHIQUE

Le numérique modifie profondément l'ensemble de nos sociétés et de nos cultures. Il a pour périmètre scientifique et technique, l'informatique, une partie des mathématiques, de la physique et des sciences cognitives ; il inclut la robotique et l'intelligence artificielle. Nous sommes tous, consciemment ou pas, volontairement ou pas, acteurs et spectateurs d'une profonde conversion numérique. Cette conversion s'appuie sur des évolutions technologiques majeures mais elle consiste surtout en une profonde révision de notre vision du monde. Ce changement se caractérise par une évolution de nos manières de penser, de nous organiser, de communiquer, de partager, de quantifier, de travailler et globalement de nous impliquer dans nos vies personnelles ou professionnelles. La pandémie due au SARS-CoV-2 a été un révélateur et un accélérateur remarquable de cet état de fait. Le numérique y a été crucial pour permettre aux plus jeunes de continuer à suivre leur classe à distance, aux plus seniors de communiquer avec leurs

proches, à tous de bénéficier directement ou indirectement du numérique pour communiquer, travailler, bénéficier des services publics et de la continuité des approvisionnements.

Un premier point remarquable de ces évolutions à l'échelle de l'histoire de l'humanité c'est la formidable rapidité de l'adoption du numérique. En moins de 25 ans (rappelons-nous que le terme « numérique » en tant que nom masculin est apparu à la fin des années 1990), le monde est devenu numérique au sens où la plupart des processus mis en œuvre par l'humanité ont intégré des procédés numériques. Ils se sont appuyés sur l'informatisation qui avait pris naissance lors de la seconde guerre mondiale en se basant sur les avancées scientifiques et technologiques de la première moitié du vingtième siècle.

Un second point remarquable est la pervasivité du numérique. En particulier les géographies physique ou politique n'ont pas le même impact sur le numérique que sur les biens ou les personnes. Même s'il est dit souvent que le numérique n'a pas de frontière, ce qui n'est pas toujours correct, il induit une révision profonde de la notion de souveraineté et impacte profondément les missions des nations habituellement considérées comme régaliennes. On se souviendra par exemple de la censure par Twitter de la communication d'un président des États-Unis d'Amérique ou de la mise en place de crypto-monnaies échappant totalement au contrôle des banques centrales nationales.

Du fait de la rapidité de sa mise en place, de l'ubiquité de ses usages et de sa relative invisibilité, la maîtrise collective du développement du numérique est difficile. L'émergence du RGPD est typique de cette situation où il a fallu réglementer à posteriori compte tenu des difficultés rencontrées sur le respect de la vie privée et des données personnelles. Cette démarche est plutôt salutaire lors du déploiement d'une technologie car réguler prématurément peut handicaper le marché et des

avancées positives. Mais dans le cas du numérique, le fait que des informations aient pu être diffusées est difficilement rétractable, voire impossible. Il est donc d'autant plus souhaitable de réfléchir à priori aux conséquences de la mise en œuvre potentielle d'applications numériques plutôt qu'en constater l'ampleur à posteriori. D'où l'importance de mettre en place des capacités de réflexion et de recommandation sur les enjeux d'éthique du numérique et de ses applications multiples.

L'éthique, pour reprendre la définition du *Dictionnaire* de l'Académie française, est un nom féminin caractérisant la « réflexion relative aux conduites humaines et aux valeurs qui les fondent, menée en vue d'établir une doctrine, une science de la morale ». C'est aussi un adjectif pour qualifier ce « qui a rapport aux conduites humaines et aux valeurs qui les fondent ».

Le comité national pilote d'éthique du numérique l'exprime dans son *Manifeste* de manière plus contextuelle et le décline pour l'éthique du numérique aussi appelée cyberéthique : « L'éthique est une réflexion continue, évolutive, nourrie par l'histoire de la pensée et liée à une culture. Elle identifie et interroge les valeurs et les normes, met en évidence leurs éventuels conflits, pour éclairer des choix individuels ou collectifs. Un questionnement éthique sur les sciences et les technologies peut se rapporter à des moments singuliers de l'existence humaine comme la procréation ou la mort, mais aussi concerner tout moment du quotidien. C'est le cas de l'éthique du numérique qui interroge en permanence nos valeurs à l'aune de situations inédites d'usage des technologies, avec une dimension radicalement nouvelle d'ubiquité et d'universalité de leurs effets ».

Les valeurs dont il est question ici peuvent être les valeurs républicaines françaises : liberté, égalité, fraternité. Ce peut être aussi les valeurs de la bioéthique élaborées au travers du

Code de Nuremberg en 1947, la déclaration d'Helsinki en 1964, le rapport de Belmont en 1978-79 et énoncés par les philosophes Tom Beauchamp et James Childress en 1979 sous la forme des quatre principes suivants, énoncés dans un ordre non significatif : (1) Principe de bienfaisance (*beneficence*), (2) Principe de non-malfaisance (*non-maleficence*), (3) Principe d'autonomie (*respect for autonomy*) et (4) Principe de justice (*justice*). D'autres valeurs peuvent également être considérées telles que la dignité humaine, la non-discrimination, la confidentialité, la confiance, la solidarité, l'équité. Quand on considère l'éthique du numérique, à ces valeurs que nous venons d'énoncer, pertinentes aussi dans le cadre du numérique, viennent s'ajouter des valeurs qui sont jugées particulièrement importantes, tant pour ce qui concerne la conception que les usages : la transparence, l'explicabilité, la durabilité ou encore la reproductibilité. La discussion de ces valeurs, leur hiérarchisation éventuelle, les liens entre valeurs, principes et exigences sont sujets de réflexion depuis au moins cinq millénaires. De manière plus récente l'explicitation des spécificités de l'éthique du numérique ont suscité des travaux tels que ceux de Raja Chatila ou Christine Froidevaux et Gilles Adda ou des cartographies des valeurs sous-jacentes à la réflexion éthique en numérique et particulièrement en IA (Intelligence Artificielle) telle que publiée en 2020.

ÉTHIQUE DE LA CONVERSION NUMÉRIQUE

L'expression « conversion numérique » introduite par Milad Doueihi est particulièrement bien adaptée pour décrire la transformation numérique que nous, humains, avons initialisée, que nous nourrissons et dont il est important d'appréhender et de mesurer l'impact.

Pour le comprendre, notons d'abord que nous sommes des systèmes de traitement d'information biologique – et que, bien sûr, nous ne sommes pas que cela. Comme l'explique remarquablement Michel Serres, toute entité, biologique ou pas, reçoit de l'information, émet de l'information, mémorise de l'information et transforme de l'information. Ces quatre

opérations fondamentales sur l'information, aussi souvent appelée « donnée », caractérisent les systèmes de traitement d'information (STI). Un caillou est un système de traitement d'information relativement élémentaire. Une cellule biologique et à plus forte raison un être humain ou un ensemble d'humains sont des STI absolument non élémentaires. La lecture de ce texte sollicite le système de traitement d'information du lecteur qui va recevoir, traiter, mémoriser et éventuellement retransmettre l'information lue. Avec l'invention et le développement du numérique, les humains ont créé de toutes pièces un nouveau type de système de traitement d'information. Comme nous l'avons vu en introduction, ce système a dans les quelque cinquante dernières années pris une place de plus en plus importante dans notre rapport à l'information. De son utilisation comme « machine à calculer » au cours de la seconde guerre mondiale pour aider au déchiffrement des communications ennemies ou pour mettre au point les armes nucléaire puis thermonucléaire, ses capacités de calcul, de mémorisation et de communication ont permis des traitements de plus en plus élaborés, rapides et ambitieux pour classer, comparer, évaluer et simuler des capacités jusqu'à la typique de l'intelligence humaine comme l'apprentissage ou la prise de décision en environnement incertain. Dans cette évolution, les relations entre les systèmes de traitement biologique et numérique se sont progressivement complexifiées. De l'utilisation par les humains des STI numériques comme calculettes, nous sommes passés à des utilisations dans lesquelles la machine augmente nos capacités de calcul et de traitement d'information pour passer ensuite à une situation où nous nous hybridons mutuellement. De manière intéressante, nous programmons les machines pour obtenir plus de services, qui nécessitent davantage de données que les machines « nous demandent » de manière plus ou moins explicite. De la coopération entre les systèmes de traitement d'information numérique et biologique nous passons à une hybridation des deux systèmes telle qu'il devient

difficile de ne plus pouvoir utiliser les STI numérique (STIN). Il est en effet difficile d'imaginer gérer la régulation de la circulation dans une agglomération comme Paris ou Tokyo sans numérique, de même que de ne plus avoir accès à notre smartphone ou de ne plus gérer numériquement la combustion des moteurs thermiques ou la distribution d'eau ou d'électricité. Nos mécanismes cognitifs sont modifiés et nous ne communiquons plus entre humains de la même façon, nous ne travaillons plus de la même manière, nous dépendons de plus en plus du numérique dans une boucle d'interdépendance et d'échanges réciproques dont la limite ne sera certainement pas atteinte à court terme. Nous avons donc créé une catégorie d'outils, les STIN, dont l'essence même est similaire à notre capacité de traitement d'information humaine et dont l'utilisation induit une conversion de nos capacités cognitives et de nos cultures pour aller vers une hybridation de ces deux modes de traitement de l'information, le numérique et le biologique.

Cette analyse suscite des questions multiples. Comment pouvons-nous caractériser cette hybridation et comment faut-il envisager ses évolutions ? Comment documenter et objectiver la prise de recul nécessaire à la maîtrise de la situation ? Les questionnements sur les enjeux d'éthique dans ce cadre sont multiples et nous en développons maintenant un exemple particulièrement important.

Enjeux d'éthique de l'intermédiation numérique

Les relations entre personnes jouent un rôle fondamental dans nos différentes cultures : l'humain est un animal social. Le numérique a pris – notons qu'il serait plus correct de dire que nous avons fait prendre au numérique – une place importante dans la gestion de ces relations. Le transfert d'informations entre deux personnes, d'abord principalement basé sur la voix et l'expression corporelle, a utilisé ensuite l'écrit puis le téléphone et maintenant l'image, les écrits, les sons, numérisés au travers d'outils comme les smartphones, le

courrier dit « électronique », les réseaux sociaux, etc. Cette numérisation de nos échanges n'est absolument pas neutre. En effet, le vecteur numérique utilisé pour communiquer est un système de traitement d'information, programmé avec des objectifs déterminés et dont toutes les finalités sont rarement rendues explicites et accessibles aux utilisateurs.

Cette programmation rendue possible par la numérisation est appelée dans de nombreux cas *ubérisation* du nom (Uber) de la plateforme de mise en relation entre les chauffeurs de taxi et leurs clients. Le remplacement du standard téléphonique d'une centrale de taxi par une plateforme numérique a transformé le métier de chauffeur de taxi ainsi que le comportement des clients. Elle a en fait changé profondément la mise en relation entre deux personnes, l'une ayant la capacité à transporter quelqu'un, l'autre pouvant planifier un tel transport avec ses outils de communication numériques usuels, un smartphone, un navigateur internet, etc. S'inscrire sur une telle plateforme d'intermédiation – programmée et supervisée par des humains – ne demande pas nécessairement d'avoir une licence de taxi : son rôle est juste de mettre en relation deux personnes pour qu'un service soit rendu. Ce service, sous le contrôle de la plateforme permet au-delà de la mise en relation de gérer l'échange d'autres informations telles que le coût éventuel du service rendu mais aussi la qualité de ce service et donc potentiellement l'évaluation par le client du chauffeur et l'évaluation par le chauffeur du comportement du client. L'utilisation d'une telle plateforme permet aussi d'optimiser le choix de la localisation du véhicule choisi pour minimiser le coût d'approche, en reliant aussi potentiellement le système de gestion de la plateforme aux autres systèmes de traitement d'information comme ceux qui gèrent les flux de circulation, etc. On voit comment l'idée simple, consistant à remplacer un service d'intermédiation rendu par des humains par un système numérique a des conséquences qui dépassent très largement l'objectif initial. L'innovation introduite transforme mais aussi convertit par son usage, elle convertit

les usagers d'un tel système avec son lot de possibilités mais aussi de contraintes, par exemple il est très commode de commander « un Uber » mais l'usage des données que l'on transmet à la plateforme lors de cette opération échappe largement au consentement explicite de l'utilisateur.

L'intermédiation dans de très nombreux domaines peut ainsi être numérisée avec des conséquences toujours aussi difficiles à apprécier que ce soit au moment de leur mise en place, à moyen ou long terme. Par exemple, l'intermédiation entre les personnes et les médecins a l'avantage certain de permettre de trouver le praticien adapté à tel ou tel besoin ou symptôme. Elle permet aussi de noter les praticiens comme les patients. Elle nécessite aussi une grande transparence sur la manière dont les praticiens sont proposés sur la plateforme aux différents patients : on sait que l'ordre de présentation est significatif, comment est-il déterminé par les algorithmes implantés sur la plateforme ? Par ailleurs, les données accumulées par la plateforme d'intermédiation sont révélatrices par exemple de la popularité des praticiens : ces données sont-elles utilisées, et si oui comment et avec quel objectif ?

Du point de vue de l'éthique, nous voyons sur ces exemples que les valeurs en jeu vont être le respect des données personnelles, la transparence des procédures utilisées, l'auditabilité des programmes mis en œuvre, la justice pour faire les bonnes associations patient-médecin sachant que, par exemple, les médecins n'admettent pas tous les mêmes moyens de paiements. Mais d'autres valeurs sont consécutives des transformations sociales ou des impacts environnementaux induits, citons-en deux, qui ne sont pas des moindres : l'autonomie (p. ex. des chauffeurs ou des patients) et la liberté (p. ex. pour tous les acteurs utilisant une plateforme d'intermédiation, il est difficile d'expliquer à une plateforme numérique la spécificité éventuelle d'une demande, avec des conséquences importantes pour les personnes en situation de

vulnérabilité). La régulation ou la législation qui apparaissent nécessaires pour gérer au bénéfice de tous la mise en place de ces plateformes doivent s'adapter aux cultures et lois locales alors qu'elles sont pour beaucoup bâties sur une base internationale, posant souvent des questions de souverainetés numérique, sanitaire ou culturelle.

Organiser la réflexion et l'éducation sur les enjeux d'éthique de la conversion numérique

Comme nous venons de le voir, la conversion numérique transforme notre représentation du monde dans de multiples aspects. Pour maîtriser cette évolution, il faut prendre en compte plusieurs éléments majeurs.

D'une part, la temporalité : nous ne sommes qu'au début de la transformation numérique et du fait de sa radicalité et de sa rapidité nous devons nous équiper intellectuellement pour penser sur le long terme tout en sachant prendre en compte les évolutions au jour le jour.

D'autre part, la modification radicale de nos approches réflexives. En effet, comme nous l'avons vu, les deux systèmes de traitement d'information biologique et numérique s'influencent : nous faisons évoluer le numérique et le numérique nous change. Cette boucle d'interaction, qui contribue à la richesse de la situation, ajoute à la complexité de la compréhension du phénomène et de son évolution. Elle ajoute aussi au fait que nos propres valeurs et leurs interrelations peuvent être modifiées par les valeurs implémentées, sciemment ou pas, dans les systèmes numériques que nous avons conçus et qui, par leur utilisation souvent quotidienne, nous influencent à leur tour.

Par ailleurs, la conversion numérique et son évolution ont un impact sur d'autres enjeux majeurs des prochaines décennies (pour le moins), la transition écologique et un impact environnemental, en particulier dû au développement du numérique dans tous ses aspects, notamment relatif aux objets connectés.

Enfin, l'ensemble des transformations numériques se joue

au niveau global de la planète et de l'ensemble de l'humanité dans une complexité géopolitique avérée.

Dans ce contexte :

- L'éducation au numérique doit être prise en compte dans sa globalité allant de l'école maternelle à l'université et, du fait de l'évolution rapide déjà soulignée, être instituée tout au long de la vie, par exemple comme élément constituant à part entière d'un projet parental. Il en est de même pour l'éducation à la réflexion éthique et à la pratique du débat et à l'argumentation rationnelle.
- La réflexion sur les enjeux d'éthique du numérique doit être promue localement (dans les entreprises, les associations, les territoires régionaux), nationalement avec notamment le développement d'un comité national d'éthique du numérique, consultatif et indépendant, et internationalement tout en sachant la difficulté intrinsèque à organiser de telles réflexions au niveau global. On pourra s'inspirer d'initiatives telles que le GPAI ou l'appui d'institutions comme l'Unesco ou encore en prenant modèle sur des initiatives comme le GIEC.
- Pour nous aider dans ces processus de réflexion et d'éducation, dont il faut souligner l'importante dépendance à la culture, il nous faut développer des éléments objectifs de compréhension et de mesure. L'élaboration d'observatoires des multiples développements du numérique sous toutes leurs formes, matérielles ou immatérielles, cognitives, environnementales est essentiel et pourra s'inspirer et développer les initiatives telles que celle de l'Obvia ou des recommandations existantes concernant la maîtrise du développement de domaines spécifiques émergents comme les réseaux de nouvelle génération (5G, 6G) ou l'internet des objets.

Le numérique et en particulier l'intelligence artificielle, est bien plus qu'un développement technologique de plus, il induit une profonde conversion de la représentation cognitive de notre monde. La maîtrise de cette transformation dont nous sommes tous acteurs repose sur le développement de notre compréhension de ses enjeux d'éthique.

BIBLIOGRAPHIE

BEAUCHAMP Tom et CHILDNESS James, *Principles of Biomedical Ethics*, 7th edn, Oxford University Press, 2013.

BERRY Gérard, *L'hyperpuissance de l'informatique. Algorithmes, données, machines, réseaux*, Paris, Odile Jacob, 2017.

CERNA, *La souveraineté à l'ère du numérique Rester maîtres de nos choix et de nos valeurs*, Octobre 2018. <http://cerna-ethics-allistene.org/Publications/%2bCERNA/index.html>

CHATILA Raja, « Bioéthique et éthique du numérique : une hybridation paradoxale », in *Pour une éthique du numérique*, Paris, Humensis, Presses Universitaires de France, 2022.

Comité national pilote d'éthique du numérique, *Manifeste pour une éthique du numérique*, 2021. <<https://www.ccne-ethique.fr/fr/actualites/manifeste-pour-une-ethique-du-numerique>>

GERMAIN Éric, KIRCHNER Claude et TESSIER Catherine, *Pour une éthique du numérique*, (dir.), Paris, Humensis, Presses Universitaires de France, 2022.

DOUEIHI Milad, *La grande conversion numérique*, La librairie du XXI^e siècle, Paris, Éditions du Seuil, 2008.

FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah, NAGY Adam et SRIKUMAR Madhulika, « Principled Artificial Intelligence : Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI », *Berkman Klein Center Research*, n° 2020-1, 2020.

KIRCHNER Claude, Anne FAURE, HARFI Mohamed, NABOULET Antoine et TRANIER Éva, *Le monde de l'internet des objets : des dynamiques à maîtriser*, France Stratégie, février 2022.

<<https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/fs-2022-rapport-iot-fevrier.pdf>>

FROIDEVAUX Christine et ADDA Gilles, « Regards croisés sur la cyberéthique et la bioéthique », in *Pour une éthique du numérique*, Paris, Humensis, Presses Universitaires de France, 2022.

LA RÉVOLUTION NUMÉRIQUE ANNONCE- T-ELLE UNE RENAISSANCE DE L'HOMME ?

par M^{mes} et MM. M. BIASI, E. CASSART, N. DEJEAN, H. El BOUSTANI, J. FERRASSE, D. GERADTS, G. GALANT, E. SURAUD¹ de l'Académie des Sciences, Inscriptions et Belles-Lettres de Toulouse

INTRODUCTION

En 2018 un jeune japonais a symboliquement épousé une image en trois dimensions, portant le nom d'Hatsune Miku². Quoi de plus virtuel ? Cela pourrait-il bientôt devenir banal, dans la colonne « Carnet » de notre quotidien favori, si tant est que la notion de journal acheté au coin de la rue ait encore un sens ? Certes, la culture japonaise, où le shintoïsme attribue, même aux objets, une dimension spirituelle, échappe en partie à notre vision occidentale du monde fondée sur une hiérarchie entre le vivant et l'inerte. Néanmoins, associer le terme de mariage, même symbolique, à une image ou une machine peut choquer. Serait-ce un signal, parmi d'autres, de la disparition programmée de l'Homme ? L'*hubris* du développement technologique et numérique finirait-il par déchaîner les foudres de Némésis au point de faire disparaître l'Homme ? L'Homme deviendrait-il un demiurge inquiet – peut-être même un Frankenstein terrorisé par sa création ?

Les craintes de l'Homme vis-à-vis des machines qu'il conçoit ne sont pas nouvelles mais semblent avoir pris un tournant incontrôlé avec la révolution numérique. Le

1 Représentant le groupe de travail interdisciplinaire des auteurs (région toulousaine).

2 https://fr.wikipedia.org/wiki/Hatsune_Miku.

développement exponentiel de la mal nommée Intelligence Artificielle (IA) bouleverse non seulement les modalités d'existence de toutes les sociétés, mais promet aussi la transformation physique des individus. Une post-humanité se profile... En moins d'une décennie l'IA a relégué, par endroits, le cauchemar prémonitoire d'Orwell à un jeu d'enfant et la science-fiction transhumaniste à une réalité à portée de main ! Devant l'ampleur et la rapidité des développements du numérique on peut se sentir « dépassé » et se demander si l'Homme ne serait pas révolu. Mais nous croyons que l'Homme a la capacité de coévoluer avec la révolution numérique, pour peu qu'il trouve sa place dans ce nouveau monde. Par ailleurs la science elle-même montre les impossibilités du tout calculable et donc les limites des IA. La révolution numérique pourrait donc finalement annoncer une renaissance de l'Homme. C'est l'hypothèse optimiste qui fonde cette réflexion.

Le texte est organisé en cinq sections. Après le constat alarmant de l'omniprésence du numérique dans notre monde (section 1) nous discuterons l'exemple de l'ambiguïté (section 2). La section 3 abordera la place de l'homme par rapport aux révolutions scientifiques et techniques. Nous proposerons ensuite des pistes pour positionner l'Homme par rapport à la révolution numérique (section 4). Enfin nous terminerons par une analyse critique en discutant des conséquences de l'existence du chaos et du hasard (section 5).

L'HOMME RÉVOLU ?

Doit-on se résoudre à un scénario catastrophe ?

Le scénario « catastrophe » de la disparition de l'Homme submergé par le numérique est courant. Et peut-être est-il effectivement fondé ? Un des mythes fondateurs de l'IA, et de façon plus générale du digital, est en effet sa supposée

infinie capacité de calcul, donc de prévision. Pour les tenants du tout digital il est « évident » qu'à terme tout deviendra calculable et donc prévisible. C'est juste une question technologique, une question de temps. Devant l'ampleur et la rapidité des développements du numérique on ne peut, de prime abord, que se sentir « dépassé » et adhérer à ce scénario catastrophe. Mais qu'advierait-il de l'homme dans un tel scénario ?

L'IA ferait disparaître l'Homme ! Elle le réduirait à un ensemble de données, indéfiniment moulinées par des supercalculateurs. Ces derniers décideraient de nos achats (à l'aide de suggestions en ligne lors d'une navigation sur leur site web), du lieu de nos vacances (en fonction par exemple des prévisions météorologiques, ou de goûts déduits de nos recherches sur internet) et du travail auquel nous pourrions prétendre (en fonction de notre capital génétique et des maladies que nous serions censés développer). Une oligarchie pourrait profiter du transhumanisme pour attendre, dans un luxe d'objets connectés, une mort tardive. Tandis que les masses, gavées de produits bon marché, feraient tourner l'économie de certains pays et de certaines entreprises privées à coups d'achats compulsifs, suite à des sollicitations ciblées, renforcées par l'obsolescence programmée des objets achetés.

Le débat politique aurait disparu, remplacé par des statistiques vertigineuses à coups de « j'aime », hors de tout contrôle et de toute réflexion. Et dans le cas improbable où apparaîtrait une idée non prévue par les IA, les milliards de caméras de surveillance vidéo (la Chine en recense déjà aujourd'hui plus d'un demi-milliard³) auraient tôt fait d'identifier le ou les auteurs de trouble, de les isoler et de les remettre au pas... d'une manière ou d'une autre. Bref, l'individu aurait perdu toute liberté et serait devenu une sorte

3 <https://www.wsj.com/articles/a-billion-surveillance-cameras-forecast-to-be-watching-within-two-years-11575565402>

de sous-machine, peu fiable (sujet à des pannes inattendues qu'ignorent les IA, sauf peut-être lors de coupures de courant ou de bugs informatiques), mais nécessaire comme « marché », à la demande des machines, et pour l'offre spéculative de ceux qui les conçoivent.

S'agit-il de science-fiction ? Non, d'une extrapolation tout au plus ! À chaque instant la vie ordinaire est accompagnée par les machines et leur arsenal algorithmique. Ces dernières gèrent insidieusement nos données issues de logiciels de surveillance et de reconnaissance faciale, nos modalités individuelles de consommation, notre santé... C'est au quotidien que la vie personnelle et collective est ainsi prise en charge, assistée par les machines : les logiciels suivent et analysent les comportements afin d'élaborer des stratégies de marketing. Dans le potentat de la victoire économique, l'humain est progressivement réduit à un vecteur de consommation, un rouage mécanique. Dans ces conditions, l'Homme tel que nous le connaissons aujourd'hui, tel que l'Humanisme le décrit encore, surprenant, incertain, énigmatique, imprévisible, perfectible, libre, soucieux d'une juste redistribution des richesses, soucieux de l'égalité qui le confronte à ses pairs, cet Homme disparaîtrait-il ? Serait-il tout simplement révolu, au sens fort du terme ?

Sommes-nous résolus par le calcul ?

Comment se résoudre à un tel scénario dans lequel les calculs singent la pensée humaine et où l'Homme y serait résolu au sens mathématique du terme ? Pourquoi résolu ? Parce que certains concepteurs de l'IA pensent que l'Homme est « calculable », c'est-à-dire réductible à un ensemble de données, un ensemble de chiffres que la machine engouffre et digère de manière toujours plus rapide. « Penser c'est

calculer » dit R. Da Silva⁴. Si la machine calcule si bien, cela tendrait à priver l'Homme de sa faculté la plus humaine et la plus spécifique, celle de penser ! Mais entendons-nous bien. Même si le calcul singe la pensée, nous n'en avons pas pour autant trouvé l'équation ultime, « l'équation de l'Homme » que l'on résoudrait à l'aide d'un supercalculateur. Non, pas d'équation. Juste un ensemble de données, d'informations que l'on recolle entre elles en espérant que ce déluge de *Big Data* sans liens apparents compensera, au poids, l'absence de compréhension globale.

Dans cette logique conquérante de l'IA, l'immense complexité de la vie et de l'Homme pourrait donc finalement être réduite, à terme, à un nombre colossal de calculs empilant des montagnes de données collectées tous azimuts. Il ne s'agirait non pas d'un modèle « intelligent » de l'Homme à proprement parler mais plutôt d'une simulation, au sens propre du terme, indéfiniment retouchée à coups de statistiques. Cet Homme « résolu » marquerait donc la fin de l'Homme « humain » pour en faire un objet « révolu ».

Est-ce la fin du chemin ?

Remplacer l'Homme par une machine est une idée ancienne. Elle se fonde sur une identification possible du fonctionnement de l'Homme à celui d'une machine. Descartes, déjà, qui comparait les animaux à des horloges, rêvait d'une approche mécaniste de l'homme et donc d'une mécanisation éventuelle de la pensée⁵. Mais cette question renvoie aussi à la théorie de la connaissance.

Au-delà de la philosophie, les sciences humaines se sont largement emparées de cette approche « mécaniste » au XX^e siècle, par exemple avec l'approche cognitiviste fondée

4 Rui DA SILVA, *Histoire de la psychologie cognitive*, Ed. Sciences Humaines « Petite Bibliothèque », Auxerre, 2012.

5 René DESCARTES, *Discours de la méthode*, (1637), Garnier Flammarion, Paris, 2016.

sur l'hypothèse d'une analogie entre la pensée et les processus de traitement de l'information. Ces approches cognitivistes s'appuient sur la théorie de l'information et tendent à décrire l'esprit comme un dispositif. Mais, bien que très pragmatiques, elles échouent à expliquer la capacité humaine à résoudre les problèmes. Il faudrait en effet disposer d'un modèle de connaissance et d'un modèle d'utilisation de ces connaissances pour choisir les actions. Or, si les machines agissent à partir de modélisations, les humains ne sont pas programmables, au moins tant qu'ils conservent leur libre arbitre. La linguistique a aussi largement soutenu cette approche cognitive notamment via les travaux de N. Chomsky qui a pu isoler des structures syntaxiques universelles⁶.

Mais l'Homme demeure « une boîte noire ». Depuis les années 1970, le connexionnisme⁷ modélise les processus mentaux de ladite boîte comme des phénomènes issus d'unités simples et interconnectées. Plus récemment, grâce aux IRM cérébrales, la neurophysiologie a mis en évidence l'activation de réseaux neuronaux qui permettent une « lecture » des pensées⁸. L'observation et l'expérimentation renforcent donc l'idée d'une réalité humaine réduite à une neurophysiologie à base de calculs, à l'instar des ordinateurs.

C'est encore sommaire mais les sciences convergeraient donc vers des caractères humains calculables et artificiellement reproductibles. Mais attention, un problème méthodologique, pour ne pas dire logique, demeure. Le raisonnement se fonde en effet sur l'idée qu'à découper un problème complexe en problèmes élémentaires, le tour est joué. La biologie moléculaire, depuis des décennies, s'est engouffrée dans cette doxa. Et la science, de manière générale, tend à considérer que de la compréhension du

6 Noam CHOMSKY, *Structures syntaxiques*, Seuil, Paris, 1979.

7 <https://fr.wikipedia.org/wiki/Connexionnisme>

8 https://fr.wikipedia.org/wiki/Imagerie_cérébrale

simple, naît, comme par génération spontanée, celle du complexe. Et pourtant, on comprend parfaitement le comportement de certains systèmes complexes, pourvu que l'on ait identifié les bonnes quantités qui les caractérisent, sans pour autant avoir besoin de connaître le comportement de chacune des parties les constituant.

La thermodynamique, par exemple, qui s'intéresse aux échanges de chaleur, s'est développée au XIX^e siècle en accompagnement de la révolution industrielle autour de la vapeur. Ceci a conduit à des prouesses technologiques sans pour autant que l'on connaisse, à l'époque, les propriétés des molécules constituant la vapeur. Si la technologie a alors parfois devancé la science, ces technologies se sont avérées justifiées sur le plan scientifique. Dans ce cas la science a donné *a posteriori* les fondements à une approche empirique. La science a ainsi repris le contrôle de la technologie. C'est essentiel du point de vue méthodologique mais aussi du point de vue des applications, même si la thermodynamique originale, pragmatique et globale, demeure la bonne méthode opérationnelle. Mais en sera-t-il de même de l'invasion des IA visant à dissoudre l'homme dans la mathématique ?

DES LIMITES AUX IA ?

Le point de vue humaniste... et sa faiblesse

L'Homme soluble dans la mathématique ? Non, du moins pas encore, car trop de questions restent inaccessibles au calcul. C'est le cas par exemple du mystère de la conscience qui émerge de notre cerveau. Ou bien, en rebondissant sur l'idée de Platon, celui de l'existence possible d'un monde invisible, à l'extérieur des réseaux sémantiques fermés de la pensée logique et du langage. Pourquoi un monde plus grand n'inclurait-il pas en effet l'espace logique ?

La philosophie interroge l'ontologie autant que l'épistémologie. Cependant, en science comme en

philosophie, la pensée logique et langagière est fermée sur elle-même, bloquée dans ses réseaux sémantiques. Mais on peut imaginer un monde hors des réseaux sémantiques du langage et auquel notre pensée ne peut avoir accès. Et donc l'Homme peut aspirer à exister hors des déterminations mathématico-physiques et philosophiques fondées sur des langages. Peut-être l'esprit humain déborde-t-il du logos en intuitions diverses, en appréhensions confuses, en désordres illogiques, en paradoxes, en indicibles divers.

Cette posture se ramène cependant plus ou moins à la position humaniste classique : « l'humain est trop complexe pour être calculé »... Il est évident pour les tenants de ce point de vue que tout calculer est impossible. Il suffit de regarder autour de nous, si ce n'est nous-mêmes ! Mais l'argument est faible. Se réfugier derrière la complexité peut paraître évident mais reste logiquement insuffisant. En effet qu'est-ce qui prouve qu'à terme on ne saura pas traiter de systèmes de plus en plus complexes et prédire leur évolution ? D'ailleurs il suffit de regarder en arrière pour voir que nombre de prédictions de cet ordre sont tombées à l'eau, du voyage sur la lune à l'ingénierie génétique... Alors, pourquoi ne saurait-on pas trouver les bonnes caractéristiques des systèmes complexes pour les réduire à des systèmes simples, complètement calculables ?

Prenons un exemple de la vie courante, l'air constituant une pièce. La vision moderne en fait un assemblage de molécules d'azote et d'oxygène en agitation constante. On sait décrire le mouvement de chacune de ces molécules. Mais ce calcul gigantesque est inutile. Il « suffit » en effet de connaître la température et la pression, que l'on sait très bien gérer avec un radiateur ou une fenêtre, et que l'on sait relier aux caractéristiques desdites molécules. Dans ce cas on passe d'un système ultra complexe à un système relativement simple et « calculable », de la complexité du monde microscopique à la simplicité de la thermodynamique. La réponse humaniste est donc un pis-aller, une sorte de retraite stratégique, mal fondée

rationnellement dans la mesure où le doute de la simplification possible subsiste. Alors, existerait-il des situations moins discutables ?

L'homme face à l'ambiguïté

Figure 1 : Le canard-lapin, plus qu'une illusion d'optique.



En fonction du point de vue, au premier comme au second degré, on voit sur cette image un profil de canard (regardant vers la gauche) et/ou un profil de lapin (regardant vers la droite). On peut se demander ce que verrait une IA devant ce dessin. Saurait-elle décider ? Douter ? Accepter un « peut-être » ou un « aussi » ? Tiré de Jastrow, *Op. cité*.

Pour illustrer notre propos nous allons prendre l'exemple du célèbre canard-lapin de Jastrow, proposé à la fin du XIX^e siècle⁹, repris et analysé quelques années plus tard par L. Wittgenstein¹⁰. Le canard-lapin est cette entité imaginaire qui regarde vers la gauche comme canard et vers la droite comme lapin (Figure 1). Qu'en est-il de l'homme face au canard-lapin ? Laissons-nous porter ici par L. Wittgenstein qui a donné du canard-lapin une analyse fort éclairante pour expliquer divers niveaux d'appréhension.

9 Joseph JASTROW, *Fact and fable in psychology*, Houghton, Mifflin and Co, Boston, 1900, accessible via <https://www.gutenberg.org/ebooks/48869>

10 Ludwig WITTGENSTEIN, *Remarques sur la philosophie de la psychologie*, vol 1, Éditions T.E.R., 1989.

Le premier niveau est informatif. C'est celui des sciences, des pensées, du langage. Lié aux perceptions, ce premier niveau pointe cet aspect important de la connaissance qui ne peut finalement être que re-connaissance. Ce premier niveau implique donc la mémoire et la reconnaissance. On connaît les lapins et les canards. On les retrouve dans l'ambiguïté du dessin. Il s'agit bien de re-connaissance car un individu qui n'aurait jamais vu de canard de sa vie, ne verrait que le lapin. Il en va de même des sciences qui sont donc de nature variable, puisqu'au fur et à mesure des progrès, ce qui est reconnaissable varie.

Le second niveau est plus subtil. Il nous fait juger de la qualité du dessin, crayon ou plume, imaginer l'auteur, le support, le contexte, la finalité. C'est un niveau plus stable de connaissance élargie jusqu'aux frontières du contenant. Le contenu est le dessin, le contenant tout le reste. Cette connaissance est métaphysique c'est-à-dire « à côté » de ce qui est physique. Et là, nous entrons dans le champ philosophique.

Au-delà s'esquisse un troisième niveau, peut-être celui de Spinoza dans sa connaissance du troisième genre¹¹. Ce n'est plus le domaine de la pensée, c'est celui de l'intuition, mélange d'imagination et de sensation. C'est aussi le domaine du lien au tout incommensurable dont le dessin s'est extrait. C'est-à-dire tout ce qui n'est pas particulier comme la feuille et le dessin, tout ce qui est imprécis, potentiel, virtuel. L'humain s'y sent engagé dans la démesure d'une globalité qui le dépasse.

L'IA face à l'ambiguïté

Que ferait une IA devant un canard-lapin ? Si l'IA ne connaît que les canards, à l'exclusion des lapins, l'IA détectera un canard (et réciproquement, si l'IA ne connaît que les lapins). Pas de surprise, l'IA a « reconnu » ce qu'elle

11 Baruch SPINOZA, *L'Éthique*, (1677), Trad. de R. Misrahi, Livre de poche, Paris, 2011.

connaissait. Ce cas-là n'est donc pas très intéressant. Imaginons maintenant que l'IA ait été entraînée à reconnaître séparément les lapins et les canards. Les deux animaux font partie du registre de ses possibilités, comme pour nous ! Alors que verra l'IA ? Un canard ou un lapin, en fonction du critère de décision qui lui est imposé ! Et si l'IA a plus de canards dans sa base de données que de lapins... elle verra vraisemblablement un canard ! Même s'il n'y a qu'un canard de plus dans sa base de données ! Ainsi vont les probabilités. Le canard-lapin est donc un cas difficile pour une IA et montre ses limites intrinsèques, notamment par rapport à l'Homme, qui voit canard et lapin et sait jouer de cette ambiguïté.

Et pourtant, en une décennie les IA ont acquis une maîtrise inégalée en reconnaissance d'images, par exemple pour le diagnostic médical, même si des applications potentiellement liberticides, en matière de sécurité, existent aussi. L'IA régnerait donc en maître incontesté dans le domaine. Mais que peut faire une IA face à une illusion d'optique comme le canard-lapin ? Par construction, elle ne peut qu'évaluer les probabilités de voir tel ou tel objet en fonction de la base de données sur laquelle elle s'appuie. C'est donc au concepteur de cette IA qu'incombe la responsabilité de lui donner telle ou telle « inclination ».

Tout cela se réduirait à un jeu sans importance si ce canard-lapin n'était qu'une illusion d'optique. Mais le canard-lapin porte, comme nous l'avons vu, beaucoup plus de sens que cela. L'IA s'efface alors devant l'homme, car elle ne peut pas traiter l'information ambiguë et encore moins s'en amuser ! L'une des spécificités de l'homme n'est-elle pas sa capacité à assumer les ambivalences, les exceptions et les dérèglements ?

Comme le soulignait Aristote l'homme « qui sait », le « sachant », a l'intuition de son ignorance, de même que le concepteur, celui qui conçoit, a l'intuition de l'inconcevable. Mais est-il concevable que la machine ait l'intuition de

l'inconcevable ? Est-il besoin de se demander si la machine peut atteindre le troisième niveau d'analyse de Wittgenstein ? À la différence de l'homme qui accède à la connaissance du troisième genre, une IA peut-elle « éprouver et sentir » une idée, plus que la « penser », faire preuve d'imagination ? Peut-elle exister en dehors des langages et des algorithmes ? Mais alors, la révolution numérique ne serait-elle rien d'autre qu'une révolution technologique de plus ?

L'HOMME FACE AUX RÉVOLUTIONS SCIENTIFIQUES ET TECHNIQUES

Trois révolutions industrielles en trois siècles

On pense d'abord aux révolutions industrielles récentes : celle de la vapeur, du train et du télégraphe au XIX^e siècle puis celle du pétrole, de l'électricité, de l'automobile et du téléphone au XX^e siècle. Selon Rifkin¹², chaque révolution industrielle s'appuie sur une association systémique entre une nouvelle source d'énergie et une nouvelle circulation d'informations. La dernière de ces révolutions serait celle que nous vivons actuellement, avec la montée en puissance des énergies renouvelables et le numérique.

À court terme, trois siècles tout au plus, et compte tenu du fait que nous vivons un bouleversement profond, nous avons le sentiment naturel que la révolution numérique surpasse largement en intensité et en impact les deux précédentes révolutions industrielles. On voit une forte continuité dans l'évolution technique des derniers siècles, y compris avant les révolutions industrielles. Les machines semblent avoir pris petit à petit une place croissante au sein de l'humanité, mais « en douceur ». L'impression de rupture qui caractérise la révolution numérique paraît plus forte sans doute à cause

12 RIFKIN, *La troisième révolution industrielle*, Les liens qui libèrent, Paris, 2011

du sentiment de forte accélération (et donc de perte de contrôle) du processus. Mais ne serions-nous pas victimes d'un problème de perspective ?

Les révolutions techniques sont profondes et modifient l'Homme

Depuis son apparition sur terre, il y a environ deux millions et demi d'années, on estime que l'humanité a vécu une dizaine de révolutions technico-scientifiques majeures. Par exemple, l'apparition du biface comme outil, il y a peu ou prou 800 000 ans, ou la maîtrise du feu, il y a environ 450 000 ans, ne représentent-elles pas des révolutions au moins aussi majeures, si ce n'est plus essentielles, que l'apparition des IA ? Et ne soulèvent-elles pas des questions aussi sensibles ? Il ne s'agit pas ici de comparer l'incomparable, juste de prendre un peu de recul.

Mettons-nous par exemple à la place de l'un de nos ancêtres qui découvre un outil. Cet outil ouvre de nouvelles perspectives, par exemple un arc et des flèches pour chasser une bête sauvage. N'est-ce pas là, pour notre ancêtre, une rupture, un bouleversement majeur qui lui fait, temporairement, perdre ses repères ? Et peut-être se dire qu'il n'est finalement plus si utile que cela tel qu'il est. Si une flèche peut tuer son gibier, y a-t-il encore besoin d'un chasseur intrépide qui affronte la bête face à face, à mains nues et au péril de sa vie ? Ou bien suffit-il de trouver une bonne position à l'abri, caché dans les broussailles, avec un arc que même un faible enfant peut manier ? Comment imaginer que notre chasseur ne se soit pas légitimement demandé quelle était sa place ?

Le point essentiel, au-delà de l'anecdote, c'est la position et l'évolution de l'Homme par rapport aux révolutions techniques. On a longtemps considéré qu'à cause de son gros cerveau l'homme était à l'origine des multiples évolutions techniques et scientifiques qui jalonnent l'histoire de

l'humanité. Il est évident qu'il y a ici un lien causal. Mais ce lien n'est pas aussi simple qu'il y paraît. Si l'Homme peut être à l'origine de la technique, celle-ci est aussi à l'origine de l'Homme tel que nous le connaissons aujourd'hui. Cette thèse a été largement développée au milieu du XX^e siècle par l'anthropologue A. Leroi-Gourhan, qui a démontré que la technique modifie le développement des capacités neurologiques¹³. En un mot, l'Homme serait non seulement le père mais aussi l'enfant de la technique qu'il développe. Mais cet environnement technique n'est pas forcément de nature « rationnelle ». Il peut résulter de faits hasardeux ou de conditions particulières sans rapport avec une détermination humaine.

Qu'est-ce à dire ? L'homme n'est pas plus intelligent parce que son encéphale est naturellement volumineux... Non. C'est le volume de l'encéphale qui a été conditionné dans son développement par les sollicitations adaptatives du milieu. Ce phénomène naturel décrit l'environnement technique comme une « matrice ontologique »¹⁴, qui enfante constamment un nouvel être... De cette matrice, naît une humanité neurologiquement modifiée par l'influence des techniques qui conditionnent la manière d'apparaître des choses. Les perceptions découlent du milieu où elles se développent. Ainsi la lune qui porte l'empreinte de pas humains, et que l'on voit aujourd'hui, n'est pas la « même », pour nous, que celle qui était regardée par les hommes du paléolithique. La perception est modifiée. De même, des études neurologiques montrent que les jeunes générations, fortement imprégnées par la révolution numérique,

13 André LEROI-GOURHAN, *Le geste et la parole*, Tomes 1 et 2, Albin Michel, Paris, 1965.

14 Stéphane VIAL, *L'être et l'écran, Comment le numérique change la perception*, PUF, Paris, 2013.

développent des stratégies cognitives différentes de celles de leurs aînés¹⁵.

Vers une renaissance de l'Homme ?

Considérer ces bouleversements dans l'histoire de l'Homme participe fondamentalement de l'idée que nous sommes les produits de mutations techniques qui ont des incidences sur la constitution humaine. L'Homme a donc constamment coévolué avec la technique. Il a ainsi subi de perpétuelles évolutions que l'on pourrait presque qualifier de révolutions. Et en jouant sur les mots on pourrait presque dire que l'Homme est constamment révolu par la technique, au sens d'être en perpétuelle (r)évolution. Mais surtout, à tout instant, l'Homme se trouve ainsi révolu au sens premier du terme, pour renaître comme nouvel Homme. Le terme de révolu ne porte donc plus le sens négatif d'achèvement mais, paradoxalement, celui d'éternel recommencement. L'idée que nous sommes ainsi naturellement « révolus » inscrit donc l'espèce dans une dynamique naturelle de modifications successives, sans rapport direct avec la « culture ».

Dans ce contexte, même si l'Homme semble temporairement et partiellement « résolu » – c'est-à-dire celui dont la pensée peut être mécaniquement traduite par des calculs – il ne s'agit peut-être que d'un simple épisode anthropologique qui nous échappe en partie. Situer la révolution numérique dans le cours plus globalisé de notre histoire permet donc de relativiser nos responsabilités en la matière et de poser les termes de notre évolution comme étant, certes inéluctable, mais naturelle et source d'espoir, donc de futur. Cette évolution est porteuse de l'harmonisation entre ce qui varie – les techniques, les savoirs qu'elles drainent – et ce qui est invariant,

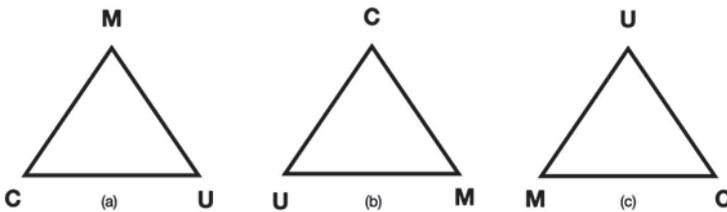
15 Guy VALLANCIEN, « Ecce Homo... artificialis », *Le Débat*, 2017/1 n° 193, pages 167 à 181.

l'adaptabilité aperceptive et essentielle de l'être que décrit Spinoza dans la connaissance du troisième genre. Quel est l'avantage de tels constats ? Celui d'envisager la révolution numérique comme une opportunité de renaissance pour l'Homme.

QUELLE EST LA PLACE DE L'HOMME ?

Le triangle CUM

Figure 2 : Le Triangle CUM comme illustration de la place de l'Homme.



Le triangle CUM (Concepteur, Utilisateur et Machine) regroupe sous forme schématique les trois acteurs de la technique. En fonction de la position symbolique relative des sommets on peut distinguer l'approche « standard » (a), de type « complotiste » (b) et l'approche « humaniste » (c) qui replace l'Homme en position dominante, mais consciente.

La question de notre relation à la révolution numérique et plus généralement aux révolutions techniques, peut se visualiser sous la forme compacte d'un triangle. Nous avons baptisé ce triangle CUM, en référence au sens latin du mot (« avec »). Ce triangle (figure 2) regroupe trois éléments clés : la Machine (M), son Concepteur (C) et l'Utilisateur (U). En jouant sur la place respective et symbolique des différents sommets on perçoit les relations entre C, U et M de différentes manières. Ainsi l'image de la vision courante de l'Homme révolu par le numérique (section 1), correspond au schéma (a) avec le M au sommet et C et U au pied. Cette disposition met l'accent sur la machine qui écrase

l'utilisateur mais fait également fi du concepteur, pourtant essentiel notamment dans la révolution numérique avec ses IA « survitaminées » aux *Big Data*. Au contraire notre brève analyse du nouvel Homme « révolu » (section 3) nous pousserait à l'imaginer au sommet du triangle suivant le schéma (c). Sans ignorer la machine ni le concepteur, l'Homme, l'utilisateur, y retrouverait sa place dominante. Enfin, on ne peut éviter d'imaginer ce que pourrait représenter le triangle avec un C en position dominante (schéma (b)), cette configuration évoquant clairement les logiques complotistes malheureusement de plus en plus d'actualité via le développement des réseaux sociaux.

Il semble donc que le nouvel Homme « révolu » puisse retrouver, en toute conscience, sa place au sommet du triangle CUM. L'Homme engendré par les progrès de ses révolutions, sans ignorer la machine, ni le concepteur, trouve ainsi la place dominante dans le libre usage de ses outils techniques. Cette posture, qui relève en dignité l'être humain, est celle d'un Humanisme renouvelé. Nous atteignons là le cœur de notre quête. Mais rien n'est donné. L'Homme doit certainement se résoudre à toujours s'interroger, et par là même, à continuer de rouler de révolution en révolution comme une boule de neige dévale la montagne¹⁶. Mais si l'Homme est alors mouvement, il n'est pas pour autant victime.

Comment faire pratiquement ?

Quelle pourrait être la mise en œuvre de cette nouvelle posture de l'Homme ? Au niveau individuel, une clé est la prise de conscience de notre place dans le monde et de notre positionnement par rapport à l'environnement, notamment par rapport aux machines. On pourrait alors envisager, avec un regard plus serein, l'intégration plus ou moins importante

16 Henri BERGSON, *L'évolution créatrice*, (1907), PUF, Paris, 2017.

de la machine dans l'humain. Nous utilisons depuis longtemps des prothèses, amovibles (cane, lunettes...) ou inamovibles (dents, broches dans les os, cœur artificiel...) dans notre corps. Nous nous sommes adaptés à ces prothèses qui améliorent notre confort et même prolongent la vie. L'intégration du numérique ajoute une dimension nouvelle mais s'inscrit dans cette logique.

Le transhumanisme pousse à l'extrême cette intégration de bouts de machines dans l'individu. Sans aller jusque-là, on peut imaginer des sortes de « métissage » avec l'environnement numérique, une coévolution harmonieuse de l'Homme avec les IA, dans un mode d'hybridation douce et contrôlée. C'est par exemple la thèse proposée par S. Vial¹⁷, mais évoquée aussi dans le pamphlet sur *La Tyrannie des algorithmes*¹⁸. Le cas de l'art est intéressant ici avec un débat sur la place possible des IA dans le domaine de la création¹⁹.

Au-delà du niveau individuel vient évidemment le niveau collectif, essentiel pour tout individu. Nos sociétés sont d'ailleurs déjà marquées par quelques signes positifs en ce sens, comme par exemple les initiatives d'économie participative ou les manifestations d'actions collectives et anonymes que l'on pourrait appeler « oasis de fraternité », selon le terme d'E. Morin²⁰. Les jardins collectifs, tris recyclés, trocs locaux, habitats participatifs, boîtes à livres, plateformes de liens entre voisins, territoires isolés connectés, bénévolats ... relèvent de la même logique qui replace l'Homme dans un nouvel environnement. On peut aussi mentionner ici la philosophie *care*²¹ qui propose une

17 Stéphane VIAL, *Op. cit.*

18 Michel BENASAYAG, *La tyrannie des algorithmes*, Éditions Textuel, Paris, 2019.

19 *Intelligence artificielle, Enquête sur ces technologies qui changent nos vies*, Dossier Libération, Champs Actuels, Paris, 2018.

20 Edgar MORIN, *La fraternité pourquoi ?* Actes Sud, Arles, 2019.

21 https://fr.wikipedia.org/wiki/Éthique_de_la_sollicitude

nouvelle disposition éthique de l'individu, enracinée dans la sensibilité. On peut voir dans ces comportements adoucis la fin de l'individualisme et une lecture optimiste du monde contemporain²².

Serait-ce aussi simple ?

Il faut cependant rester circonspect et ne pas se contenter d'un optimisme béat, ou prendre pour argent comptant tous les arguments. Les plus simples, qui finalement se fondent sur des pratiques (section 4-b), pourraient être qualifiés d'arguments de ferveur. L'échec de certaines approches cognitivistes relève aussi un peu de cette logique (section 1-c). Rien ne prouve de manière absolue qu'elles n'aboutiront jamais. Face à une situation somme toute peu dramatique et séduits par l'essor d'initiatives coopératives, nous pourrions néanmoins nous sentir rassurés.

Mais on peut aussi s'appuyer sur des arguments plus rationnels, fondés par exemple sur l'analyse des révolutions techniques (section 3). Cependant, ces arguments relèvent finalement plus de l'autorité que de la raison. Le fait que l'Homme se soit adapté aux multiples révolutions techniques ne dit rien sur sa capacité d'adaptation et de survie face à la révolution numérique. Il n'y a rien de déterministe là-dedans. Même si aucun argument fort ne permet de conclure que la révolution numérique diffère, dans sa nature, des révolutions techniques précédentes. L'analyse de Wittgenstein du canard-lapin relève aussi, en partie, de cette logique d'autorité. C'est une analyse séduisante, certes argumentée et rationnelle, mais limitée par construction puisque fondée sur une part d'arbitraire. Rien ne prouve en effet de manière

22 Michel MAFFESOLI, *Être postmoderne*, Collection Idées, Éditions du Cerf, Paris, 2017.

Luc FERRY, Nicolas BOUBOU, *Contre les idéologies du déclin*, Frémeaux & Associés, La librairie Sonore, 2019.

absolue qu'une machine soit incapable d'aller au-delà du premier niveau informatif.

UNE ANALYSE CRITIQUE : ET SI LA RÉPONSE VENAIT DU CHAOS... OU DU HASARD ?

Le chaos ou la fin des IA triomphantes ?

Figure 3 : Surface avec courbure négative.



Exemple de surface sur laquelle Hadamard a montré qu'une géodésique reliant « au plus court » deux points de la surface conduit au chaos. Le mouvement le long d'une telle ligne n'est donc complètement prévisible que moyennant une connaissance exacte des conditions initiales (position et vitesse au départ), ce qui est strictement impossible tant au niveau des mesures que d'une représentation numérique de ces données.

Sur le plan strictement rationnel, les limites de l'exercice sont donc peut-être atteintes. Mais n'allons pas trop vite. Argumenter que les IA n'arriveront pas à résoudre l'Homme à cause de nos initiatives ou de notre histoire, est insuffisant. Mais tout n'est pas strictement calculable, au sens de prévisible, comme nous l'avons implicitement supposé. Un bon exemple est celui de la prévision météorologique, évoquée au début de ce texte. Les modèles de climat sont assez bien connus et on sait en écrire les équations et les résoudre. Mais ces équations sont chaotiques : la moindre

incertitude sur le point de départ rend imprévisible le futur, d'où l'incapacité des prévisions météorologiques au-delà de quelques heures, quelques jours tout au plus. C'est le fameux effet « papillon »²³.

Le problème pourrait néanmoins tenir à la modélisation, toujours perfectible, ou à un problème de capacité de la machine. En effet on peut se demander si les « équations du climat » sont les bonnes et si leur complexité ne reflète pas notre manque de connaissance. En fait la question est sans objet parce que certains systèmes ultrasimples sont chaotiques par nature, indépendamment de tout modèle. Le mathématicien français J. Hadamard a ainsi démontré en 1898 que le mouvement sur certaines surfaces était chaotique... quoi qu'on fasse²⁴. Un exemple est celui d'une surface en forme de double tore (double *donut*) qui n'est pas sans rappeler le symbole mathématique de l'infini d'ailleurs. Ici pas de problème de modèle ou de précision du calcul. Le mouvement suivant une géodésique, c'est à dire la courbe la plus courte qui relie deux points sur une surface (un bout de droite dans le plan, un arc de cercle sur une sphère)... est chaotique !

Prévoir l'évolution exacte d'un système chaotique nécessiterait une précision infinie dans la connaissance de l'état initial ce qui est techniquement impossible du point de vue expérimental mais aussi numériquement puisque dans un ordinateur la précision est finie. Devant le chaos une IA est donc impuissante... pas plus, mais surtout pas beaucoup

23 « L'effet papillon » est une métaphore désignant l'extrême sensibilité aux conditions initiales en théorie du chaos. Elle reprend le titre d'une conférence donnée par E. Lorenz en 1972 : « Le battement d'ailes d'un papillon au Brésil peut-il provoquer une tornade au Texas ? », voir par exemple https://fr.wikipedia.org/wiki/Effet_papillon

24 Jacques Hadamard a démontré l'instabilité de toutes les géodésiques de certaines surfaces particulières, voir par exemple, https://en.wikipedia.org/wiki/Hadamard%27s_dynamical_system

moins qu'un homme. De là à imaginer que certains de nos comportements relèvent du chaos, avec les conséquences que l'on peut en tirer, il n'y a qu'un pas que nous laissons le lecteur franchir, ou non. Le point important ici est l'absence d'hypothèses arbitraires liées à un modèle.

Le chaos sonne certainement le glas du mythe d'une IA pouvant tout calculer. Mais on pourrait aussi considérer qu'il ne concerne que certaines situations particulières, ce qui est en partie vrai. Rien ne nous dit que l'univers est un double *donut* ! Pire, la nature semble bien éloignée d'un tel objet mathématique. Et quant au comportement chaotique (ou pas) de certains systèmes complexes, la réponse est, elle-même, complexe. La plupart des systèmes complexes présentent en effet des régimes de fonctionnement chaotique qui coexistent avec des comportements périodiques. Et nous avons vu que réduire un système complexe à un plus simple ne résout pas forcément le problème. C'est en fin de compte la question sensible du rôle de la modélisation et du choix des variables pertinentes qui soulève la plus grande difficulté.

Revenons donc un instant directement à la nature, pour éviter d'emblée cette question de la modélisation. Que nous dit la nature ? Tout est-il vraiment modélisable et calculable ? Des éléments de réponse solides existent. C'est la mécanique quantique et certains de ses paradoxes qui éclairent cette réponse... via le hasard.

Et le hasard ?

La version communément acceptée de la mécanique quantique, celle de l'École de Copenhague autour de N. Bohr, repose sur une interprétation de la nature en termes de probabilités, ce qui implique une perte partielle de la notion habituelle de causalité et une place importante faite au hasard. Cette interprétation du monde peut heurter et a donné lieu à des débats multiples depuis près d'un siècle. Un des exemples célèbres est le paradoxe EPR proposé dans les

années 1930 par Einstein, Podolsky et Rosen²⁵. Ce paradoxe réfute l'idée que l'état de deux particules microscopiques initialement corrélées (par exemple mises en rotation sur elles-mêmes en sens inverse l'une par rapport à l'autre) ne peut être connu avant qu'on le mesure. Dans le cas de notre exemple il s'agirait de savoir quelle particule tourne dans quel sens. Précisons : il est évident que la mesure donne la réponse. Mais, pour faire simple, EPR considère que cet état existe avant la mesure et n'est que révélé par la mesure. La mécanique quantique suppose que cet état n'existe pas tant qu'il n'a pas été mesuré. Pour éviter tout biais la mécanique quantique considère donc que l'état du système est un mélange de toutes les possibilités (avec les probabilités idoines) et que la mesure lui fait « choisir » un des états possibles. L'expérience de pensée EPR pointe donc la question de l'existence de la réalité et de la causalité. Le débat est resté ouvert pendant plus d'un demi-siècle, jusqu'à ce que des expériences montrent enfin la voie. Après une quarantaine d'années supplémentaires de vérifications et de raffinements de ces mesures le résultat est sans appel, au détriment d'EPR et au bénéfice de la mécanique quantique. Le hasard l'aurait-il donc emporté ?

Revenons un instant sur ce terme de hasard qui peut prêter à confusion. Il faut distinguer d'abord entre hasard « subjectif » et hasard « objectif ». Le hasard subjectif, c'est celui qu'on connaît bien. Par exemple, lancer une pièce en l'air et noter pile ou face se modélise parfaitement par les probabilités, une chance sur deux pour chaque face d'une pièce parfaite. Dans ce cas la probabilité ne fait que simplifier un problème que l'on pourrait résoudre par ailleurs si on le voulait, puisque l'on sait parfaitement décrire le mouvement de la pièce. Il en va de même de situations où la complexité, par exemple des interactions, est avantageusement modélisée

25 Franck LALOË, *Comprenons-nous vraiment la mécanique quantique ?*, 3^e édition, CNRS Éditions, Paris, 2018.

par des variables aléatoires. Ce hasard subjectif est en fait un « faux » hasard, un hasard d'ignorance ou de simplification, mais qui ne remet pas en cause notre possibilité de compréhension du mécanisme sous-jacent ni surtout son existence. Ce hasard subjectif est donc parfaitement compatible avec la notion courante de causalité.

Dans le cas d'un hasard objectif les fondements sont moins stables. On a remplacé une compréhension par un chiffre, la probabilité, qui est le pur résultat d'une observation sans compréhension et qui n'implique pas forcément la possibilité même de cette compréhension... au-delà de la probabilité. Un exemple typique de hasard objectif potentiel est celui de la radioactivité qui conduit à la désintégration de certains noyaux. On sait parfaitement décrire la radioactivité par des probabilités. On caractérise d'ailleurs un processus de radioactivité par sa demi-vie, c'est à dire l'instant auquel la moitié des noyaux d'un échantillon s'est désintégrée. Cette demi-vie est caractéristique du noyau composant l'échantillon et le traitement probabiliste parfaitement adapté. Par contre on ne sait pas quand un noyau donné va se désintégrer ! Le processus individuel de désintégration est imprévisible. En l'état de nos connaissances on est sans doute face à un effet de hasard objectif.

La notion de probabilité recouvre les deux aspects, hasard subjectif et hasard objectif sans les distinguer... On pourrait considérer cela comme dangereux car faussement rassurant. Mais on peut aussi considérer que la probabilité permet de restaurer de manière effective une dimension qui nous échappe et que nous avons tendance à occulter, celle du hasard objectif. D'une certaine manière, en recourant à une probabilité, on perd donc une partie de la connaissance mais on en gagne une autre. L'incertitude liée à la probabilité permet un gain par rapport à la certitude ! Le hasard dont il est question en mécanique quantique pourrait être objectif. Le conditionnel tient en partie à la question de savoir si la

théorie quantique est complète. C'était précisément l'une des questions soulevées par EPR avec comme conséquence l'existence de possibles variables « cachées ». Les expériences ont tranché et aucune ne remet aujourd'hui en cause la validité de la théorie quantique. Le hasard quantique pourrait donc être objectif...

Qu'en conclure ?

Si effectivement la nature recèle une part de hasard objectif, ce qui en l'état de nos connaissances semble avéré, l'idée du tout calculable, sur la base d'une représentation mécanique de la nature n'est donc pas fondée. Le hasard objectif écorne en effet la causalité au sens où on l'entend habituellement. Cette dernière fonde les sciences modernes et demeure sous-jacente au tout calculable de l'IA. On ne discutera pas ici plus avant de cette causalité écornée.

On pourrait aussi se dire que le hasard, même objectif, pourrait très bien être simulé par une IA. D'ailleurs combien de jeux numériques utilisent le hasard ! Mais attention le hasard généré par un ordinateur n'a de hasard que le nom... Les ordinateurs génèrent des séquences de nombres (dits pseudo-aléatoires) qui présentent, au moins sur les échantillons considérés, toutes les caractéristiques de vrais nombres aléatoires. Mais ce sont des ersatz, parce que ces séquences de nombres sont construites de façon parfaitement déterministe et reproductible... et surtout périodique. Au bout d'un nombre suffisant de « tirages numériques » on sait mathématiquement que l'on va retrouver le nombre de départ et régénérer la même séquence de nombres. Entendons-nous, dans toutes les applications courantes on n'atteint jamais ce « retour à la case départ » parce que la période associée est gigantesque. Par contre, d'un point de vue logique, ces nombres pseudo-aléatoires ne constituent pas un vrai hasard et encore moins un hasard objectif. Comme avec le chaos l'IA rencontre donc, hors de tout doute lié à une quelconque

modélisation, un mur infranchissable. Tout n'est pas calculable par une IA. Les mathématiques s'allient à la nature pour délivrer ce constat indiscutable.

ÉLÉMENTS DE CONCLUSION

Conscients des craintes suscitées par la révolution numérique, nous croyons néanmoins fondé de considérer que l'Homme n'est ni révolu ni obsolète, bien au contraire. C'est peut-être même pour lui une opportunité inattendue d'atteindre une nouvelle place dans le monde qu'il devrait être capable de préserver s'il fait preuve d'un minimum de conscience et de sens critique.

Le fait que l'Homme se soit adapté aux multiples révolutions techniques ne dit rien sur sa capacité d'adaptation, voire de survie, dans un monde organisé par les chiffres et écologiquement très menacé. Cependant, peu d'arguments autorisent une conclusion radicalement défaitiste. La révolution numérique ne semble pas différer, dans sa nature, des révolutions techniques précédentes. Repenser l'humanisme à travers la connaissance du troisième genre pourrait aussi être rassurant. Elle permettrait de quitter la posture passive, typique du premier niveau de la connaissance fondé sur la simple opinion, pour replacer l'Homme dans une disposition active, mêlant rationalité et intuition.

En tous les cas, il reste un fait : tout n'est pas strictement calculable, au sens de prévisible, d'autant plus que la nature recèle une part de hasard inaccessible au calcul. Le chaos et le hasard objectif semblent donc bien laisser la porte ouverte à la nouvelle place de l'homme face aux IA. Mieux, ils fournissent des éléments objectifs et indiscutables de l'incapacité des IA face à eux. Or il est fort probable que le comportement de l'homme, système complexe par nature, relève en partie du chaos. Et la nature dans laquelle évolue l'homme semble elle-même propice à ce hasard inaccessible

aux IA. La place de l'homme au sommet du triangle CUM est donc bien réelle même si parfois le sommet de ce triangle peut paraître bien pointu.

Continuons donc d'agir au mieux, portés par l'idée irrépressible de nous transcender dans le meilleur, comme au sein d'oasis de fraternité. Nous sommes bien face à une énième renaissance de l'Homme, révolu une fois encore pour mieux renaître et évoluer. L'Homme révolu, passif, n'est pas à l'ordre du jour. Au contraire, saisir nos limites et demeurer éblouis par l'incommensurable est la posture humaniste qui nous préserve en dignité, en éveil et en perpétuelle renaissance.

INTELLIGENCE ARTIFICIELLE, TRANSHUMANISME ET JUSTICE PÉNALE

par M. Jean TOUZET

*Procureur général honoraire près la Cour d'appel de Reims
Ancien Président de l'Académie nationale de Reims*

Il nous a été demandé, en cette année académique, de réfléchir sur les problèmes nouveaux que posent ces concepts encore inimaginables il y a peu de temps et dont l'ampleur et les conséquences sur notre monde et sur notre mode de vie sont encore impossibles à mesurer et même simplement à évaluer : il s'agit de l'intelligence artificielle d'une part, du transhumanisme d'autre part. Je n'ai ni connaissance théorique ni expérience pratique de ces domaines mais leur existence même m'interpelle sous l'angle de la profession qui fut la mienne pendant quarante-deux années (4 mai 1953 – 30 juin 1995), celle de magistrat spécialisé dans la justice pénale, douze années ayant été consacrées aux fonctions de juge d'instruction et vingt-sept à celles de l'accusation publique ou, comme on dit encore : du ministère public ou du parquet.

Que l'ensemble constitué par la législation pénale – extrêmement abondante, trop abondante sans doute –, par la jurisprudence qui l'interprète et l'applique et par la doctrine qui commente l'une et l'autre, ne produise pas des résultats satisfaisants, chacun le sait ; il suffit de considérer l'ampleur de la délinquance et de la criminalité sous toutes leurs formes et l'ampleur du phénomène de récidive et de multirécidive pour se rendre compte que, malgré le dévouement de tous, magistrats, avocats et fonctionnaires – car on travaille beaucoup dans les palais de justice et dans les services qui en dépendent – et malgré le souci constant et partagé de faire pour le mieux, le résultat n'est pas celui que les magistrats eux-mêmes, les justiciables, les citoyens et la société toute entière attendent légitimement ; ce déficit n'est pas nouveau car

jamais, en France comme ailleurs, ni la délinquance ni la criminalité n'ont été véritablement maîtrisées. On invoque volontiers, dans la justice française, le manque de moyens et, certes, il serait vain de le nier : avec onze juges pour cent mille habitants, la France se trouve dans les trois derniers des quarante-sept pays réunis sous l'égide du Conseil de l'Europe ; mais les pays les mieux dotés peinent autant que nous à maîtriser la violence sous toutes ses formes, l'improbité sous toutes ses formes, la malversation sous toutes ses formes.

Aussi, lorsque j'entends parler de conquêtes nouvelles, déjà en voie de réalisation telle l'intelligence artificielle ou encore à l'état de réflexion tel le transhumanisme, je ne peux m'empêcher de m'interroger : « Cela serait-il, en tout ou en partie, applicable à la justice pénale ? Cela serait-il bénéfique ? Cela résoudrait-il, ne fût-ce que partiellement, des problèmes demeurés sans solution satisfaisante depuis des siècles ? ». Bien entendu, mon âge, au cœur de sa quatre-vingt-treizième année, ne me laisse aucune chance de connaître les réponses, mais cela ne m'empêche pas de me poser les questions et ce sont simplement ces questions que je voudrais évoquer ici.

LES ÉLÉMENTS DU PROBLÈME

La justice pénale exige, d'abord et avant tout, la connaissance de la vérité, quant à la matérialité des faits et quant à l'identité de leur auteur, et ensuite, la détermination de la solution adéquate, c'est-à-dire à la fois conforme à la loi, adaptée à la personne du coupable et bénéfique pour la société. La première question se rapporte au passé puisque la justice pénale n'est saisie que de faits déjà consommés et la seconde question se rapporte à l'avenir puisqu'il s'agit des effets de la solution, généralement en forme de sanction, sur le comportement futur du délinquant ou du criminel ; or connaître le passé est déjà loin d'être simple, quant à connaître l'avenir... Paul Valéry, il est vrai, conciliait les deux en les rejetant dans les mêmes ténèbres avec cet aphorisme :

« L'historien est au passé ce que la tireuse de cartes est à l'avenir mais la sorcière s'expose à une vérification, pas l'historien » ; soit... on pourrait en disserter longuement¹ mais ce qui est incontestable c'est la grande difficulté de connaître le passé et de construire l'avenir et, lorsqu'il s'agit de justice pénale, nul n'ignore les conséquences fâcheuses de toute erreur dans un domaine comme dans l'autre.

La connaissance de la vérité

Sur le premier point : la matérialité des faits et l'identité de l'auteur, toute insuffisance ou toute appréciation erronée entraîne l'erreur judiciaire, locution qu'on tend à réduire au seul cas de la personne déclarée coupable alors qu'elle ne l'est pas et condamnée alors qu'elle ne devait pas l'être ce qui constitue, sans contestation possible, le pire qui puisse arriver. Mais une procédure clôturée sans déclaration de culpabilité ni condamnation à la charge de quiconque – que ce soit par classement sans suite du parquet, par ordonnance de non-lieu du juge d'instruction, par jugement de relaxe du tribunal correctionnel, par arrêt de relaxe de la cour d'appel ou par arrêt d'acquiescement de la cour d'assises ou de la cour criminelle (là où cette dernière juridiction est actuellement en place) –, constitue pareillement une erreur judiciaire beaucoup plus fréquente et qui, infiniment moins dramatique qu'une condamnation infondée, n'a cependant rien d'anodin ni de négligeable puisque l'ignorance de la vérité exclut toute réponse adéquate à l'égard de l'auteur de l'infraction comme de la communauté dont il est membre. Quant à la victime, si elle peut, sous certaines conditions, être financièrement indemnisée, l'argent ne résout pas nécessairement les problèmes psychologiques et moraux que peuvent engendrer certaines infractions.

¹ C'est précisément ce qui avait été demandé, au titre de l'épreuve de culture générale, au concours d'accès aux fonctions de magistrat, en 1951.

La détermination de la sanction

Quant au second point : l'erreur dans la détermination de la sanction appliquée au coupable exactement identifié, comment l'éviter, tant le problème est complexe ? Certes, on fait la part des choses et – qu'on requière la peine en tant que magistrat du ministère public ou qu'on la prononce en tant que juge – on ne cherche pas la réponse dans la même gamme de sanctions pour l'assassinat d'une personne ou pour le vol d'un crayon à bille ; mais la question est celle du champ de solutions dont on dispose qu'on en ait peu voire trop peu, qu'on en ait beaucoup, voire trop. Avec l'ancien *Code pénal* de 1810² dans sa version originale, le champ était fort étroit et pour ne prendre que cet exemple, le vol simple, commis de jour, sans arme ni violence ni effraction ni état de récidive, était puni d'un emprisonnement de un an à cinq ans, emprisonnement ferme puisque le sursis ne fut institué que bien plus tard, en 1891. Quant aux circonstances atténuantes autorisant une peine inférieure au minimum légal, elles ne pouvaient être accordées que si le préjudice était inférieur à vingt-cinq francs, somme qui, à l'époque, n'était sans doute pas négligeable, mais enfin, pour un vol de trente francs, la peine minimale, pour un délinquant primaire, était d'une année d'emprisonnement ferme, étant précisé qu'à cette époque et en l'absence de toute possibilité légale de libération anticipée, la peine subie était nécessairement la peine prononcée. Avec le *Code pénal* de 1994 – maintes fois modifié depuis sa promulgation et toujours par adjonctions – j'avais recensé pour un vol, une trentaine de réponses possibles, soit un champ pratiquement illimité (et encore élargi depuis lors), chose qui eut fait frémir Montesquieu et chose susceptible d'engendrer deux dérives : ou bien cette quasi-indétermination de la peine, le magistrat se l'approprie, cas auquel l'inévitable subjectivité qu'il est

2 Le dernier des cinq codes promulgués par Napoléon I^{er}, après le *Code civil* (1804), le *Code de procédure civile* (1805), le *Code de commerce* (1806) et le *Code d'instruction criminelle* (1808).

impossible d'éliminer totalement peut aboutir, à l'égard de plusieurs justiciables, pour des faits analogues et des situations personnelles comparables, à des sanctions différentes ; ou bien, n'intégrant pas toutes les possibilités, le magistrat rétrécit le champ des possibles et se focalise sur un nombre restreint de solutions, cas auquel il rend certes homogène sa propre jurisprudence mais ne peut-on craindre que ces politiques pénales, conduites par des juges cultivant des tendances opposées, ne conduisent à des réponses pénales dispersées au point de défier le principe de l'égalité des citoyens devant la justice ?

Rien de tout cela ne peut satisfaire et on aspire tellement à des solutions meilleures que lorsqu'on entend parler d'innovations ou de découvertes, on pense aussitôt : « Et si cela réussissait pour la justice pénale ? » ; et c'est pourquoi, personnellement, je me pose ces deux questions : « Que pourrait apporter l'intelligence artificielle ? » et « Que pourrait apporter le transhumanisme ? ».

QUE POURRAIT APPORTER L'INTELLIGENCE ARTIFICIELLE ?

Envisageons successivement les deux domaines que nous venons d'évoquer : celui de la recherche de la vérité et celui de la détermination de la sanction.

Le premier domaine : la recherche de la vérité

De tout temps, la recherche de la vérité, en matière de justice pénale, a été une immense difficulté que les hommes ne savaient comment aborder et encore moins comment résoudre tout en étant conscients de leur incapacité. Aussi s'en remirent-ils d'abord aux dieux ou à Dieu par le moyen des ordales ; tel était-il accusé de vol et le niait-il ? On lui plongeait la main dans l'eau très chaude, on confectionnait un pansement fermé par un scellé et puis, trois jours plus tard, on levait le scellé et

on avait la réponse : si la main était en bon état, la preuve était faite de l'innocence ; si elle suppurait, la preuve était faite de la culpabilité. Une femme était-elle accusée d'adultère et le niait-elle ? Il suffisait de la jeter à l'eau et selon qu'elle coulait ou qu'elle demeurait à la surface, on connaissait la réponse. Excellentes méthodes, certes, mais les meilleures choses ont une fin et, en 1215, le IV^e Concile de Latran mit un terme à ces modes de conviction, il fallait trouver autre chose et, en cette Terre de chrétienté, quoi de plus pertinent que la question c'est-à-dire la torture ? Car Dieu est juste ou plutôt Dieu est Justice, et il soutient l'innocent dans l'épreuve tandis qu'il abandonne le coupable à ses propres capacités de résistance dès que l'intensité de la torture en a atteint les limites. C'était bien pensé mais les Lumières et la Révolution française ont – fort heureusement d'ailleurs – aboli ces abominables méthodes. Alors que faire ? Il n'y eut plus désormais – et il n'y a toujours – que deux gisements de vérité : d'une part, les paroles : celles du suspect – s'il en est un –, celles de la victime et celles des témoins – s'il en existe et si on en trouve – et, d'autre part, les choses : l'état des lieux, les traces de pas, les objets découverts sur place ou en un domicile et les empreintes diverses (digitales ou génétiques) qu'ils peuvent porter. Recueillir ces éléments : paroles et choses, les rassembler et les interpréter, tel est le travail des enquêteurs qu'ils appartiennent à la Police nationale ou à la Gendarmerie nationale, tel était mon travail comme juge d'instruction et tels étaient les éléments dont je disposais lorsque j'exerçais l'accusation publique. Mais l'expérience a montré depuis longtemps les limites et les imperfections de la méthode. Certes, il existe – heureusement – des cas simples, beaucoup de cas simples, où le rapprochement des données débouche tout de suite ou très rapidement sur l'évidence, sur la preuve ; mais il n'en va pas toujours ainsi et souvent manquent des éléments qui conduiraient à cette certitude, cas auquel, à défaut de preuve topique, on doit se contenter d'un « ensemble de présomptions graves, précises et concordantes, équipollentes à

la preuve », formule qui ne va pas sans difficulté car telle personne est auteur de l'infraction ou ne l'est pas, ce qui situe le raisonnement dans la logique binaire du oui ou du non tandis que, avec des notions dépourvues de limite incontestable comme celle de gravité, voire celle de précision, on se trouve dans une logique floue, une logique du plus ou du moins et passer de celle-ci à la logique du oui ou du non n'est pas simple ; c'est alors dans ce processus que peut se glisser l'erreur, celle qui fait reconnaître comme auteur de l'infraction celui qui ne l'est pas, ou ne permet pas de reconnaître celui qui l'est. À ce sujet, soulignons que le principe fondamental exprimé par l'adage : « *In dubio prosit reo* » (« Dans le doute que (cela) profite à l'accusé » ou simplement : « Le doute profite à l'accusé ») conduit infiniment plus souvent à l'erreur négative qui laisse l'auteur impuni qu'à l'erreur positive qui punit l'innocent ; et si l'erreur n'est jamais une bonne chose, elle est toujours un échec, mieux vaut, certes, que l'erreur négative domine et, en fait, elle domine beaucoup.

En ce qui concerne la collecte des éléments dans le domaine de la parole, je ne suis pas sûr qu'on ait fait beaucoup de progrès et je me demande si on interroge mieux un suspect, si on entend mieux une victime ou un témoin qu'on ne le faisait jadis, sauf – et c'est important – lorsqu'il s'agit d'enfants pour lesquels la meilleure connaissance de la psychologie infantile chez des enquêteurs spécialement formés a certainement apporté une amélioration. En revanche, on a fait d'énormes progrès dans le domaine des choses, qu'il s'agisse d'en percevoir l'existence, de les relever sans les altérer et de les interpréter : empreintes digitales, empreintes génétiques, analyses et opérations techniques de toute sorte, les progrès sont extraordinaires³. Quant à exploiter rationnellement les éléments acquis, quelle que soit leur source, le problème est de

3 On n'imagine plus une seconde de nos jours la bévue de ce sympathique gendarme qui, constatant une empreinte de pas bien nette, entreprend de la protéger en attendant l'arrivée du photographe en la couvrant de brindilles qui la brouillent complètement et la rendent inexploitable.

les rapprocher judicieusement pour construire un ensemble non seulement cohérent mais encore et surtout véridique car, malheureusement, on peut élaborer une construction intellectuelle séduisante au possible dont le seul défaut est de ne pas correspondre à la réalité ce qui, en matière de justice pénale, est bien le pire qui puisse arriver.

Mais déjà, en parlant du rapprochement judicieux d'éléments factuels épars, nous parlons d'intelligence si nous définissons celle-ci comme l'aptitude à percevoir un rapport ou des rapports entre des éléments distincts. Cependant, l'intelligence humaine a ses limites et il arrive que nous ne parvenions pas à établir ces relations, peut-être parce que, face à des éléments trop nombreux, trop dispersés, nous sommes incapables de percevoir l'ensemble, peut-être parce que, face à des éléments trop hétérogènes, nous sommes incapables de percevoir les lignes de convergence et les points de recoupement ; et pourtant ce sont là choses nécessaires pour construire une image conforme à la réalité. Les époux Cornec, aujourd'hui décédés – fondateurs de la FCPE (Fédération des conseils de parents d'élèves) souvent appelée : Fédération Cornec –, avocats pénalistes l'un et l'autre, racontaient que, pour construire leur argumentation, ils faisaient répliquer par leur propre secrétariat, au moyen de la machine à écrire et du papier carbone (car la photocopie n'existait pas encore), et en multiples exemplaires, l'unique copie du dossier que leur avait délivrée le greffe ; puis, ils découpaient les différents passages de chaque document et utilisaient les fragments comme les éléments d'une sorte de puzzle, non certes pour reconstituer l'image du dossier judiciaire qu'ils voulaient précisément combattre, mais pour constituer une autre image, plus favorable à la défense, mais surtout dont l'existence même démontrait que les éléments du dossier étaient susceptibles de constructions différentes et donc d'interprétations différentes, propres à engendrer au moins un doute quant à l'image présentée par l'accusation puisque, à partir des éléments du dossier, cette image n'était pas la seule possible. C'était une

méthode pour percevoir des rapports que la simple lecture ne mettait pas spontanément en évidence mais n'était-ce pas déjà une forme rudimentaire – Ô ! combien... –, une préfiguration lointaine de l'intelligence artificielle ?

L'informatique telle qu'elle est de nos jours, avec ses ordinateurs, ses logiciels et ses algorithmes, s'avère évidemment, pour une opération analogue, beaucoup plus efficace que la méthode des époux Cornec avec leurs papiers découpés et, bien entendu, elle est déjà utilisée à cette fin. La Gendarmerie notamment, avec le Service central du renseignement criminel (SCRC) s'est beaucoup investie dans ce domaine ; l'opération consiste à entrer la totalité du dossier dans la mémoire d'un ordinateur dans l'espoir que l'utilisation judicieuse d'algorithmes appropriés mettra en évidence des corrélations, des rapports que les enquêteurs n'avaient pas perçus et, si cela permet seulement de relancer sur de nouvelles bases une enquête au point mort, ce n'est déjà pas négligeable. Au demeurant, le système est en voie de développement et d'approfondissement avec la création récente, au sein de la Gendarmerie nationale, de la Division des affaires non élucidées (DIANE) regroupant une vingtaine de personnes et commandée par une lieutenant-colonelle ; et on peut espérer que cette utilisation de l'intelligence artificielle permettra d'aller plus loin en mettant en évidence non seulement des rapports que les hommes n'avaient pas perçus mais des rapports qu'ils n'auraient pas imaginés, et qui, en certains cas, pourraient s'avérer décisifs. Il va de soi que la Police nationale met en œuvre des méthodes comparables.

Encore faudrait-il que les magistrats, et particulièrement les juges d'instruction, fussent formés et ouverts à la compréhension de ces systèmes, raison pour laquelle vient d'être créé un pôle judiciaire à compétence nationale spécialement dédié à ces affaires – à ces *cold cases*, comme on dit –, installé au Tribunal judiciaire de Nanterre avec – pour commencer tout au moins – un groupe de trois juges

d'instruction, trois greffiers et un magistrat du ministère public et qui doit recevoir dès à présent – en ces premiers jours du mois de mars 2022 –, des dossiers venant de toutes les juridictions de France où se trouvent de ces affaires figées. Au demeurant, certains domaines d'activité semblent plus que d'autres propres à l'application de ces méthodes ; on souhaiterait certes que ce fût la grande criminalité contre les personnes – les homicides en série notamment – mais on ne sera pas surpris d'apprendre que, pour l'heure tout au moins, le domaine économique et financier s'y prête mieux et c'est ainsi notamment que l'AMF (Autorité des marchés financiers) en fait grand et fructueux usage pour détecter les opérations boursières frauduleuses.

Cela dit, le concept d'intelligence artificielle fait encore l'objet de débats et sur sa définition même, sur sa fiabilité et sur nombre de problèmes tant techniques qu'éthiques. On parle déjà d'intelligence artificielle forte comme de la possibilité pour une machine de comprendre son propre raisonnement au point d'être suffisamment performante pour concevoir une nouvelle machine plus performante qu'elle-même. En 2016, on a vu aux États-Unis d'Amérique surgir la suggestion – heureusement non suivie d'effet ! – qu'une machine dotée d'intelligence artificielle se présentât aux élections pour tester le potentiel de l'intelligence artificielle en politique ! L'idée même d'un État gouverné par un robot avait donc déjà pu germer dans l'esprit de certains ! Ne nous perdons pas dans les nuages mais la performance déjà considérable et sans cesse accrue des ordinateurs ouvre tout de même de vastes perspectives.

Cela ne signifie pas que tout ira sans problème et, à mon avis, se posera une question : celle de la nature juridique du recours à l'intelligence artificielle. S'agira-t-il de la mise en œuvre d'une méthode d'étude des dossiers comme une autre et parmi d'autres, ou s'agira-t-il d'une opération technique extérieure ? L'étude d'un dossier par un magistrat est une

opération intellectuelle qui relève de la nature même de la mission du ministère public, du juge d'instruction et du juge du fond et de la méthode que le magistrat utilise : qu'il prenne des notes, qu'il établisse des plans, qu'il esquisse des schémas, qu'il confectionne des fiches, qu'il travaille dans un accompagnement de musique ou qu'il ait besoin de silence, qu'il s'enferme dans son bureau ou qu'il réfléchisse en marchant, etc., tout cela ne regarde que lui et il en va de même pour les officiers de police judiciaire qu'il a éventuellement délégués à cette tâche, qu'ils appartiennent à la DIANE ou à tout autre service existant ou à créer. Mais quid, si venait le jour où le magistrat soumettrait le dossier à une machine dotée d'intelligence artificielle mais dont lui-même – ni les officiers de police judiciaire par lui délégués – n'aurait pas la maîtrise ni la disposition, lesquelles appartiendraient, par exemple, à une entreprise spécialisée ? S'agirait-il encore d'une méthode de travail parmi d'autres ? Ou ne s'agirait-il pas plutôt d'une opération technique ordonnée, certes, par la justice mais confiée à un tiers, ce qui est la définition même de l'expertise ? En tel cas ne faudrait-il pas mettre en œuvre la procédure d'expertise prévue par le *Code de procédure pénale* puisque cette procédure n'est que l'application à une opération technique d'un principe fondamental : celui de la contradiction et du libre exercice des droits de la défense ? La réponse à cette question ne dépendrait-elle pas des caractéristiques mêmes du système d'intelligence artificielle mis en œuvre ? Si, par pure hypothèse, il n'existait, pour une telle application, qu'un seul et unique algorithme ou un seul ensemble d'algorithmes, si la technique était telle que, sur un ensemble déterminé de données, toute mise en œuvre de l'intelligence artificielle ne pourrait aboutir qu'à un seul et même résultat, alors la machine ne serait qu'un instrument de travail comme un autre dont le ministère de la Justice devrait doter ses juridictions, en assurant évidemment les formations adéquates. Mais au contraire, si, du fait de techniques de mise en œuvre différentes, du fait de l'utilisation d'algorithmes différents, des

résultats différents pouvaient être obtenus, alors, je ne vois pas comment on pourrait faire l'économie de la procédure légale et contradictoire d'expertise et de contre-expertise. Qu'en sera-t-il ? Il va de soi que je n'en sais absolument rien et c'est bien pour cela que je me pose la question.

Le second domaine : la détermination de la sanction

Déterminer la sanction est loin d'être facile. Évidemment, il faut tenir compte d'abord du fait lui-même : crime ou délit, des circonstances qui l'ont entouré, de ses conséquences pour la victime et pour la communauté et, là, on est dans le domaine des données objectives, voire sous certains aspects, quantifiables ; mais il faut tenir compte également de la personne du criminel ou du délinquant puisque c'est à lui que s'appliquera la décision à prendre, il faut considérer les causes endogènes ou exogènes qui ont conduit à l'infraction et il faut enfin évaluer l'effet qu'aura sur l'auteur, la peine prononcée : le fera-t-elle monter vers le bien et sortir de la délinquance ? Cela arrive, fort heureusement ; ou au contraire, le fera-t-elle descendre un peu plus – ou beaucoup plus – vers le mal et l'enfoncera-t-elle davantage dans la délinquance, voire dans la criminalité ? Cela arrive aussi, fort malheureusement. Et, pour compliquer le problème, s'ajoute à ces questions déjà fort ardues, celle de l'exemplarité, c'est-à-dire celle des possibles effets de la décision sur les criminels ou les délinquants potentiels. Le problème n'est pas nouveau et saint Augustin lui-même – qui, comme tous les évêques de son temps, exerçait des fonctions judiciaires – l'exprimait mieux que quiconque :

« Quand il s'agit de punir ou de ne pas punir, quel embarras ! Tout ce que nous voulons, c'est contribuer au salut de ceux à l'égard desquels nous nous décidons à user de rigueur ou d'indulgence. Et la mesure à garder dans la punition – car il faut tenir compte non seulement de la nature et du nombre des fautes, mais des forces morales de chacun, de ce que le délinquant peut ou non supporter ;

autrement c'est l'arrêter dans ses progrès et même l'exposer à des chutes – quelle tâche mystérieuse, énigmatique... »

Et il ajoutait :

« J'avoue que dans ces matières, je me trompe tous les jours...⁴ »

Autrement dit, face à la délinquance et à la criminalité considérées non comme des concepts sociologiques mais comme des réalités incarnées en des êtres humains, saint Augustin ne savait pas quoi faire et, seize siècles plus tard, nous ne le savons pas mieux que lui. Pourtant, nous disposons de beaucoup plus d'éléments d'appréciation : dans la plupart des affaires de quelque gravité – toutes les affaires criminelles, notamment –, nous avons des renseignements de comportement, des enquêtes de *curriculum vitae* depuis l'enfance jusqu'au jour de l'infraction, des enquêtes sociales, des expertises psychologiques, des expertises psychiatriques, sans parler, évidemment, des renseignements complets sur les antécédents judiciaires ; mais le problème est celui de l'utilisation de tous ces éléments, rapprochés des éléments constituant ou entourant l'infraction elle-même, pour déterminer la sanction la plus adéquate, la solution la meilleure à choisir dans le champ presque illimité – comme nous l'avons vu – des solutions juridiquement possibles.

Le problème n'a jamais été simple. La sanction en tant que telle repose sur trois fondements théoriques : la vengeance, la purification et l'utilité mais gardons-nous ici de toute discussion d'ordre philosophique et ne considérons qu'un seul objectif pratique : l'utilité. L'utilité pour le condamné qui consiste à l'extraire du cercle infernal de la délinquance et de la criminalité où il risque de s'enfermer, où il est peut-être déjà enfermé pour son plus grand mal, et l'utilité pour la société qui

4 Cité par Henri MARROU, *Saint Augustin et l'augustinisme*, Paris, Éd. Seuil, coll. Les maîtres spirituels, 1956.

consiste à protéger ses membres de la délinquance et de la criminalité en évitant, d'une part, que le coupable en cause ne récidive – d'où l'individualisation de la sanction – et, d'autre part, que d'autres ne soient tentés de l'imiter – d'où l'exemplarité de la sanction, individualisation et exemplarité qui ne font pas toujours bon ménage et saint Augustin l'écrivait déjà :

« En punissant un coupable parfois on le perd et en ne le punissant pas parfois on en perd un autre. »

Et en ce domaine encore, que pourrait apporter l'intelligence artificielle ? Pourrait-elle nous aider comme elle nous aurait aidés en la recherche de la vérité, soit en nous proposant des solutions construites et intégrant toutes les données utiles ou peut-être, seulement et plus modestement, en proposant des orientations ? Peut-on envisager l'intervention de la machine en ce domaine ? À priori : pourquoi pas ? La pratique serait la même : entrer dans la mémoire de la machine dotée d'intelligence artificielle tous les éléments disponibles, d'une part sur la matérialité des faits et d'autre part sur les caractéristiques personnelles de leur auteur : son psychisme en ses différentes composantes, son quotient intellectuel, son quotient émotif, sa volonté, etc., et sur ses caractéristiques sociologiques : son éducation, son milieu, son activité, sa famille, etc. pour obtenir, grâce à des algorithmes appropriés, une réponse adaptée.

Sur ce point, selon des renseignements qui m'avaient été fournis lors d'une conférence faite par un magistrat et confirmés sommairement par des recherches sur Google, il existait au moins une nation qui avait commencé à utiliser ces procédés dans un cadre judiciaire réel et c'était l'Estonie ; il était prévu, semble-t-il, que pour les délits ayant entraîné un préjudice inférieur à sept mille euros, le soin de décider de la culpabilité de la personne mise en cause serait confié à l'intelligence artificielle secrétée par une machine – avec cependant la possibilité d'en appeler devant des juges ; le but officiel n'était pas, apparemment, d'obtenir une justice plus

sûre mais d'exonérer les magistrats du soin des affaires les moins importantes pour leur permettre d'accorder davantage de temps et de soin aux affaires les plus graves et, vu sous cet angle, le projet paraissait surtout d'ordre managérial. Pour être documenté de manière plus précise, je me suis adressé à l'ambassade d'Estonie mais à ma missive postale – fort civile, vous vous en doutez – j'ai reçu une réponse télématique non exempte de quelque sécheresse et que je reproduis ici fidèlement et dans son intégralité : « Cher Monsieur, Faisant suite à votre courrier concernant l'application de l'IA aux décisions de justice dans la République d'Estonie, je vous informe qu'à l'heure actuelle l'intelligence artificielle n'est pas appliquée dans les systèmes d'information de justice et toutes les décisions sont prises par les humains dans les affaires pénales et civiles et dans tous les stades de la procédure. – Bien cordialement. ». Ai-je mauvais esprit ? Mais cette réponse m'a quelque peu murmuré à l'oreille l'idée que des essais avaient été faits et n'avaient pas donné satisfaction.

Une autre expérience est conduite actuellement aux États-Unis d'Amérique dans quelque soixante juridictions avec un logiciel dit COMPAS (*Correctionnal Offender Management Profiling for Alternative Sanctions*) fondé sur un questionnaire comportant cent quarante-sept questions différentes appelant chacune une réponse sous forme de cotation de 1 à 10 et ce afin de calculer, d'évaluer mathématiquement, le risque de récidive, étant précisé que la décision est, en tout état de cause, prise par un juge et en aucun cas par le système informatique lui-même. Or il semble que les résultats soient décevants au point qu'on a pu accuser le système de racisme, la prévision du risque de récidive s'étant avérée deux fois supérieure pour les justiciables d'origine afro-américaine ! Était-ce imputable au choix de questions ? au choix des algorithmes ? Je ne sais mais, quelle que soit la cause, l'expérience n'a rien de rassurant.

On m'objectera peut-être qu'une procédure pénale entièrement automatisée et sans intervention humaine existe en

France dans un domaine très limité de la circulation routière : celui de l'excès de vitesse où l'infraction est relevée sans intervention humaine par un radar fixe ou embarqué, les données transmises automatiquement à un centre de traitement unique où tout se règle par voie informatique avec des logiciels et des algorithmes appropriés : identification du titulaire de la carte grise, retrait des points sur le permis de conduire, envoi de l'avis d'infraction avec indication de l'amende à payer, acquittement de l'amende par le contrevenant au moyen d'un timbre à apposer sur le formulaire à retourner ce qui met un terme définitif à la procédure, une procédure où l'intervention humaine est inexistante : la preuve de l'infraction a été rapportée par la captation photographique opérée par le radar, l'identité de l'auteur de l'infraction a été faite automatiquement par rapprochement avec le fichier national des immatriculations, les sanctions : l'amende et le retrait de points étaient l'une et l'autre prédéterminées et fixes. Peut-on, en ce domaine parler d'intelligence artificielle ? Si, encore une fois, on définit l'intelligence comme l'aptitude à déceler des rapports, certes on aura établi un rapport entre une vitesse donnée et une vitesse permise, entre une plaque minéralogique et le propriétaire de la voiture, entre l'infraction et la sanction mais, chaque opération est, en elle-même, tellement simple, rudimentaire presque, que j'hésite vraiment à y reconnaître la marque de l'intelligence artificielle. Et il n'en reste pas moins – notons-le au passage – que cette procédure peut être contestée par celui qui en est l'objet et, en ce cas, l'affaire est de la compétence d'un magistrat, d'un juge du tribunal judiciaire du lieu de l'infraction avec la procédure normale prévue pour les contraventions. Dans ce type d'affaires, il n'y a donc pratiquement rien à juger et pourtant le recours à un juge reste possible. Mais si on tient à y voir une application de l'intelligence artificielle, alors je me demande si la simplicité et la sécheresse du processus mis ici en œuvre ne souligne pas, *a contrario*, l'inadéquation totale d'un tel processus pour des délits, voire pour des crimes, dont l'élucidation peut être

complexe et dont la sanction exige la prise en compte d'un bon nombre de paramètres d'ordre matériel et surtout d'ordre humain.

C'est pourquoi on ne peut que s'interroger sur les problèmes qui se poseront si, un jour, une méthode fondée sur la mise en œuvre de l'intelligence artificielle peut fonctionner, ce qui n'est pas garanti à l'avance ; certes, en quelques décennies, l'informatique a fait des progrès incroyables, et certes, dans les décennies à venir, elle en fera de nouveaux tout aussi incroyables, mais cela ne signifie pas qu'elle pourra tout faire ni tout régler et, notamment, dans le domaine qui nous intéresse. Certes les robots assimileront un nombre croissant de variables exploitées avec des algorithmes d'une sophistication sans cesse accrue mais, tout de même, cela ne suppose-t-il pas l'existence d'un certain déterminisme dans les phénomènes traités ; or précisément, le comportement humain et donc le comportement de l'auteur de l'infraction n'est pas entièrement déterminé car, s'il l'était, l'être humain – et par conséquent le juge lui-même tout comme le délinquant – ne serait qu'une machine comme une autre, un peu plus perfectionnée en certains domaines, nettement moins en d'autres, mais, enfin, une machine. Or, entre un robot et un homme, il y a – me semble-t-il – une différence non de degré mais de nature, car l'Être humain possède une liberté intérieure, un libre-arbitre qui lui confère sa dignité de personne, de sujet de droit et non d'objet de droit ; *et a priori* je ne vois pas très bien comment la liberté intérieure et le libre-arbitre seraient intégrables dans les algorithmes et dans les logiciels. En outre, s'agissant non plus de la culpabilité qui relève du passé mais de la sanction qui relève de l'avenir, on ne peut oublier que le physique et le psychisme de l'homme ne sont pas statiques mais évolutifs et l'évolution comporte toujours quelque chose d'aléatoire ; il suffit parfois d'une rencontre, d'ordre amical ou amoureux, pour réorienter le psychisme d'une personne et changer ses habitudes et son

mode de vie et, à mon avis, une machine ne pourra jamais garantir le succès d'une sanction qu'elle proposerait, pas plus qu'un juge n'a jamais pu et ne peut le garantir, quel que soit le soin qu'il apporte à élaborer sa décision.

Au demeurant se poserait ici le problème que nous avons rencontré précédemment : celui de la nature du recours au robot : méthode de travail interne à la juridiction ? Ou expertise judiciaire avec la procédure qu'elle comporte et qu'elle doit comporter ?

Mais enfin, et toujours par pure hypothèse, admettons encore un instant que l'application à la sanction pénale de l'intelligence artificielle puisse, un jour, fonctionner en ce double domaine de la recherche de la vérité et de la sanction : si un robot suffit à assumer ces deux tâches, alors, il n'y aura plus d'audience (et donc plus de robes... aïe ! La Maison Bosc et la Maison Gérin ! Elles vont devoir se reconvertir dans les blouses de chirurgiens ou les scaphandres pour cosmonautes). Cela ne veut pas dire qu'il n'y aura plus de défense mais le rôle de celle-ci sera précisément d'apporter des éléments que la machine devra intégrer comme elle intégrera les autres éléments du dossier et notamment ceux apportés par l'accusation ; tout sera écrit – et même dématérialisé – et comme le disait, il n'y a pas si longtemps, M. Badinter, actuel sénateur, ancien garde des Sceaux, ministre de la Justice et qui avait été un grand avocat et un grand orateur de cour d'assises : « La plaidoirie de Papa, c'est fini... ». Pour l'heure et indépendamment de la question de l'intelligence artificielle – ce semble être déjà la tendance au moins en matière civile où il est de plus en plus question de plaider par observations... Mais il s'agit là d'un tout autre domaine étranger à mon propos ; il n'en est pas encore question en matière pénale mais si l'intelligence artificielle du robot devait, un jour, se substituer à l'intelligence naturelle du magistrat, s'il ne s'agissait plus que d'introduire des données dans une machine, l'expression orale devenue inadaptée serait bannie des

juridictions alors que, jusqu'à présent, le principe juridique fondamental, en matière pénale, est celui de l'oralité des débats, le dossier n'étant que le support à partir duquel se construit le débat oral, celui à l'issue duquel et en fonction duquel est prise la décision. Qu'on songe notamment à la cour d'assises, où seul le président détient et connaît le dossier qui lui est évidemment nécessaire pour la conduite des débats, mais dossier auquel n'ont accès ni les assesseurs ni les jurés au point que, lorsque la cour et les jurés se retirent en chambre du conseil pour délibérer sur la culpabilité et éventuellement sur la peine, le président dépose publiquement et solennellement le dossier entre les mains du greffier qui le conserve dans la salle d'audience. Et si l'oralité est, aujourd'hui encore, un principe fondamental, ce n'est pas sans raison ; en toute matière, les dossiers – qu'on appelait jadis des sacs – existent depuis des siècles mais, en matière pénale, l'oralité est une conquête de la Révolution française et depuis ce temps, le législateur a toujours tenu à ce que la justice des hommes présente un caractère humain et c'est parce que la relation entre des êtres humains est plus dense, plus proche, plus humaine en un mot, si elle s'exprime dans le face à face, avec une diction, une éloquence, un schème mélodique, une gestuelle – amplifiée par la robe –, avec l'accent de la colère ou l'accent de la compassion, avec les mille nuances qui font la richesse de la relation humaine. Mais il est bien évident que si la décision doit être prise par une machine, tout cela sera évacué. Alors justice déshumanisée ? À coup sûr... Mais si, toujours par pure hypothèse, le robot doté d'intelligence artificielle devait rendre des décisions objectivement meilleures que les décisions rendues par les hommes alors le choix ne serait-il pas entre une meilleure justice robotique et une moins bonne justice humaine ? Face à une telle alternative – en supposant qu'elle puisse se produire – on ne pourrait formuler de grief à l'encontre de celui qui choisirait la première solution et pourtant ce serait le dernier pas, et un pas de géant, vers une déshumanisation totale de l'activité

judiciaire, déjà amorcée d'ailleurs. Car, dans les palais de justice actuels, dans certains tout au moins, la nouvelle organisation due précisément à l'informatisation croissante des tâches procédurales et secrétaires, distend les relations entre les acteurs professionnels : magistrats, avocats, fonctionnaires, au point que, dans le tout nouveau et très avant-gardiste palais de justice du Tribunal judiciaire de Paris ⁵, ces relations s'étaient tellement effritées d'abord, détériorées ensuite et aigries finalement, comme cela n'avait jamais été, qu'il a fallu organiser des réunions interprofessionnelles pour rétablir un contact supportable et on peut donc s'interroger avec inquiétude sur une justice robotisée en tout ou en partie.

Bien entendu, dans une justice pénale rendue par une machine définie comme infaillible, et bien que le double degré de juridiction soit un principe fondamental de notre procédure pénale, expressément confirmé par la Convention européenne des droits de l'homme, il n'y aurait plus de place pour les voies de recours car, en appeler d'une machine infaillible à une autre qui, par définition, le serait tout autant, n'aurait aucun sens ; il en découlerait donc une suppression des juridictions d'appel devenues inutiles ! Et même de la Chambre criminelle de la Cour de cassation ; autrement dit, bouleversement complet de l'organisation judiciaire ! Et bien entendu, l'institution du jury disparaîtrait puisque, par définition, il n'y aurait plus d'audiences. Mais enfin, cela, c'est vraiment l'hypothèse extrême, du rêve, de l'utopie... : un monde où la décision judiciaire sur la culpabilité et sur la peine serait rendue par un robot tenu pour infaillible.

Est-il possible d'envisager des solutions moins radicales ? Nous en avons déjà évoqué une, celle où le recours au robot serait considéré comme une procédure d'expertise avec les

⁵ La *Cour d'appel de Paris* et la *Cour de cassation* sont demeurées dans l'île de la Cité.

possibilités qui en découleraient de contre-expertise. Mais on peut aussi imaginer que le robot n'ait qu'un rôle de proposition et que sa proposition soit soumise à un juge ou à une juridiction composée de juges qui auraient alors le pouvoir ou d'entériner la décision informatique ou de la rejeter purement et simplement ou même encore, de la modifier. Au fond, le robot jouerait alors le rôle d'une sorte d'instance de premier degré par rapport à laquelle le juge ou les juges conserveraient le rôle qui est normalement le leur : celui de rendre la justice ; ce serait peut-être quelque peu analogue – *mutatis mutandis* – à l'actuelle procédure dite de plaider-coupable où la personne poursuivie du chef d'une infraction comparaît devant le procureur de la République, non pas pour négocier la peine comme cela se fait aux États-Unis d'Amérique, mais pour accepter ou refuser la peine proposée par le magistrat et, au cas d'acceptation, l'affaire est obligatoirement déférée à un juge qui entérine ou non l'accord qu'on lui présente. Peut-être pourrait-on imaginer que le magistrat chargé de l'accusation publique et la personne poursuivie, l'un et l'autre satisfaits de la proposition formulée par le robot, aient le pouvoir de renoncer à la comparution devant le juge et en ce cas, la proposition du robot se trouverait, de plein droit, convertie en décision de justice, mais encore faudrait-il ménager, pour le ministère public comme pour la personne poursuivie, un temps de réflexion : soit qu'on leur accordât un délai entre la proposition du robot et leur décision de comparaître ou de ne pas comparaître devant les juges, soit qu'on exigeât qu'ils se prononcent immédiatement mais en leur accordant un délai de repentir pour pouvoir, éventuellement, revenir sur leur décision.

Peut-être aussi pourrait-on imaginer de scinder les deux aspects du procès ? Par exemple, le robot statuant sur la déclaration de culpabilité, le coupable acceptant cette sentence et, ensuite, le juge ne statuant que sur la sanction ? ou l'inverse : le juge statuant sur la culpabilité et, celle-ci une fois acquise, le robot proposant une sanction que l'intéressé

pourrait accepter ou refuser, cas auquel le juge statuerait ? Mais rappelons-nous, un instant, que l'expérience faite actuellement aux États-Unis d'Amérique et ci-avant rapportée n'a absolument rien de rassurant. Bien sûr, on peut tout imaginer puisque rien ne nous interdit de délirer et pourtant, de nouvelles techniques encore plus extraordinaires pointent déjà et je veux parler de nouvelles formes d'intelligence artificielle.

Les nouvelles formes d'intelligence artificielle : les jumeaux numériques et Métavers

L'intelligence artificielle nous apparaît déjà comme presque irréaliste tant nous avons de la peine à concevoir que des machines puissent imiter qualitativement et dépasser quantitativement les possibilités de l'esprit humain. Et pourtant la question se pose : cette toute récente intelligence artificielle n'est-elle pas dès maintenant en voie d'être dépassée et peut-être même dépassée et par les jumeaux numériques et surtout par *Métavers*.

Les jumeaux numériques

Ce qu'on appelle jumeaux numériques existe déjà et comporte des applications pratiques : c'est la copie numérique conforme d'une réalité qui permet des études et des expérimentations infiniment moins onéreuses et moins risquées en cas d'échec. C'est ainsi que *Dassault-Systèmes* n'emploie plus de maquettes matérielles, ni pour des études d'aérodynamisme ni pour des études d'usure. Mais cet exemple ne doit pas faire croire que la méthode ne s'applique qu'à l'ingénierie mécanique ; elle peut s'appliquer à tout, indépendamment des problèmes d'échelle et on peut aussi bien créer le jumeau numérique d'un organe humain que le jumeau numérique d'une grande ville ; actuellement des villes comme Angers, Rennes, Dijon ou, à une toute autre échelle, Singapour utilisent cette technique à des fins d'études urbanistiques, pour la gestion de chantiers, l'optimisation de la consommation

d'énergie, la lutte contre la pollution, etc., et, dans un tout autre domaine, la médecine commence à créer des jumeaux numériques de patients pour l'étude de prothèses.

Peut-on envisager des applications en matière de droit pénal ? À une époque où on parle de traitement pénologique autant que de peine, plutôt que de peine, ne pourrait-on imaginer une application consistant à créer un jumeau numérique du délinquant comme la médecine crée des jumeaux numériques de patients pour observer l'action de telle ou telle mesure, éliminer celles qui s'avèreraient inadaptées voire nocives et choisir celle qui paraîtrait la plus adéquate ? Ainsi prononcerait-on une peine à coup sûr ou presque ce qui, jusqu'à ce jour est un idéal inaccessible car on ne sait jamais les effets que comportera la mesure qu'on choisit. À cela, certes, on ne manquera pas d'objecter que les problèmes ne sont pas les mêmes, que les problèmes médicaux relèvent de la physiologie et de la nosologie et les problèmes judiciaires de la psychologie et de la sociologie, encore qu'il existe bien des interpénétrations entre les divers domaines. Science-fiction ? Science, je ne sais pas..., fiction, oui, certes, à ce jour, mais dans quelques décennies ? Pour l'heure, ces méthodes exigent d'énormes puissances de calcul très voraces en énergie et il y a encore beaucoup de chemin à parcourir pour disposer d'algorithmes adaptés à tous les problèmes mais nous ne sommes peut-être et même sans doute qu'au tout début. Quand j'ai commencé ma carrière, en 1953, les mots : informatique, ordinateur, logiciel, algorithmes étaient inconnus ; mais lorsque j'ai terminé ma carrière en 1995, j'avais un ordinateur sur mon bureau et je l'utilisais tous les jours. On ne peut donc aujourd'hui apporter aucune réponse ; tout au plus est-on en droit de considérer la question comme potentiellement pertinente et de la formuler.

Métavers

Ce qu'on appelle Métavers, pour l'heure, semble encore relever du domaine de la fiction et pourtant certaines firmes

envisagent déjà la mise en œuvre de ce système dans lequel elles perçoivent le successeur d'Internet. Le concept paraît encore quelque peu flou, sans définition univoque ni universellement acceptée ; on parle de réalité virtuelle ce qui me paraît constituer un oxymore, ou de réalité augmentée, ce dernier concept exprimant la superposition d'objets virtuels dans un environnement réel. Toutefois, une définition plus précise a été donnée : « Vaste réseau de mondes et de simulations 3D persistant, rendus en temps réel, qui assurent la continuité de l'identité des objets de l'histoire, des paiements et des droits, et qui peuvent être expérimentés de manière synchrone par un nombre effectivement illimité d'utilisateurs, chacun ayant un sentiment de présence individuelle. » Autrement dit une chose à la disposition de tous mais que chacun peut s'approprier ? On a déjà rencontré des définitions moins prolixes et plus claires ! Concrètement, Métavers permettrait aussi bien d'assister à un concert virtuel, que de faire un voyage en ligne, ou d'essayer des vêtements numérisés, et cela sans quitter son domicile ni même son fauteuil à la condition d'être équipé d'accessoires haptiques telles les lunettes de réalité augmentée ou même – et cela existe déjà – des gants permettant la sensation de toucher, voire de saisir, des objets purement virtuels. Certains vont encore plus loin puisqu'ils imaginent une cryptomonnaie virtuelle engendrant une économie non moins virtuelle et autosuffisante. Où en est-on ? Chacun s'accorde à dire qu'à ce jour, l'immersion dans la réalité virtuelle est loin d'être assez performante pour obtenir autre chose et plus qu'une sorte de jeu-vidéo perfectionné. Pourra-t-on aller plus loin ? Certains en doutent et considèrent que ce projet relève du rêve et de l'utopie tandis que d'autres, même s'ils voient la route encore longue et non exempte de difficultés, sont persuadés qu'une dizaine d'années de recherche conduira à bon port. Et déjà certains s'alarment des risques et dangers inhérents à un tel système : l'évolution dans un monde virtuel et donc irréel serait-elle sans retentissement sur le psychisme ? Le retour au

monde réel ne pourrait-il déstabiliser le sujet ? Les passages répétés de l'un à l'autre ne seraient-ils pas susceptibles d'engendrer des troubles de la personnalité ? Et pourtant, sans plus tarder, Facebook a mobilisé des capitaux considérables pour s'engager dans l'aventure et la mener à bien, mais déjà il n'est plus seul et Microsoft vient d'acquérir pour soixante-neuf milliards de dollars un éditeur mondial de jeux-vidéo, ce secteur devant jouer « un rôle-clef dans le développement des plateformes du Métavers », sachant que Facebook et Microsoft cultivent, pour l'instant tout au moins, des visions différentes des évolutions prochaines.

Là encore, admettons, toujours par hypothèse, que Métavers réussisse ; pourrait-on imaginer des applications pour la justice pénale ? Une me vient à l'esprit : au cas de peine privative de liberté prononcée par la juridiction de jugement ou de détention provisoire ordonnée par le juge des libertés et de la détention, ne pourrait-on pas concevoir une incarcération virtuelle dans un milieu réel ? Autrement dit, la personne condamnée ou mise en détention provisoire resterait physiquement chez elle et, cela, nous savons déjà le faire et nous le faisons déjà : songeons au contrôle judiciaire avec assignation à domicile dans lequel le contrôle est assuré par le bracelet électronique ; mais une application adéquate de Métavers, plongerait la personne condamnée à une peine privative de liberté ou placée en détention provisoire dans une situation virtuelle qui serait celle du milieu carcéral.

Ainsi, alors que l'environnement réel mais non perçu serait bien celui du domicile, l'environnement virtuel mais perçu serait celui de l'établissement carcéral. Économiserait-on ainsi la construction et l'entretien – l'une et l'autre fort onéreux – de bâtiments pénitentiaires ? Économiserait-on pareillement le financement – non moins onéreux – des agents de l'administration pénitentiaire qui seraient remplacés par un nombre limité d'acteurs évoluant dans des décors ? Encore faudrait-il s'assurer que le pseudo-détenu conserverait bien sur lui les équipements haptiques sans lesquels ne serait plus perçu

l'environnement virtuel qui, seul, aurait un caractère carcéral et donc pénal mais ce problème serait certainement soluble du point de vue technique. Déjà le bracelet électronique permet de s'assurer de la présence de l'intéressé à son domicile et la technique permettrait sans doute de s'assurer qu'il garde sur lui les équipements haptiques tout au moins pendant un nombre d'heures quotidien suffisant pour que l'incarcération virtuelle ait du sens et soit ressentie comme punitive. Il va sans dire qu'à ces questions et à quantité d'autres qui se poseraient, je suis tout à fait incapable d'apporter le moindre commencement de réponse. Mais, encore une fois, lorsque, par hypothèse, on admet l'avènement de Métavers, il est légitime de s'interroger sur toutes les applications concevables et, donc, en ce qui me concerne, sur l'application à la privation de liberté qui, malgré les nombreux modes de diversification des sanctions, demeure encore – tout au moins pour l'instant – le mode central, l'outil essentiel de la politique pénale ; n'a-t-on pas appelé la prison : « la bonne à tout punir » ? Mais quand j'écris : « lorsque, par hypothèse, on admet l'avènement de Métavers », ne suis-je pas déjà en retard ? Une chose me frappe : lorsque j'ai commencé à travailler sur ce sujet, en novembre 2021, Métavers était présenté comme une idée à approfondir et pour certains presque comme une utopie ; or nous sommes en mars 2022 et il ne se passe pas de jour où la presse ne nous informe que telle firme, tel organisme, travaille à des applications concrètes de Métavers dans son domaine de compétence ; les choses semblent progresser à une vitesse folle.

Mais passons au second thème : le transhumanisme.

QUE POURRAIT APPORTER LE TRANSHUMANISME ?

« Bonne question », dirai-je, mais, auparavant, il faut, comme toujours, définir les mots qu'on emploie et, s'agissant du transhumanisme, ce n'est pas simple.

Qu'entendre par transhumanisme ?

Certes, l'étymologie est claire, la préposition latine et le préfixe latin : *trans* n'exprimant aucune idée de supériorité ni d'infériorité mais une idée de latéralité, de déplacement horizontal d'un domaine à un autre : *transport*, *transfert*, *translation*, ce qui peut parfois se traduire par une idée d'*au-delà* mais toujours dans un plan horizontal : *transalpin*, *transatlantique*. Le *transhumanisme* ne saurait donc consister à créer un *hyperanthrope*, un *surhomme*, un *superman* mais un *homme autre*, agissant et réagissant *autrement* et, si possible : *mieux* que l'homme actuel, sans quoi l'opération serait sans intérêt. Vu ainsi, le concept est plutôt rassurant et séduisant car quoi de plus légitime, pour chacun et pour tous, qu'une amélioration des caractéristiques physiques et psychiques de tous et de chacun ? Et c'est bien ce que proclame l'Association française transhumaniste (je cite) : « Prôner l'usage des sciences et des techniques afin d'améliorer les caractéristiques physiques et psychiques humaines. » Mais vu ainsi, cela n'a rien de nouveau : c'était déjà ce qu'on faisait ou ce qu'on espérait faire lorsqu'on administrait à un enfant de l'huile de foie de morue ou un fortifiant quelconque ou lorsqu'on le punissait pour améliorer son comportement. Tout dépend de ce qu'on veut faire et de la manière dont on veut le faire ; lorsqu'on lit qu'il s'agit d'améliorer la résistance à la maladie – ce que font déjà les vaccins et la vitamine C –, d'améliorer le contrôle des désirs, des humeurs et des états mentaux – ce qu'on peut tenir pour synonyme de maîtrise de soi –, c'est très bien ; lorsqu'on complète l'énumération par les mots – je cite encore : « jeunesse éternelle », on commence à se méfier ; mais lorsqu'il est question – je cite toujours – de « soustraire l'homme à la mort », on se demande si on n'aborde pas le royaume du délire et, de fait, on peut dire tout et n'importe quoi en bâtissant un vocabulaire ésotérique, en parlant d'extropianisme et de bien d'autres choses jusques et y compris – je cite toujours – l'invention d'un dieu-post-humain (?) et, plus concrètement aux USA de supersoldats (!). Quant

à la manière dont on veut le faire, quand on parle d'utiliser les nanotechnologies, les biotechnologies, les neurosciences, les électrodes et la fusion informatique, on commence légitimement à s'interroger sinon à s'inquiéter. Ce qui paraît certain, c'est qu'actuellement rien n'est encore établi, fixé, normé ; tant qu'il s'agit de publier des articles et de faire des conférences, tout, absolument tout peut être envisagé et si l'auditeur ou le lecteur est d'un naturel optimiste, il se prend à fredonner l'*Hymne à la joie* sur la musique de Beethoven et s'il est d'un naturel pessimiste, il se dit – reprenant l'expression d'Albert Londres – que « Dante lui-même n'avait rien vu », cependant que d'autres sceptiques par nature parlent purement et simplement d'imposture.

À ce jour, à mon avis, on ne sait donc rien ou pas grand-chose ; mais ce qui est à peu près certain, c'est qu'il y aura des avancées en ce sens, voire des réalisations concrètes dans les décennies qui viennent et que certaines au moins seront bonnes ; et voilà qui m'amène à me poser la question : Pourra-t-on utiliser ces méthodes ou certaines d'entre elles dans le cadre de la justice pénale pour changer le comportement des condamnés et éviter la récidive pour le plus grand bien du condamné lui-même et de la société qui l'entoure ?

Transhumanisme et sanction pénale

Au risque de surprendre, je dirai que cela, la justice a toujours voulu le faire. Évidemment, lorsqu'elle prononçait une peine capitale ou une peine de travaux forcés à perpétuité – ou même à temps – qui envoyait le criminel mourir sous le ciel de l'équateur et dans les marécages de la Guyane, elle n'avait pas pour objectif de transformer le criminel en honnête homme ; mais, déjà, lorsqu'elle prononçait une peine métropolitaine temporaire (l'emprisonnement) ou ponctuelle (l'amende), elle espérait bien ne pas revoir le condamné sous les traits d'un récidiviste ; au contraire, elle espérait que la peine – c'est-à-dire la souffrance – qu'elle infligeait servirait de leçon et corrigerait le délinquant comme

l'indiquaient les épithètes toujours en usage dans les locutions de tribunal correctionnel et de peine correctionnelle.

Et corriger le délinquant, faire de lui un citoyen vivant dans la société sans susciter de troubles, c'était bien le but recherché pour le délinquant lui-même, certes, mais également pour la société car je ne connais pas de meilleur moyen, pour débarrasser la société d'un homme malhonnête, que de le transformer (voilà le préfixe : trans !) en homme honnête. Le malheur c'est qu'on n'a jamais su le faire ; non certes que cela ne se soit jamais produit : bien des délinquants primaires n'ont pas récidivé et certains récidivistes ont fini par sortir de la délinquance et, certes, ce n'était pas, ce n'est pas à négliger ni à dédaigner car même si cela ne touche qu'un petit pourcentage – si j'ose m'exprimer en termes de statistique – c'est déjà considérable car ce pourcentage, si modeste soit-il, représente des êtres humains et le sauvetage moral et social d'un être humain n'a pas de prix ; mais, sur le plan de la société, le compte n'y est certes pas car, pour que la justice pénale fût socialement efficace, il faudrait qu'elle pût sauver sinon tous les délinquants – ne rêvons pas ! – mais au moins un grand nombre, le plus grand nombre ; or elle ne l'a jamais fait parce que – encore une fois – elle n'a jamais su le faire et saint Augustin, déjà cité, s'en plaignait : « quelle tâche mystérieuse, énigmatique... dans ces matières, je me trompe tous les jours » ; au moins avait-il le mérite de reconnaître son incapacité, ce qui n'a pas toujours été et n'est pas toujours le cas alors que nous ne réussissons pas mieux que lui.

De tout temps, on a parlé d'amender le délinquant ou le criminel c'est-à-dire de le rendre autre par ablation de quelque chose, mais on a fini par faire un contresens sur la notion d'amendement, ce mot suscitant une idée d'apport, d'adjonction alors que c'est plutôt l'inverse ; bien sûr, on amende un sol trop acide en apportant des marnes ou de la chaux, on amende une loi, généralement en y ajoutant quelque chose et, au Parlement, les innombrables amendements proposés par les parlementaires lors du vote des textes

législatifs consistent presque toujours en l'adjonction d'un article ou d'un paragraphe. Mais étymologiquement le mot amendement est construit à partir du préfixe latin *ab* qui exprime une idée d'éloignement, de suppression comme dans le mot ablation (littéralement : action de porter hors de...) et à partir du substantif : *mendum* qui signifie : défaut ; amender une chose, c'est la débarrasser d'un défaut et cela peut se faire par adjonction comme l'élimination de l'acidité par apport d'une base, mais cela peut se faire aussi par enlèvement.

Qu'en est-il dans le sujet qui nous occupe ? Certes l'amendement, entendu comme l'ablation d'un défaut ayant pour propriété de pousser à une forme de délinquance, s'est fait parfois et longtemps par ablation physique ; en France même et à certaines époques ou en certaines coutumes, le viol était puni par la castration et en Islam le vol a été puni et semble l'être encore dans certains pays où règne la *Charia*, par amputation de la main. Parfois il est vrai, l'ablation n'était que symbolique : telle fut, à la Libération en 1944 (1945 pour certaines régions) la tonte des femmes à qui on reprochait leur inconduite avec les occupants, la faute qu'on leur imputait étant d'ordre sexuel et la chevelure féminine étant un symbole sexuel, l'ablation physique de la chevelure symbolisait l'ablation morale du facteur de mauvais comportement. Pourtant, avec cette sorte de *catharsis* symbolique, on n'était plus vraiment dans le domaine de l'amendement puisque la récidive n'était plus à craindre du fait du départ des occupants et on était plutôt dans une idée de vindicte, notion qui, elle aussi, a sa place en droit pénal. Mais tout cela n'est plus à l'ordre du jour et le problème quotidien est de savoir comment amender les voleurs en leur ôtant l'appétence pour le bien d'autrui, comment amender les incendiaires en leur ôtant l'appétence pour le feu, comment amender les agresseurs en leur ôtant l'appétence pour la violence, étant bien entendu, qu'il faut éliminer les cas d'ordre purement pathologique qui

relèvent non de la justice pénale mais de la médecine psychiatrique.

En ce qui concerne la justice pénale, il y a longtemps qu'elle cherche à transformer le délinquant. Pour ne parler que des temps modernes et contemporains, partons du *Code pénal* de 1810 qui, lui, ne voyait qu'une méthode : la sévérité de la peine. Je l'ai dit : le vol d'un objet ou d'une somme excédant vingt-cinq francs, sans arme, sans violence, sans effraction et en primo-délinquance⁶ était puni – au minimum – d'un an d'emprisonnement ferme sans possibilité de libération anticipée et lorsqu'on sait ce qu'étaient les prisons à l'époque – qu'on pense à ce qu'était, à Reims, la prison de la Bonne Semaine sise sur l'actuelle rue Vauthier-le-Noir –, il était évident que celui qui avait subi cela une fois n'avait aucune envie de voler à nouveau ! Mais l'ennui, c'est que, justement, il recommençait parce que cette peine excessive dans ce régime carcéral sordide, loin de l'avoir amendé, l'avait dégradé un peu plus et, loin de lui avoir enlevé ses défauts, les avait aggravés ou en avait ajouté d'autres. On mit beaucoup de temps sinon à le comprendre ou, tout au moins, à en tirer les conclusions : on commença au milieu du XIX^e siècle à créer les libérations anticipées sous condition de non-récidive – et cela, il faut le souligner, sur la suggestion du Procureur de Roi de Reims, Arnaud Bonneville de Marsangy, le premier président de notre Académie après le Cardinal Gousset, son fondateur – et on instaura à la fin du siècle le sursis à l'exécution de la peine, deux mesures excellentes et bénéfiques qui sauvèrent assurément des délinquants – ce qui, encore une fois, n'a pas de prix – mais qui n'empêchèrent pas la délinquance de poursuivre son chemin. En même temps qu'on instituait ces mesures de mansuétude, on en instituait d'autres en sens

6 Notons qu'à cette époque, en l'absence de casier judiciaire, la récidive était difficile à établir au-delà de la circonscription du tribunal, les *sommiers judiciaires* tenus au ministère de la Justice s'avérant d'un usage malaisé et peu performant.

inverse pour punir plus sévèrement la récidive et notamment la terrible relégation. En toute logique, il était impossible de mieux faire : indulgence pour la primo-délinquance et sévérité contre la récidive. Et tout continua comme avant et on ne fit plus rien de notable jusqu'à la Seconde Guerre mondiale.

Les temps qui suivirent la Libération virent apparaître des méthodes nouvelles, dites sous régime progressif, pour amender les délinquants et on procéda de manière rationnelle en expérimentant sur de jeunes condamnés de dix-huit à vingt-cinq ans, et pour ce faire on créa un centre de détention spécial, une prison-école, très différente des prisons classiques, à Oermingen (Bas-Rhin). C'est là qu'en octobre 1952, j'ai fait mon stage d'études pénitentiaires et je puis en témoigner : si jamais expérimentation fut faite dans les conditions psychologiques idéales, ce fut là, en ce moment ; il régnait dans cet établissement un enthousiasme, une foi, un état d'esprit extraordinaires, non certes chez les surveillants réduits à la garde des portes, mais chez les éducateurs, les enseignants, les moniteurs ; vraiment, oui... pour la première fois on avait abordé et traité le problème comme il le fallait. Comment n'y avait-on pas pensé plus tôt ? La cause de la délinquance étant enfin dûment identifiée en une carence éducative liée au milieu, il s'agissait en toute logique de combler ces carences, de compenser tout ce qui avait été mal vécu, de reconstruire le délinquant, de le rendre autre, de le transformer si bien que celui qui sortait n'était plus le même que celui qui était entré. Et cela qu'était-ce ? sinon une préfiguration du transhumanisme même si on n'avait pas inventé le mot. Entre la sortie de la prison-école et la liberté pure et simple, il y avait une phase intermédiaire de semi-liberté à Maxéville (Meurthe-et-Moselle) et là, nous avons été surpris par le pessimisme de l'éducateur responsable du centre, en parfait contraste avec l'enthousiasme qui régnait à Oermingen ; il nous avait presque choqué cet homme, mais hélas ! c'était lui qui voyait clair parce qu'il avait sous les yeux les détenus sortant de la prison

et commençant à retrouver la liberté et, de fait, la récidive fut la même qu'ailleurs ! La prison-école fabriquait d'excellents détenus, elle n'était pas plus capable qu'une autre de fabriquer des citoyens libres ! Je ne sais ce qu'il en est à l'étranger mais, en France, il est impensable pour une institution de reconnaître ses erreurs et on ne trouva rien de mieux à faire que d'étendre le système à toute la France, où il fut souvent appliqué dans des conditions beaucoup moins bonnes qu'à Oermingen et la récidive et la multi récidive prospérèrent comme auparavant. Ce ne fut qu'en 1980 qu'une circulaire du garde des Sceaux, ministre de la Justice, annonça officiellement l'abandon du régime progressif qui avait – précisait la circulaire – rendu de grands services mais qui arrivait à l'obsolescence. De grands services... Bien sûr, il y avait eu des réussites individuelles mais globalement, sociologiquement, le système avait échoué et on s'orienta vers une conception diamétralement opposée, un système à la carte, un système d'individualisation. En effet, le régime progressif percevait la délinquance comme reposant sur des causes sociétales, essentiellement des carences éducatives, et le but était de compenser ces carences en travaillant sur des groupes de détenus ; or d'une part si les causes sociétales ou exogènes existent – il serait stupide de le nier – les causes personnelles ou endogènes existent aussi et c'est toute l'opposition entre l'héritage des deux penseurs italiens du XIX^e siècle : Cesare Lombroso avec l'anthropologie criminelle et Enrico Ferri avec la sociologie criminelle et, évidemment les causes individuelles ou endogènes ne s'accommodent pas nécessairement d'un traitement uniforme appliqué au sein d'un groupe hétérogène.

La nouvelle politique consiste donc à appliquer à chacun un traitement différencié, individualisé. Cette nouvelle approche s'exprime d'abord par le champ quasiment illimité que j'ai évoqué ci-dessus des réponses pénales possibles et ensuite par le suivi individuel de l'application de la sanction par des services spécialisés, les Services pénitentiaires d'insertion et

de probation (les SPIP) sous le contrôle de nouvelles juridictions : les juges de l'application des peines (JAP)⁷, les tribunaux de l'application des peines (TAP) et, au niveau des cours d'appel, les chambres de l'application des peines (CAP). Bref, le système progressif, hyper rationnel a été remplacé par un système hyper pragmatique et à la théorie a succédé enfin la réalité. Et quelque vingt ans plus tard, en ces années 2020, le résultat est là, désolant : la délinquance, la récidive et la multi récidive fleurissent plus que jamais et atteignent des niveaux inédits. Bien sûr, là encore et comme toujours, il y a des réussites individuelles et on ne s'en réjouira jamais assez mais globalement, c'est l'échec. Sans aller plus loin, il suffit de lire chaque semaine dans le journal champenois *L'Union*, édition rémoise, le compte rendu des audiences du Tribunal correctionnel de Reims : pas une semaine, pas une audience, où ne compare un délinquant dont le casier judiciaire ne mentionne douze, quinze condamnations voire davantage alors que le condamné était libre puisqu'il avait commis de nouveaux méfaits.

Il faut d'ailleurs savoir que ces casiers judiciaires où les condamnations sont énumérées par ordre chronologique de leur prononcé ne présentent pas une liste mais un inextricable maillage car, pour des raisons tenant à la logistique procédurale et judiciaire, l'ordre des condamnations n'est pas nécessairement l'ordre des méfaits si bien qu'une condamnation prononcée après une condamnation avec sursis ne révoque pas le sursis si les faits qui motivent la seconde condamnation sont antérieurs aux faits qui avaient motivé la première ; en outre, la révocation du sursis n'est plus automatique ; il peut y avoir des dispenses de révocation ou des révocations partielles tandis que des sursis probatoires dont le mécanisme de révocation est différent de celui des

7 Les juges de l'application des peines, expérimentés en Alsace, à Oermingen et à Haguenau dès 1950, avaient été, eux, étendus à toute la France dès 1959.

sursis simples sont intervenus dans l'intervalle. Bien entendu, souvent, le délinquant finit par ne plus savoir où il en est : il est délinquant, c'est son état, de temps à autre il va en prison – il y est habitué – de temps à autre il sort de prison – il en est bien content et commet un autre méfait dont il ne sait pas très bien s'il le fera retourner en prison ; il vit ainsi, c'est son mode de vie, c'est tout. De temps à autre, il est convoqué par un agent de probation du SPIP et, en général, il se rend à la convocation peut-être pas toujours aussi fréquente ou aussi prompte qu'il serait souhaitable. En effet, l'effectif budgétaire des éducateurs et agents de probation a été calculé au plus juste et chacun a à charge un nombre de condamnés élevé voire trop élevé ; en outre, l'effectif budgétaire n'est pas toujours au complet : il n'est rien de plus normal qu'un agent de probation obtienne sa mutation pour une autre ville plus en accord avec ses convenances personnelles ou familiales, mais son poste peut très bien rester vacant plusieurs mois avant que son successeur ne soit nommé ; qu'un autre bénéficiaire pendant ce même temps d'un congé parental peut fort bien arriver et, finalement, ceux qui sont disponibles se répartissent la tâche au mieux mais chacun doit suivre un nombre de condamnés trop élevé pour pouvoir accorder à chacun tout le temps qui serait nécessaire. De temps à autre, le condamné comparait devant le JAP. ou devant le TAP, sans très bien savoir pourquoi ni très bien comprendre pourquoi on l'écroue à ce moment plutôt qu'à un autre et il tâche de bien se comporter en prison de manière que, dès que ce sera possible, le même JAP le rende à la liberté, au besoin avec un bracelet électronique ou une obligation d'aller pointer à la gendarmerie ou à la police, ce qui ne l'empêche pas de commettre un nouveau délit.

Que risque-t-il ? Une seizième condamnation et un retour en prison ? Il y est habitué et ce n'est pas cela qui va lui abattre le moral. C'est triste, c'est même affreux mais nous ne sommes pas plus avancés que saint Augustin qui avait au moins le mérite de reconnaître qu'il se trompait tous les jours (ce qui,

au demeurant, ne signifiait pas qu'il se trompait toujours ni dans toutes les affaires, mais enfin qu'une journée ne se passait pas sans qu'il ne fit quelque erreur, ce qui, à la fin de l'année, faisait déjà un joli total !). En notre XXI^e siècle, ce genre d'aveu est plus rare. Et les statistiques délictuelles et criminelles continuent de progresser.

Alors, quand on entend parler de transhumanisme fondé sur des biotechnologies, des nanotechnologies, des neurosciences, on se prend à rêver. Et si cela marchait pour le domaine de la délinquance ? Certes, ce serait une bénédiction mais qu'est-ce qui marcherait ? Et que serait-il possible d'appliquer ? Juridiquement, la formule ne serait pas difficile à trouver car elle existe déjà ne serait-ce, pour ne prendre que cet exemple, que dans le cadre du sursis probatoire ou, pour s'exprimer de meilleure façon, du sursis avec mise à l'épreuve. L'épreuve imposée par le juge peut être très diverse : ne parlons pas d'épreuves d'ordre administratif (répondre aux convocations, prévenir des changements de résidence, etc.), mais d'épreuves adaptées à la personne et aux circonstances et qui constituent le fond même du traitement pénologique. Elles sont nombreuses⁸ et parmi elles le juge détermine, pour chaque cas particulier, celle qui lui paraît ou celles qui lui paraissent le mieux adaptées et voici celle qui m'intéresse : « se soumettre à des mesures d'examen médical, de traitements ou de soins même sous le régime de l'hospitalisation ».

Ainsi donc on peut d'ores et déjà, en toute légalité, imposer à un condamné une mesure qui doit agir non sur ses droits comme le fait l'emprisonnement qui abolit son droit d'aller et de venir, comme le fait l'assignation à résidence qui limite son droit de se déplacer, comme le fait la suspension du permis de conduire qui abolit son droit d'user d'un véhicule sur la voie publique, comme le fait l'incapacité électorale qui éteint son droit de vote, comme le fait l'amende qui affecte son

8 Article 132-45 du *Code pénal*.

patrimoine, comme le fait l'interdiction de fréquenter les débits de boissons qui lui interdit d'entrer chez un cafetier, comme le fait le travail d'intérêt général qui l'astreint à des tâches et on pourrait multiplier les exemples. Mais aucune de ces interdictions ou obligations n'est pour lui invasive – si on ose dire – en ce qu'elle n'agit pas sur son corps, ni sur son anatomie ni sur sa physiologie. Alors que l'obligation de suivre un traitement médical – ambulatoire ou avec hospitalisation – l'oblige à accepter l'introduction dans son corps, dans son être, de substances médicamenteuses ou de radiations propres à modifier son organisme ou le fonctionnement de son organisme, son psychisme ou le fonctionnement de son psychisme.

L'exemple le plus simple est la cure de désintoxication alcoolique ; je ne sais pas comment elles sont faites de nos jours ni avec quelles méthodes, mais je sais que les premières méthodes, celles qui étaient en usage dans les années mil neuf cent cinquante, étaient abominables puisque leur principe était d'établir un lien mental entre l'alcool et le mal-être épouvantable provoqué par la cure, de manière que l'idée même d'alcool éveillât de tels souvenirs qu'elle en devenait insupportable et c'était la méthode de la cure de dégoût ; cela fonctionnait certes, pendant un certain temps et puis, peu à peu, le souvenir de l'abominable mal-être s'estompait, l'appétence pour l'alcool revenait et il fallait recommencer. Non seulement la méthode était épouvantable, mais en outre elle était dangereuse et un médecin m'avait dit que l'alcoolique ne survivait pas toujours à la troisième ; et – croyez-moi – c'était dans la Bretagne profonde, et l'alcoolisme, on connaissait ! Mais, à ce moment-là, le sursis avec mise à l'épreuve et les autres mesures d'effets analogues n'existaient pas et la justice ne pouvait donc ordonner aucun traitement médical de quelque nature qu'il fût et les cures de désintoxication relevaient de l'efficacité persuasive du médecin de famille. Cependant, si le sursis avec mise à l'épreuve avait existé, la justice aurait-elle pu ordonner une cure de désintoxication, aussi éprouvante

dans son vécu, voire aussi dangereuse ? Aurait-elle pu ordonner une mesure susceptible d'exposer le condamné à un risque vital ? De nos jours certes, les méthodes ont changé, mais enfin l'ingestion de médicaments n'est jamais anodine et exige, en déontologie médicale, le consentement du malade. Le premier *Code de déontologie médicale*, le code élaboré par le Professeur Portes, prévoyait que le médecin pouvait imposer le traitement, ce qui n'existe plus avec le code actuel qui est au contraire celui de la contractualisation fondée sur le consentement éclairé du patient. Alors, la justice pourrait-elle imposer dans le cadre d'un sursis avec mise à l'épreuve, un traitement médical, au besoin sous le régime d'hospitalisation impliquant un traitement sévère et invasif ? Eh bien, non, de nos jours, la justice ne le peut pas. Certes, elle peut ordonner une épreuve qui peut être, évidemment, une cure de désintoxication – ou toute autre thérapie – mais si le condamné refuse de s'y soumettre, alors, le sursis est révoqué et l'incarcération s'ensuit pour le temps qui a été fixé si bien qu'en définitive, le condamné peut préférer la privation d'aller et de venir à l'épreuve qui lui a été imposée, surtout si cette épreuve consiste en une action thérapeutique car la privation d'aller et de venir s'exerce de l'extérieur au moyen des murs qui entourent le condamné, alors que l'action thérapeutique s'exerce à l'intérieur, au moyen de substances ou de radiations introduites dans son organisme et en définitive le choix appartient toujours au condamné.

Alors, *quid* du transhumanisme dont on nous parle maintenant sans que, comme nous l'avons vu, ses contours soient encore bien déterminés ? Cela dépendrait de ce qu'il aurait à proposer dans chaque cas et, pour l'heure, j'ai l'impression qu'on ne le sait guère. Certes, l'utilisation de biotechnologies a déjà des réalisations à son actif par exemple pour établir, au profit de personnes physiquement handicapées, des prothèses qui permettent de reconstituer des fonctions physiologiques de mouvement, de déplacement, etc.,

et cela est certainement très positif mais n'a rien à voir avec notre problème. Pour ce qui est à venir, il est question de beaucoup de choses floues et incertaines. Lorsqu'on parle – je cite – de « repousser les limites de l'humain physiquement et moralement », ou bien – je cite encore – « de développer le niveau d'intelligence et de créativité », de quoi, concrètement est-il question ? Certes, parler d'immortalité ou plutôt, selon d'autres, « d'amortalité » (je ne l'invente pas), de mort choisie et non subie paraît relever du fantasme et, en tous cas, n'a rien à voir avec la politique pénale ; de même parler – je cite encore – « d'humanité virtuelle sans enveloppe physique », laisse rêveur. En revanche, développer l'empathie et l'altruisme au profit de la société, voilà exactement ce qu'il nous faut, mais on est bien obligé de reconnaître que cela n'a rien de nouveau et c'était l'objet même de la morale qu'on enseignait dans mon enfance au cours de morale, à l'école primaire préparant au Certificat d'études et au catéchisme préparant à la communion solennelle ; c'était excellent mais n'a jamais empêché qu'il y eût des voleurs ou des pères qui abusaient de leur fille.

La question est donc de savoir comment le transhumanisme s'y prendrait pour réussir là où précisément cette empathie et cet altruisme sont absents et sont remplacés par l'agressivité, la délinquance, voire la criminalité. On parle d'implantation de puces électroniques qui modifieraient les circuits neuronaux ; si cela marchait et donnait de bons résultats – ce qu'on ignore pour l'instant –, on pourrait s'en réjouir. Mais, juridiquement, on rejoindrait le type d'épreuve ci-dessus évoqué : le traitement et la cure médicale, si ce n'est qu'au lieu de l'administration de substances chimiques médicamenteuses, il s'agirait d'intervention physique, en bref de chirurgie au lieu de médecine. Si cela devait exister un jour, le problème resterait le même : on ordonnerait cette intervention à titre d'épreuve dans le cadre d'un sursis probatoire ou d'une contrainte pénale, voire de toute autre procédure judiciaire à inventer, mais le mécanisme juridique fondamental serait le même et finalement la décision du juge consisterait à proposer

plus qu'à imposer puisque l'intéressé pourrait toujours refuser l'épreuve et préférer l'incarcération. Quant à inventer une procédure nouvelle qui imposerait, qui contraindrait, il n'y faut, à mon avis, pas penser car – indépendamment des problèmes pratiques très concrets qui se poseraient et qui ne seraient pas minces – imposer une intervention invasive touchant le corps pour modifier l'esprit, ne serait-ce pas disposer de l'être humain comme d'une chose : *objet de droit* et non plus le traiter comme une personne : *sujet de droit* ? Ce serait alors un tel bouleversement dans la conception même du droit que cela impliquerait, supposerait, un véritable changement de civilisation ; je ne dis pas que cela n'arrivera pas – je ne suis pas prophète – mais enfin, pour l'heure nous sommes toujours dans notre civilisation occidentale, telle qu'elle est, résultant de la confluence et de la fusion de l'humanisme gréco-romain et de la spiritualité judéo-chrétienne et tant que nous y serons – et, personnellement, je souhaite qu'on y reste ! – ce genre de procédure ne me paraît pas envisageable.

En définitive, il me paraît vain de dissenter sur les applications judiciaires et pénales du transhumanisme tant que ses modalités concrètes – ou certaines modalités concrètes – n'auront pas été parfaitement définies. Combien de temps sera nécessaire ? Je n'en sais rien et je crains que nul n'en sache rien car le cheminement plus ou moins rapide des possibilités s'inscrit dans l'histoire, dans le court terme, dans le moyen terme, dans le long terme ; comment le prévoir ? De plus, face aux différentes découvertes qui pourront intervenir et aux différentes possibilités concrètes qu'elles offriront, comment le droit réagira-t-il ? Le droit n'accepte pas d'emblée tout ce qui est matériellement possible. Depuis combien de temps la procréation médicalement assistée est-elle matériellement possible ? Cela fait des décennies et cependant ce n'est qu'en l'année 2022 que la loi a accordé la PMA pour toutes alors que celle-ci avait toujours été limitée aux couples mariés infertiles.

Depuis combien d'années la gestation pour autrui est-elle matériellement possible ? Or en cette année 2021, elle vient encore d'être rejetée par le législateur. Depuis combien de temps sait-on pratiquer l'euthanasie ? Or elle est toujours interdite en France. Certes, on m'objectera que ces trois pratiques sont licites dans bon nombre de pays étrangers et qu'il n'a jamais manqué de Français pour se rendre dans les pays – certains très voisins, voire limitrophes – où cela est permis. Certes, mais il s'agit de choses sollicitées par les personnes elles-mêmes, dans le cadre de leur liberté individuelle. Or la problématique serait fort différente avec les techniques transhumanistes appliquées à des fins judiciaires puisqu'il ne s'agirait plus ici d'accorder une possibilité sollicitée par une personne, mais d'imposer une obligation à une personne qui ne la solliciterait nullement. On me dira que c'est exactement ce qu'on fait lorsqu'on prononce une peine d'emprisonnement ferme, mais encore une fois, l'atteinte à la liberté de la personne n'est pas de même nature ; dans l'emprisonnement, il s'agit d'une limitation de la liberté physique réalisée par un dispositif – les murs – extérieur à la personne ; dans le cas d'une intervention médico-chirurgicale de type transhumaniste, il s'agirait d'une modification de l'intériorité de la personne par une action invasive, intérieure à la personne et la question n'est pas du tout la même.

Autrement dit, si l'emprisonnement, même s'il entraîne malheureusement des répercussions psychiques souvent néfastes, affecte l'être humain de l'extérieur, les applications transhumanistes l'affecteraient, elles, de l'intérieur et cela, finalement, me ramène au premier problème évoqué ici : la recherche de la vérité. Pour celle-ci, tout ce qui est possible est permis sauf ce qui affecte l'intériorité de la personne soupçonnée et c'est – ou plutôt ce fut – le problème du célèbre sérum de vérité, célébrité usurpée d'ailleurs car, en fait, l'efficacité de la scopolamine, du penthotal ou d'autres substances, n'a toujours été que fort incertaine ; ces substances – semble-t-il – modifiaient temporairement le psychisme ce

qui ne contraignait certainement pas le patient à dire la vérité mais, simplement, lui ôtait la lucidité et l'agilité d'esprit nécessaires pour mentir sans se contredire et donc, rendant le mensonge impraticable, elles étaient censées faire nécessairement apparaître la vérité.

En fait, la fiabilité était plus que douteuse car, en admettant même que la substance administrée rendît impossible une politique de mensonge pensée, organisée, structurée, il semble qu'elle pouvait aussi gêner la perception des frontières entre le réel et l'imaginaire car, entre le mensonge et la vérité, il y a aussi place pour l'erreur, pour l'illusion. Cela étant, si par hypothèse, on découvrait une substance efficace et fiable, je n'imagine pas qu'on pût, juridiquement, imposer à la personne interrogée, une injection intraveineuse qui modifierait temporairement son psychisme, l'empêcherait radicalement de mentir et donc abolirait temporairement sa liberté intérieure ; le *Code de procédure pénale* actuel ne prévoit certes pas cet acte procédural ni pour la police judiciaire ni pour le juge d'instruction ni pour le juge des libertés et de la détention ni pour les juges du fond ; une disposition législative serait nécessaire et – à supposer que le Parlement la votât ce qui est loin d'être acquis à l'avance – je doute fort que le Conseil constitutionnel la déclarerait conforme à la Constitution ou, plus exactement au « bloc de constitutionnalité » et que la Cour européenne des droits de l'homme, siégeant à Strasbourg, la déclarerait conforme à la Convention européenne des droits de l'homme, tout au moins en l'état actuel des textes et des principes. Mais je doute fort qu'à moins d'un bouleversement révolutionnaire et d'un renversement total de civilisation, les choses changent sur ce point. Et ce qui est vrai en ce domaine de recherche de la vérité me paraît l'être, pareillement, *mutatis mutandis*, dans le domaine de la sanction pénale et je doute fort que le Conseil constitutionnel d'une part, la Cour européenne des droits de l'homme d'autre part, donnent leur agrément à une disposition législative qui permettrait d'imposer et de réaliser, contre la volonté du

condamné, une intervention d'inspiration transhumaniste qui, finalement, ferait de lui un autre, meilleur peut-être, mais enfin : un autre.

Terminons ce trop long exposé en évoquant le problème de la responsabilité.

AUTRES PERSPECTIVES FUTURISTES ET RESPONSABILITÉ

J'ai fondé les considérations qui précèdent sur ce qui existe aujourd'hui concrètement en matière d'intelligence artificielle et virtuellement en ce qui concerne le transhumanisme ; mais ces perspectives dépendent, pour leur avènement et leur développement, des instruments disponibles en matière d'informatique et il est probable sinon certain que le jour où ces instruments deviendront plus performants ou différemment performants, ces considérations – tout au moins certaines d'entre elles – seront périmées. Or, en ce domaine des machines, les perspectives actuelles sont déconcertantes. Mais qui dit technique, qui dit instrument ne peut s'empêcher de penser que les techniques et les instruments peuvent un jour défaillir et jusqu'ici, toute défaillance matérielle ou humaine a toujours suscité des problèmes de responsabilité ; qu'en sera-t-il ?

Les perspectives futuristes

À ce jour, – mais les choses peuvent changer très vite – deux problèmes me paraissent posés : l'un touchant les logiciels et les algorithmes et c'est le problème du web3 et l'autre touchant les machines et c'est le problème de l'ordinateur quantique.

Le problème du web3

On parle aujourd'hui du web3 capable de changer complètement la face de la télématique et cela à brève échéance. Je suis bien incapable d'expliquer ce dont il s'agit mais l'enjeu doit être important si on en juge par les

controverses qui animent actuellement la Silicon Valley et qui portent encore sur la définition même de ce concept encore entouré de bien des incertitudes. On parle d'un ensemble de technologies et de nouveaux services permettant de contrôler son « soi virtuel », son « identité digitale » et ses « biens digitaux » ce qui, je l'avoue, ne m'éclaire pas beaucoup ; dans ce nouvel et encore incertain domaine, la notion qui semble jouer un rôle majeur est celle de *blockchain* dont la seule définition construite que j'ai trouvée est celle-ci : « Base de données distribuée qui gère une liste d'enregistrements protégés contre la falsification ou la modification par les nœuds de stockage ; c'est donc un registre distribué et sécurisé de toutes les transactions effectuées depuis le démarrage du système réparti ». Soit... Certains voient dans le web3 le risque d'une utilisation pour une nouvelle concentration des pouvoirs tandis que d'autres souhaitent l'avènement d'une forme de stabilité juridique (*sic*) encore à découvrir. Pour l'heure la querelle fait rage ; selon Elon Musk, le PDG de Tesla : « Cela a tout l'air d'être une sacrée... (ici un terme à traduire en bon français par un vocable qui ne saurait trouver sa place dans un texte publié – s'il doit l'être – par l'Académie des Sciences, Arts et Belles-Lettres de Caen) » ; en fait, web3 ne serait qu'un « terme de marketing en vogue plus qu'une réalité ». Il n'empêche qu'à cet aimable scepticisme fait écho, d'ores et déjà, la mobilisation par des firmes d'importance majeure de capitaux considérables pour financer la réalisation et la mise sur le marché du web3. Qu'en sera-t-il ? Je ne sais mais si ce concept doit un jour plus ou moins proche entrer dans le domaine de la réalité, la question pourra sans doute se poser de son introduction dans la justice pénale avec des conséquences que je suis bien incapable d'imaginer.

Le problème de l'informatique quantique

Je n'évoquerai que brièvement une circonstance dont la survenance est désormais considérée comme certaine :

l'avènement de l'informatique quantique. Si, pour toutes les applications et surtout les plus complexes, la performance des logiciels et des algorithmes est de première importance, la performance du matériel, c'est-à-dire la puissance des ordinateurs, ne l'est pas moins. Il est inutile de développer des considérations sur les progrès qui ont été faits et le sont encore en fonction de la loi de Moore selon laquelle la capacité des microprocesseurs double tous les dix-huit mois ; cette loi n'a pas encore été démentie même s'il est certain qu'elle atteindra un jour une limite liée à la longueur d'onde de la lumière, à quoi certains envisagent pour remède l'intervention de l'ultra-violet⁹. Mais peut-être n'est-ce déjà plus la question puisque plus personne, désormais, ne semble douter de l'arrivée de l'informatique quantique fondée sur des bases physiques différentes et dont l'avènement est prévisible pour les années 2030 à 2035. Qu'on ne me demande pas d'en expliquer le principe, encore moins le mécanisme, car la physique quantique n'entre pas dans mon domaine de compétence.

Ce que j'ai cru comprendre, en revanche, c'est que cette nouvelle informatique, fondée sur des bases entièrement différentes, sera, à l'informatique actuelle, à peu près ce qu'est Internet au télégraphe de Claude Chappe et les conséquences prévisibles sont telles que certains parlent déjà d'une nouvelle révolution industrielle. L'intelligence artificielle et ses applications s'en trouveront alors multipliées par un coefficient actuellement inimaginable puisque, entre l'informatique actuelle et l'informatique quantique, la différence sera de nature et non plus de degré et des calculs qui, actuellement, requièrent un certain temps pour les ordinateurs les plus perfectionnés trouveront leur solution presque instantanément ; bref, ce sera un autre univers. Alors, si on imagine que, dans de telles machines, on introduit des logiciels et des algorithmes aussi performants que ceux que

9 C'est ce qui m'a été expliqué par une personne qui, du fait de sa profession, connaît bien ces problèmes.

promet Métavers, on en arrive presque à se demander s'il y aurait une limite aux possibilités de ces systèmes. En fait, il y en aurait une parce que la limite est dans la nature des choses, mais elle serait si éloignée de ce qu'on connaît actuellement qu'on ne voit pas pourquoi les applications en matière de justice pénale – et pour la recherche de la vérité et pour la détermination de la sanction – n'en bouleverseraient pas d'une manière quantitativement et qualitativement inimaginable l'activité judiciaire.

Les problèmes de responsabilité

Et justement, revenons à la justice pénale. Imaginons – par hypothèse – la machine dotée d'intelligence artificielle capable d'identifier un coupable là où tous les enquêteurs et les juges d'instruction auraient échoué, une justice pénale où il n'y aurait plus de *cold cases* et une justice pénale capable d'apporter, pour chaque délinquant ou à chaque criminel, la solution propre à le faire sortir de son enfer personnel et propre à mettre la société à l'abri de nouvelles exactions de sa part. Mais supposons qu'un jour, la machine quantique dotée d'intelligence artificielle forte vienne à errer et défaille, ce qui me paraît une hypothèse envisageable puisque, jusqu'à présent tout au moins, il n'a jamais rien existé sur terre qui ne puisse errer et défaillir ; quelle serait la conséquence ?

J'imagine que, d'abord, avant tout et au moins à titre préventif, il existerait une instance composée d'êtres humains chargée d'agrèer les systèmes et les méthodes et de contrôler leur application ; de nos jours, on connaît déjà la Commission nationale informatique et liberté (la CNIL) à vocation très générale et bien des organismes à vocation plus particulièrement ciblée telle l'Autorité de contrôle des fichiers de police judiciaire et de rapprochement judiciaire. Il paraît donc infiniment probable qu'une instance serait créée pour agrèer les systèmes et en contrôler les applications ; quelle serait sa composition ? Quels seraient ses méthodes et ses moyens d'action ? Je n'en ai aucune idée.

Mais revenons à la machine, ou plutôt à ces machines, et supposons d'abord qu'elles aient été acquises par le ministère de la Justice et fassent partie des biens d'équipement des juridictions ; il n'y aurait aucun problème particulier, l'État serait responsable du dommage causé par son robot judiciaire, comme il est responsable de tout dommage causé soit par ses préposés – fussent-ils magistrats – soit par tout bien d'équipement utilisé dans ses services ; la victime de l'erreur judiciaire pourrait alors se retourner – comme de nos jours – contre l'État pour solliciter et obtenir réparation du préjudice causé, par exemple par le temps passé injustement en détention.

Mais supposons maintenant que l'utilisation de l'intelligence artificielle ait fonctionné en mode de recours à un tiers, en mode d'expertise, hypothèse que j'ai envisagée ci-dessus ; là encore il n'y aurait pas de problème insoluble puisque l'expert serait responsable du matériel qu'il aurait utilisé, dans les termes du droit commun.

Mais imaginons maintenant que cette machine ait reçu la personnalité juridique, hypothèse qui, à juste titre, paraît saugrenue – que dis-je ? – stupide... démente... et qui, pourtant n'est plus impensable comme nous allons le voir. En effet, le Parlement européen a chargé récemment une commission d'étudier la possibilité qu'un robot doté d'une intelligence artificielle puisse être considéré comme une personne juridique pour le cas de dommage causé par ledit robot, qui pourrait alors être condamné à réparation. Le Parlement européen a envisagé l'hypothèse de conférer une personnalité électronique (*sic*) à tout robot prenant des décisions autonomes au même titre qu'une personne morale et physique. En tant que juriste, je ne sais pas du tout ce que pourrait être une personnalité électronique, mais je sais ce qu'est une personne juridique : c'est un sujet de droit, c'est-à-dire une entité titulaire de droits qui agit par le droit sur une autre entité qui, elle, subit cette action et qui est, elle, objet de droit ; jusqu'à ce jour, le sujet de droit a toujours été une

personne, soit une personne physique, c'est-à-dire un être humain, soit une personne morale, c'est-à-dire un ensemble juridiquement organisé de personnes physiques – et donc d'êtres humains –, tandis que l'objet de droit a toujours été une chose : un immeuble, un meuble, un animal, une machine, une marchandise. Que la personnalité juridique pût un jour être conférée à une machine, à un robot, serait un bouleversement civilisationnel inimaginable puisque cela impliquerait qu'un bien matériel, un objet, fût – sur certains plans tout au moins – l'égal d'un être humain, cet être humain qui, encore une fois, est le seul être à posséder la personnalité juridique puisque, jusqu'à ce jour tout au moins, les efforts des antispécistes pour faire reconnaître la personnalité juridique de l'animal et même (pour certains mouvements) la personnalité juridique du végétal, n'ont pas abouti. Ainsi, reconnaître la personnalité juridique d'une machine, d'un robot, serait un ébranlement des bases les plus fondamentales du droit. C'est pourquoi, apprenant cette décision du Parlement européen de créer cette commission d'étude, je me disais que ledit Parlement ne pensait certainement pas à l'application de l'intelligence artificielle à l'activité judiciaire mais, bien plus probablement, à l'automobile autonome ou à l'automobile à conduite déléguée qui, elle, est bien en cours d'élaboration. Mais il n'empêche que, si la question posée à la Commission recevait un jour une réponse positive, il faudrait en tirer les conséquences en matière judiciaire comme ailleurs. Ainsi, une machine dotée de la personnalité juridique qui aurait, d'un ensemble de documents judiciaires, tiré des conclusions qui, par la suite, se seraient avérées inexactes, pourrait être déclarée responsable d'une erreur judiciaire et condamnée à réparer le dommage tout comme, aujourd'hui, l'État peut l'être du même chef en cas d'erreur commise par des juges. Au demeurant, je ne voyais pas du tout la nécessité, pour obtenir cette réparation légitime, de doter la machine de la personnalité juridique puisque, de toutes manières, cette machine serait bien la propriété de quelqu'un – fût-ce celle de l'État – et que, comme

je viens de le dire, ce propriétaire serait responsable des erreurs dans les termes du droit commun comme tout propriétaire d'un équipement est responsable des dommages causés par celui-ci et dont il a – selon la terminologie juridique – l'usage, le contrôle et la direction.

J'en étais là de ces réflexions lorsque le ciel m'est tombé sur la tête à l'écho du fait suivant : en 2020, Stephen Thaler, ingénieur, citoyen des États-Unis d'Amérique, avait inventé un robot qui lui-même, en toute autonomie, avait inventé un type entièrement nouveau de récipient propre à conserver la chaleur et Stephen Thaler tenait à ce que le brevet, pour ce récipient, fût inscrit non pas à son nom patronymique mais au nom du robot, à quoi les autorités compétentes avaient opposé un refus, lui objectant qu'un brevet ne pouvait être inscrit qu'au nom d'un être humain, qu'au nom d'une personne, seule sujet de droit reconnue comme telle par le droit. Mécontent, Stephan Thaler avait tenté les mêmes démarches auprès des autorités d'autres nations – européennes notamment – et s'était vu opposer le même refus jusqu'au jour où il s'était adressé aux autorités d'Afrique du Sud qui, elles, n'avaient vu aucun obstacle à inscrire le brevet au bénéfice du robot, lequel robot se trouve donc maintenant, comme tout inventeur, titulaire d'un brevet, ce qui signifie que, pour la première fois au monde, une machine est titulaire de droits, n'est plus objet de droit mais est devenue sujet de droit. Et si le récipient isotherme rencontre un grand succès, les redevances ne pourront être versées qu'à la machine, seule titulaire du brevet et donc seule créancière des redevances, ce qui impliquera, pour cette machine, la constitution d'un patrimoine. Mais imaginons qu'un jour, le récipient inventé par cette machine s'avère défectueux, non pas par un vice de fabrication mais par un vice de conception, la seule responsable du dommage ne pourrait être que la machine, titulaire du brevet, qui paierait les dommages-intérêts sur le patrimoine qu'elle se serait constitué avec les redevances versées à son nom ; ainsi, titulaire de

droits, créancière de redevances, propriétaire d'un patrimoine, responsable des dommages, cette machine, ce robot aurait bien la personnalité juridique et c'est très probablement cette circonstance qui a ému le Parlement européen et l'a conduit à confier l'étude du problème à une commission ! Car il n'est pas impossible et il est même probable que d'autres inventeurs, à l'instar de Stephen Thaler, s'adresseront à leur tour aux autorités d'Afrique du Sud et solliciteront l'inscription d'autres brevets au bénéfice d'autres machines et, alors, l'Afrique du Sud ne restera sans doute pas longtemps seule, car d'autres États l'imiteront et finalement le Parlement européen sera bien obligé de conclure à la nécessité de modifier ou de remplacer la Convention de Munich pour permettre d'accorder des brevets non à des personnes mais à des machines, c'est-à-dire à des choses qui, jusqu'alors et par définition n'auraient été que des objets de droit et qui deviendraient ainsi des sujets de droit. Ce faisant d'ailleurs, ils ne feraient que rejoindre les animaux pour lesquels les antispécistes réclament à cor et à cri la personnalité juridique, que la loi du 30 novembre 2021¹⁰ ne leur a certes pas accordée mais vers laquelle elle a tout de même fait un grand pas.

Je me trompe peut-être mais il me semble que le jour où il n'y aura plus d'objets de droit mais uniquement des sujets de droit nous aurons changé de monde, changé d'univers et, surtout, changé de civilisation.

COMMENT CONCLURE ?

Au terme de ce travail comme de tout autre, il faut conclure mais le malheur est que j'en suis bien incapable pour la simple raison qu'il s'agit de ce qui se passera demain, ce que je ne puis ni deviner ni prévoir. Au demeurant, on dit que « L'avenir ne se prévoit pas, il se construit », mais la construction se fera

10 Loi du 30 novembre 2021 *visant à lutter contre la maltraitance animale et à conforter le lien entre les animaux et les hommes.*

sans moi. Je constate simplement qu'à ce jour ni l'intelligence artificielle ni le transhumanisme n'ont encore rien apporté à la justice. Apporteront-ils quelque chose ? et, dans l'affirmative, qu'apporteront-ils ? Je ne sais. En ces quarante-deux années où se déroula ma carrière, que de choses ont changé ! Que de réformes de statut, de structure, d'organisation, de lois de fond et de lois de procédure ont été promulguées ! Le vieux *Code d'instruction criminelle* de 1808 a été remplacé en 1959 par le nouveau *Code de procédure pénale*, et le vieux *Code pénal* de 1810 a été évincé par le moderne *Code pénal* entré en vigueur le 1^{er} mars 1994, et pourtant... la justice pénale reste affrontée aux mêmes problèmes.

Plus de choses encore changeront sans doute dans les quarante années qui vont venir : « Une justice pénale nouvelle sortira-t-elle du fond du désert, brillante de clartés, et portant sur le front une marque immortelle ? » ¹¹, de tout cœur, je l'espère et le souhaite ; ou bien les années à venir vérifieront-elles l'aphorisme quelque peu cynique et désabusé du Guépard : « Il faut que les choses changent pour que tout demeure pareil » ? Je ne le souhaite assurément pas, mais peut-on l'exclure ?

11 Qu'on veuille bien me pardonner cette parodie approximative de RACINE : *Athalie* (Acte III – scène 7).

CONTRIBUTIONS DE L'IMAGERIE CÉRÉBRALE A UNE NOUVELLE SCIENCE DE LA MÉMOIRE

par M. Francis EUSTACHE

de l'Académie des Sciences Arts et Belles-Lettres de Caen

DE L'HISTOIRE AUX NEUROSCIENCES

Quand on arrive à Caen et plus largement dans différentes villes et sites de Normandie, que l'on connaisse avec plus ou moins de précision l'Histoire, on entre sur des terres de mémoire, avec une impression particulière, différente de celle que l'on ressent dans d'autres lieux. Le château ducal de Guillaume le Conquérant, les deux grandes abbayes et les nombreuses églises, les quartiers datant du Moyen-âge en sont des exemples visibles, que l'on se plaît à découvrir ou à redécouvrir. Mais surtout, on ressent la présence de cette ville reconstruite après les bombardements alliés de 1944. Ces bâtiments, qui datent de quelques décennies, nous placent dans un entre-deux, entre un passé qui remonte à nos parents et à nos grands-parents, le présent et, bien sûr, nous font réfléchir au futur de cette ville. Logiquement, on peut comprendre que ces lieux passionnent les historiens et les chercheurs de disciplines connexes, mais aussi les citoyens du monde entier. Le Mémorial de Caen, musée consacré à l'histoire du ^{xx}e siècle et à la paix, en est un témoin érudit, incontournable. À quelques kilomètres tout autour, des échappées dans la campagne normande et sur les plages du Débarquement donnent à réfléchir plus encore, si l'on complète le recueillement dans l'un des grands sites de la bataille de Normandie, comme le cimetière militaire américain de Colleville-sur-mer, au-dessus de la plage d'Omaha Beach, avec ses 70 hectares et ses 9 386 tombes.

Je suis né en Normandie, dans le Cotentin, plus précisément. Je ne suis pas devenu historien, mais spécialiste de la mémoire humaine. Ces chemins de la mémoire, que j'ai parcourus avec ma collaboratrice Béatrice Desgranges et avec de nombreux autres collègues, nous ont d'abord conduits vers la neuropsychologie et les neurosciences cognitives¹. Il s'agissait de comprendre, d'analyser la mémoire et au-delà les fonctions cognitives chez des personnes présentant une affection cérébrale ou psychiatrique. Mon expérience clinique et les travaux de recherche prenant leur source dans les années 1970-1980, les neurosciences y contribuent aussi largement, car cette période correspond à un renouveau très important de ces disciplines. Les recherches menées en imagerie cérébrale en constituent l'exemple le plus démonstratif. À Caen, le développement de l'imagerie cérébrale est superposable à la création et à l'évolution de la plateforme Cyceron. Cyceron est un groupement d'intérêt public (GIP), fondé dans les années 1980 pour accueillir les premières grandes techniques d'imagerie comme la Tomographie par émission de positons (TEP) avec des utilisations dans différents domaines et notamment les neurosciences (fig 1). L'Imagerie par résonance magnétique (IRM) est ensuite venue compléter la TEP pour comprendre les bases cérébrales de la mémoire.

1 EUSTACHE et DESGRANGES, 2010, 2020.



Figure 1 – Première caméra à positons utilisée au Centre Cyceron, dès le milieu des années 1980, la TTV03, de fabrication française (*Photo GIP Cyceron*).

Nous évoquerons ici différents travaux qui ont permis de différencier les structures cérébrales impliquées dans les grands systèmes de mémoire (comme la mémoire épisodique ou mémoire des souvenirs), ou encore dans les processus de mémoire (encodage *vs* récupération) chez le sujet sain. Les travaux d'imagerie cérébrale ont aussi beaucoup contribué à décrire les bases cérébrales des troubles de la mémoire, la maladie d'Alzheimer ayant donné lieu au plus grand nombre d'études. Dans un premier temps nous développons les travaux réalisés dans le domaine de la recherche clinique sur les maladies neurodégénératives. Ensuite, nous évoquons des travaux actuels en imagerie cérébrale, aux confins des sciences humaines et sociales, qui révèlent les substrats cérébraux de la mémoire collective, et d'autres qui soulignent les mécanismes de protection contre un trouble de stress post-traumatique, ce qui montre à la fois l'évolution des techniques d'imagerie et au-delà de nos disciplines.

ATROPHIE CÉRÉBRALE DANS LA MALADIE D'ALZHEIMER

L'IRM anatomique a permis de décrire avec une grande précision le profil d'atrophie dans la maladie d'Alzheimer. Les travaux ont d'emblée souligné la perte de substance grise dans de nombreuses régions corticales et ont permis par le développement de nouvelles procédures de traitement d'images. Ainsi, la méthode VBM (pour *Voxel Based Morphometry*) a été utilisée pour la première fois sur groupe important de patients souffrant de maladie d'Alzheimer dans notre laboratoire². Il a ainsi été montré qu'à un stade démentiel léger à modéré, l'atrophie touche non seulement la région hippocampique, mais aussi le cortex temporo-pariétal, le cortex cingulaire postérieur et le précuneus (Fig. 2). De même, chez des patients qui souffrent de troubles isolés de la mémoire épisodique et présentent un risque accru de développer la maladie d'Alzheimer (appelés MCI pour *Mild Cognitive Impairment*), les études en IRM ont mis en évidence une atrophie du lobe temporal interne, ainsi que d'autres régions cérébrales, même si la perte de substance grise y est moins marquée que chez les patients répondant aux critères diagnostiques de maladie d'Alzheimer. Bénéficiant de diverses avancées techniques et notamment de logiciels permettant de segmenter la région de l'hippocampe, essentielle pour la mémoire et point de départ de la maladie d'Alzheimer, des travaux étudiant les sous-parties histologiques de l'hippocampe suggèrent que le champ CA1 (qui fait partie de la corne d'Ammon, l'une des trois sous-structures de l'hippocampe, avec le subiculum et le gyrus denté et qui correspond à sa partie antérieure) serait le plus précocement et spécifiquement altéré dans la maladie d'Alzheimer par rapport au vieillissement normal (LAJOIE *et al.*, 2010).

2 BARON *et al.*, 2001.

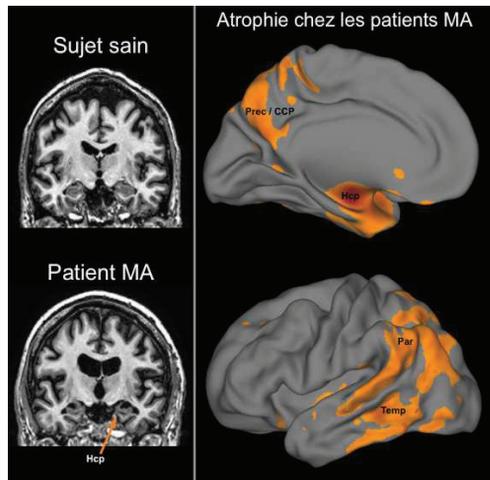


Figure 2 : Représentation de l'atrophie de la substance grise dans la maladie d'Alzheimer (MA, analyse VBM de données IRM ; UMR_S 1077).

Les modifications morphologiques de la substance grise sont donc très importantes au niveau du lobe temporal interne. Avec l'avancée de la maladie, l'ensemble du cortex présente des altérations structurales et l'atrophie s'accroît dans les régions premièrement atteintes. Ces altérations semblent mimer la progression des dégénérescences neurofibrillaires (enchevêtrements de fibrilles dans le corps cellulaire des neurones dont le principal constituant est la protéine Tau), suggérant que l'IRM anatomique est assez sensible pour surveiller *in vivo* la progression de la pathologie.

HYPOMÉTABOLISME DANS LA MALADIE D'ALZHEIMER

La mesure du métabolisme cérébral du glucose en TEP chez des patients au repos (ne réalisant aucune activité particulière) permet de quantifier les modifications fonctionnelles qui reflètent principalement les altérations synaptiques. Cette mesure, utilisant un radio-isotope à demi-vie longue, reflète une consommation cérébrale moyenne de glucose sur une durée d'une heure, chez des personnes ne recevant aucune

consigne particulière, autre que de rester éveillées. Dans la maladie d'Alzheimer, l'hypométabolisme du cortex temporo-pariétal et du gyrus cingulaire postérieur constitue l'anomalie la plus précoce et la plus fréquente (fig. 3). Cette baisse du métabolisme s'étend ensuite au cortex frontal, alors que le métabolisme du cortex primaire moteur et sensoriel, des noyaux gris centraux et du cervelet est relativement préservé. Ce profil métabolique est en accord avec l'altération des fonctions cognitives et la préservation des fonctions sensorimotrices observées chez la majorité des patients souffrant de maladie d'Alzheimer. Les patients à des stades plus précoces présentent aussi un hypométabolisme au niveau du gyrus cingulaire postérieur, y compris à l'échelle individuelle. De plus, son atteinte est plus importante chez les patients qui développent ensuite une maladie d'Alzheimer que chez les patients dont les troubles restent stables, mais c'est le dysfonctionnement du cortex temporo-pariétal qui différencie le mieux ces deux populations de patients.

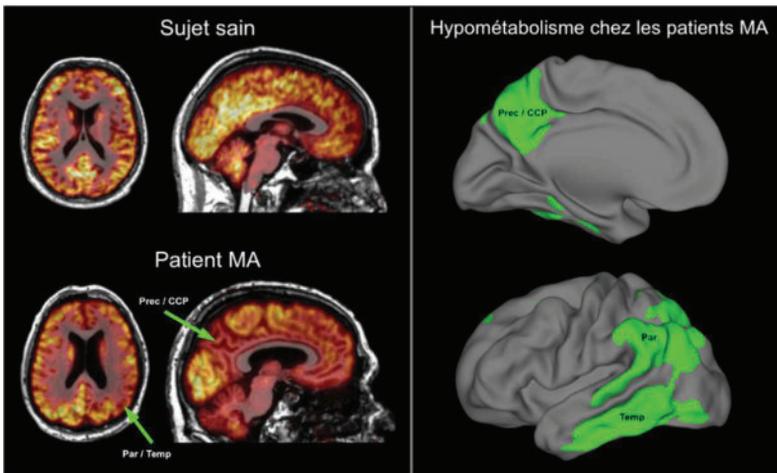


Figure 3 : Hypométabolisme cérébral dans la maladie d'Alzheimer (MA, analyse SPM de données TEP-FDG ; *UMR_S 1077*).

IMAGERIE AMYLOÏDE DANS LA MALADIE D'ALZHEIMER

Les années 2000 ont vu l'apparition de nouveaux marqueurs TEP permettant de visualiser *in vivo* les dépôts fibrillaires de protéine β -amyloïde, une des anomalies neuropathologiques caractéristiques de la maladie d'Alzheimer. Le PiB (pour *Pittsburg compound B*), marqué au carbone 11 (radio-isotope de demi-vie d'environ 20 minutes) a été l'un des premiers radio-traceurs utilisés. Le développement plus récent de composés marqués au fluor 18 (radio-isotope de demi-vie plus longue, de l'ordre de 110 minutes) a permis l'utilisation de cette technologie en recherche clinique. Notre laboratoire et la plateforme Cyceron ont largement contribué au développement de ces technologies en France et à leurs applications en recherche clinique³

Dès les premières publications, ces marqueurs ont montré une augmentation globale de la quantité de dépôts amyloïdes cérébraux chez les patients souffrant de maladie d'Alzheimer par rapport aux sujets âgés contrôles. Les données d'imagerie amyloïde sont concordantes avec celles de la neuropathologie, montrant des dépôts diffus au sein des cortex associatifs, en particulier les aires frontales (fig. 4) ; elles le sont moins avec le profil clinique des patients.

Chez les patients MCI, les dépôts amyloïdes semblent toucher les mêmes régions mais de manière moins importante, tout du moins au niveau du groupe. En réalité, les données se présentent sous la forme d'une distribution bimodale avec des patients MCI présentant un marquage élevé, quasi-identique aux patients MA (on parle de profil amyloïde-positif) tandis que pour les autres, le marquage faible est dit amyloïde-négatif. Enfin, certains sujets âgés sans déficit cognitif présentent un profil de type amyloïde-positif, témoignant de la présence de dépôts amyloïdes.

3 CAMUS *et al.*, 2012.

Utilisé conjointement avec d'autres mesures, ce marqueur pourrait favoriser le diagnostic précoce et faciliter le diagnostic différentiel puisque, à l'exception des démences à corps de Lewy, aucun dépôt amyloïde n'est généralement observé dans les autres types de démence. Cependant à l'heure actuelle, cette technique n'est pas utilisée en dehors du cadre de la recherche et, si elle l'était, il serait essentiel de prendre garde aux aspects éthiques compte tenu de la possibilité d'observer un profil amyloïde-positif chez un sujet sans aucun symptôme.

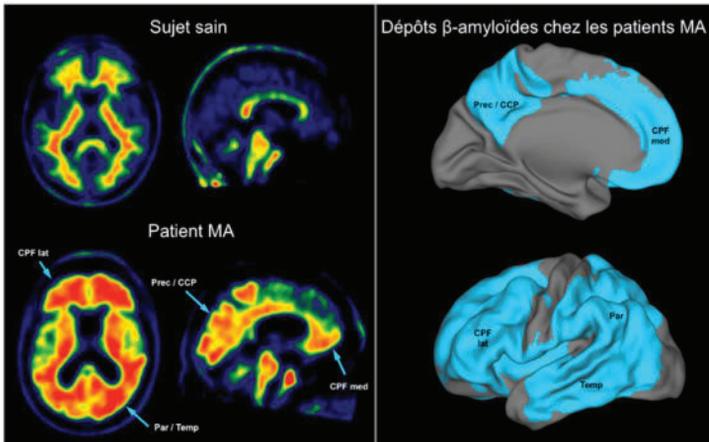


Figure 4 – Dépôts de protéine – amyloïde au niveau de la substance grise dans la maladie d'Alzheimer (MA, analyse SPM de données TEP-Florbétapir ; *UMR_S 1077*).

DISCORDANCE ENTRE ATROPHIE ET HYPOMÉTABOLISME DANS LA MALADIE D'ALZHEIMER

Au début de ces travaux, dans les années 1990, nous nous attendions à établir des cartes d'atrophie et d'hypométabolisme, en partie superposables, avec des zones hypométaboliques plus étendues que les régions atrophiées, mécanisme déjà décrit dans les pathologies par lésions focales, comme certaines lésions vasculaires. Les résultats obtenus ont

révélé une situation beaucoup plus complexe : alors que l'atteinte structurale de l'hippocampe est précoce et marquée, son atteinte fonctionnelle n'est pas constamment observée. Au contraire, le cortex cingulaire postérieur, peu atrophié, et qui est une région pourtant fort éloignée de la formation hippocampique, est hypométabolique. Cette discordance a suscité de nombreux débats et les interprétations ne sont pas closes : comment expliquer que le cortex cingulaire postérieur soit si touché sur le plan fonctionnel alors qu'il ne fait pas partie des régions les plus atrophiées ? Et comment expliquer que l'atrophie précoce et massive de l'hippocampe ne se traduise pas par une diminution importante de son métabolisme ?

Pour répondre à la première de ces questions, une hypothèse a été posée dès les années 1990, celle d'un effet à distance (ou diaschisis ; mécanisme décrit de longue date et remis sur le devant de la scène grâce aux études en TEP) : l'atrophie de l'hippocampe exercerait un effet délétère sur le fonctionnement du cortex cingulaire postérieur. Nous avons apporté des arguments forts en faveur de cette hypothèse en montrant des liens entre les anomalies de chacune de ces deux structures et le cingulum, faisceau de substance blanche qui les connecte. Par la suite, nous avons montré, chez des patients aux stades précoces, que l'atrophie hippocampique initiale est reliée au degré d'évolution de l'atrophie du cingulum, au cours du suivi de 18 mois, et que l'atrophie initiale du cingulum est elle-même reliée au degré d'évolution du métabolisme au sein du cortex cingulaire postérieur⁴. Ces résultats sont en faveur de relations de cause à effet entre ces différentes altérations, l'atrophie hippocampique se situant au début de cette séquence temporelle. Ils permettent de comprendre la progression de lésions et leurs conséquences sur le fonctionnement de régions cérébrales parfois éloignées, notamment des symptômes comportementaux qui dépassent les troubles de la mémoire.

4 Villain *et al.*, 2010.

Quant à la relative préservation du fonctionnement de la région hippocampique, elle pourrait résulter d'une compensation fonctionnelle liée à la plasticité élevée de cette région. Si la nature exacte de ce phénomène reste à comprendre, sa découverte a engendré non seulement de nouvelles recherches, mais aussi l'espoir que cette plasticité cérébrale soit mise à profit pour élaborer des thérapeutiques à même de lutter contre le processus dégénératif.

Au-delà de leur intérêt fondamental, ces résultats sur la disconnexion hippocampo-cingulaire ont permis de caractériser un phénomène qui pourrait s'avérer un marqueur de la maladie d'Alzheimer. En effet, tant dans le vieillissement normal que dans diverses pathologies cérébrales, les anomalies morphologiques et fonctionnelles se superposent assez bien. Ainsi, la démence sémantique entraîne des altérations dans le cortex temporal externe (fig. 5)⁵. De plus, même si d'autres pathologies cérébrales peuvent entraîner une atrophie hippocampique comparable à celle observée dans la maladie d'Alzheimer, cette atrophie n'est pas associée à un hypométabolisme du cortex cingulaire postérieur.

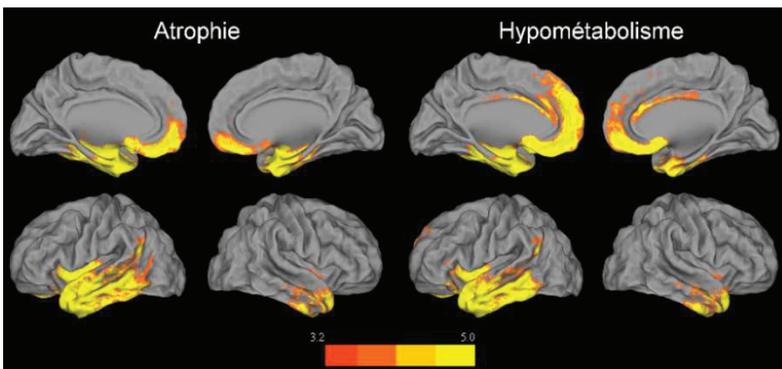


Figure 5 – Atrophie et hypométabolisme dans la démence sémantique (inspiré de Bejanin *et al.*, 2017 ; UMR_S 1077)

5 DESGRANGES *et al.*, 2007 ; BEJANIN *et al.*, 2017.

UNE MULTIPLICITÉ D'APPROCHES POUR ÉTUDIER LES LIENS ENTRE CERVEAU ET COGNITION

Une première approche, que nous avons largement développée au sein du laboratoire avec la TEP, consiste à établir des liens – sous forme de corrélations – entre des altérations cérébrales et des troubles neuropsychologiques. Elle permet ainsi d'identifier les régions cérébrales dont le dysfonctionnement est responsable de troubles cognitifs spécifiques. De même, avec l'IRM anatomique, des liens entre l'atrophie de l'hippocampe et des déficits de mémoire épisodique ont ainsi été mis en évidence chez des patients atteints de maladie d'Alzheimer. Chez des patients MCI, l'analyse automatique voxel par voxel, dans l'ensemble du cerveau, a également montré des corrélations significatives entre la densité de substance grise de l'hippocampe et des scores de mémoire épisodique.

L'approche corrélatrice a surtout été utilisée avec des mesures métaboliques à l'aide de la TEP. Ces travaux pionniers ont montré que le dysfonctionnement de différentes régions cérébrales était responsable des troubles de différents systèmes de mémoire⁶. Les troubles précoces de la mémoire épisodique sont sous-tendus par le dysfonctionnement de la région hippocampique, tandis que les troubles de la mémoire sémantique sont liés au dysfonctionnement du cortex temporal gauche. Dans la maladie d'Alzheimer à un stade modéré, cette approche a aussi permis de souligner l'existence de mécanismes compensatoires mis en jeu dans une tâche de mémoire épisodique. Ces mécanismes seraient sous-tendus par des régions néocorticales temporales, normalement dévolues à la mémoire sémantique.

La deuxième approche est celle des études d'activation en TEP ou en IRMf. La plupart s'intéressent à la mémoire

6 DESGRANGES *et al.*, 1998.

épisodique et convergent vers une diminution des activations hippocampiques chez les patients atteints de maladie d'Alzheimer par rapport aux témoins, tant lors de l'encodage que de la récupération d'informations. Chez des patients à risque de développer la maladie d'Alzheimer, l'activité de la région hippocampique semble au contraire augmentée, par rapport à des sujets témoins. Cette hyperactivation hippocampique pourrait jouer un rôle compensatoire et être transitoire, ne se manifestant qu'à un stade précoce de la maladie.

Un autre résultat émerge des études d'activation dans la maladie d'Alzheimer : une augmentation des activations chez les patients par rapport aux sujets sains âgés, principalement dans le cortex frontal. Ce résultat a d'emblée été considéré comme un support des mécanismes compensatoires. Cette idée était plausible d'une part parce que le cortex frontal est atteint plus tardivement que le cortex associatif postérieur dans cette maladie et d'autre part parce que cette région est impliquée dans la mise en œuvre des ressources attentionnelles et des fonctions exécutives, qui seraient davantage sollicitées chez les patients que les sujets âgés sains pour effectuer la même tâche. Quelques études ont conforté cette hypothèse en montrant d'une part une corrélation négative entre le volume hippocampique et l'activité corticale au sein du lobe frontal (moins l'hippocampe est gros et plus l'activité frontale est forte) et d'autre part des corrélations positives entre les activations frontales et les performances mnésiques (plus les activations sont élevées, meilleures sont les performances).

Ainsi, l'imagerie fonctionnelle permet de montrer qu'en dépit d'atteintes structurales majeures, les patients recrutent des réseaux supplémentaires qui, à défaut de compenser les performances déficientes, témoignent d'une plasticité cérébrale longtemps ignorée dans les pathologies neurodégénératives. Concernant la mémoire épisodique, au stade précoce de la maladie, la région hippocampique semble être le siège de modifications fonctionnelles compensatoires,

tandis qu'à un stade plus avancé, les mécanismes compensatoires seraient pris en charge par des régions néocorticales, principalement frontales, parfois temporales.

DE LA MÉMOIRE À L'HISTOIRE ET AUX NEUROSCIENCES SOCIALES

Même si les travaux consacrés à l'imagerie multimodale des maladies neurocognitives ont été les plus nombreux depuis les débuts du centre Cyceron, des études ont porté sur d'autres pathologies de la mémoire comme le syndrome de Korsakoff, qui provoque une amnésie sévère le plus souvent consécutive à un abus d'alcool et à une dénutrition. D'autres encore portent sur le mystérieux ictus amnésique, syndrome transitoire dont on comprend encore très mal la physiopathologie. D'autres enfin portent sur des pathologies non-cérébrales, comme le cancer du sein. Pourtant, le retentissement qu'il provoque chez les femmes qui en souffrent, entraîne des modifications du fonctionnement cognitif et de la mémoire et l'imagerie cérébrale nous aide aussi à comprendre les mécanismes qui permettent à ces femmes de faire face à la maladie, avec plus ou moins de force et d'efficacité. Par ailleurs, les études d'activation, menées chez le sujet sain ou dans diverses pathologies, n'ont jamais cessé. Dans les premières études, utilisant la TEP, nous avons étudié les mécanismes d'encodage et de récupération en mémoire épisodique (fig. 6)⁷. Progressivement, en utilisant l'IRM fonctionnelle, nous avons approché des dimensions de la mémoire plus complexes

⁷ BERNARD *et al.*, 2001.

comme la mémoire autobiographique et la projection dans le futur⁸.

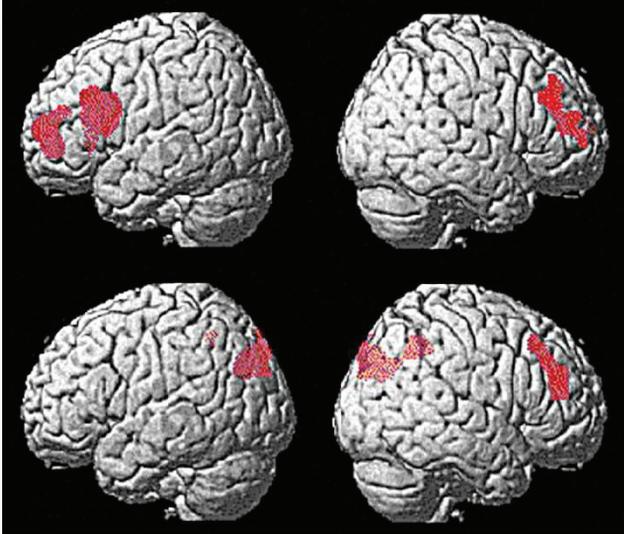


Figure 6 : Activations mises en évidence en TEP dans un groupe de sujets jeunes : l'encodage en mémoire épisodique active de façon préférentielle le cortex préfrontal gauche tandis que la récupération active de façon préférentielle le cortex préfrontal droit (inspiré de Bernard *et al.*, 2001 ; UMR_S 1077).

Plusieurs publications récentes tirent leur originalité de liens étroits avec les sciences humaines, permettant ainsi d'approcher certains aspects du fonctionnement de la mémoire qui n'auraient pas pu être envisagés il y a quelques années. Une étude porte sur la manière dont la mémoire collective façonne la construction des souvenirs personnels, y compris leurs représentations cérébrales⁹. Pour les sociologues, nos souvenirs sont en effet modelés par la mémoire collective de notre communauté ; le fonctionnement de la mémoire des individus ne peut pas être compris sans prendre en compte leur appartenance à un groupe. Ces théories issues des sciences

⁸ DEGEILH *et al.*, 2015 ; VIARD *et al.*, 2007.

⁹ GAGNEPAIN *et al.*, 2020.

sociales ont été testées dans une étude en imagerie cérébrale portant sur la Seconde Guerre mondiale qui s'est déroulée en plusieurs phases. Une première analyse a porté sur le contenu de 30 ans de reportages sur cette guerre, diffusés entre 1980 et 2010 à la télévision française. À l'aide d'un algorithme, nous avons identifié des groupes de mots utilisés pour évoquer de grandes thématiques, comme le Débarquement allié en Normandie. Par ailleurs, des participants à l'étude ont visité le Mémorial de Caen et ont observé des photos de cette période, accompagnées de légendes. À partir de ces légendes, nous avons défini le degré de relation entre les photos et les différentes thématiques de la mémoire collective identifiées dans les reportages sur la guerre diffusés dans les médias. L'étape suivante a permis d'évaluer si ce degré de proximité influençait les souvenirs individuels de la visite du Mémorial. À cet effet, les volontaires ont bénéficié d'un examen en IRM, pendant qu'ils se remémoraient les images vues la veille au Mémorial. Ces différentes analyses ont permis de croiser les représentations collectives de la guerre dans les médias et les activations en imagerie cérébrale. Les résultats ont souligné l'implication du cortex préfrontal médian, zone du cerveau associée à la cognition sociale. Ce travail met en valeur une approche transdisciplinaire de la mémoire qui permet d'approcher les confins des représentations individuelles et collectives.

Une autre étude récente s'inscrit dans un programme de recherche plus large, intitulé 13-Novembre, qui porte sur la construction des mémoires individuelles et collectives des attentats qui sont survenus à Paris en novembre 2015¹⁰. Cette étude s'intéresse aux réseaux cérébraux impliqués dans le Trouble de stress post-traumatique (TSPT). Dans un groupe de près de 200 participants, elle explore les effets d'un événement traumatique sur les structures et le fonctionnement du cerveau. Le but est d'identifier des marqueurs neurobiologiques du

¹⁰ MARY *et al.*, 2020.

stress post-traumatique mais aussi de la résilience au trauma : pourquoi certaines personnes ayant vécu un traumatisme souffrent-elles de TSPT alors que d'autres ne développent pas ce trouble ? L'étude réalisée en IRMf, chez des personnes ayant développé un TSPT et chez des personnes résilientes, ayant toutes vécu le même traumatisme, montre que la résurgence intempestive des images et des pensées intrusives, serait liée à un dysfonctionnement des réseaux cérébraux impliqués dans le contrôle de la mémoire, c'est-à-dire une défaillance des mécanismes qui permettent de supprimer et de réguler l'activité des régions de la mémoire lors d'une intrusion.

Au total, les quelques résultats résumés dans cet article montrent que l'imagerie cérébrale morphologique et fonctionnelle a profondément modifié la recherche et la pratique clinique en neuropsychologie. Elle a permis de mieux comprendre la nature et l'origine des déficits mnésiques des patients. Nous avons développé l'exemple de la maladie d'Alzheimer, maladie parfois présentée comme une « accélération » du vieillissement normal. Au contraire, nombreux sont les points de rupture entre vieillissement normal et maladie d'Alzheimer. D'autres travaux récents abordent des thématiques qui n'étaient pas envisageables il y a quelques années, comme l'importance des représentations collectives dans nos souvenirs personnels ou encore les mécanismes qui nous protègent contre les souvenirs traumatiques. Les interactions entre travaux fondamentaux et cliniques en imagerie cérébrale sont déterminantes pour le progrès des connaissances avec des applications cliniques très attendues.

Tout au long de cette communication et dans l'entretien croisé avec Jean-Michel DERLON, nous avons souligné la place singulière qu'occupe Cicéron dans la recherche en neurosciences. Depuis sa création dans les années 1980, cet établissement n'a cessé de se moderniser, de s'agrandir, d'acquérir de nouveaux équipements, le plus récent étant une

IRM de nouvelle génération (fig 7). Ces équipements à la pointe de la technique, la présence sur site de nombreux jeunes chercheurs et le soutien des organismes de recherches et des collectivités territoriales assurent à Cyceron sa pérennité et en font une force pour la recherche en sciences cognitives et cliniques.



Figure 7 : Imageur IRM 3T de marque *General Electric* utilisé à ce jour au Centre Cyceron (*Photo GIP Cyceron*).

Remerciements : Cette conférence est dédiée à ma collègue, Béatrice DESGRANGES, récemment disparue, qui a contribué à nombre de travaux présentés tout au long de ces lignes. L'auteur remercie également les membres des unités de recherches du Centre Cyceron et tous les autres collaborateurs dont les études sont ici citées.

BIBLIOGRAPHIE

- BARON, J.-C., CHÉTELAT, G., DESGRANGES, B. *et al.*, « In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease », *Neuroimage*, 2001, 14 p. 298-309.
- BEJANIN A., CHÉTELAT G., LAISNEY M. *et al.*, « Distinct neural substrates of affective and cognitive theory of mind impairment in semantic dementia », *Soc Neurosci*, 2017, 12 p. 287-302.
- BERNARD, F., DESGRANGES, B., PLATEL, H. *et al.* « Contributions of frontal and medial temporal regions to verbal episodic memory : a PET study », *Neuroreport*, 2001 13 p. 1737-41.
- CAMUS, V., PAYOUX, P., BARRÉ, L. *et al.*, « Using PET with 18F-AV-45 (florbetapir) to quantify brain amyloid load in a clinical environment », *Eur J Nucl Med Mol Imaging*. 2012, 39 p. 621-31.
- DÉGEILH, F., GUILLERY-GIRARD, B., DAYAN, J. *et al.*, « Neural correlates of self and its interaction with memory in healthy adolescents », *Child Dev*, 2015, 86, 6 p. 1966-83.
- DESGRANGES, B., BARON, J.-C., DE LA SAYETTE, V. *et al.*, « The neural substrates of memory systems impairment in Alzheimer's disease. A PET study of resting brain glucose utilization », *Brain*, 1998, 121 p. 611-31.
- DESGRANGES, B., MATUSZEWSKI, V., PIOLINO, P. *et al.*, « Anatomical and functional alterations in semantic dementia : a voxel-based MRI and PET study », *Neurobiol Aging*, 2007, 28 p. 1904-13.
- EUSTACHE, F., DESGRANGES, B., *Les chemins de la mémoire*, Paris, Le pommier, 2010.
- EUSTACHE, F., DESGRANGES B., *Les nouveaux chemins de la mémoire*, Paris, Le pommier, 2020.
- GAGNEPAIN, P., VALLÉE, T., HEIDEN, S. *et al.*, « Collective memory shapes the organization of individual memories in the medial prefrontal cortex », *Nat Hum Behav*, 2020, 4 p. 189-200.

LAJOIE, R., FOUQUET, M., MÉZENGE, F. *et al.*, « Differential effect of age on hippocampal subfields assessed using a new high-resolution 3T MR sequence », *Neuroimage*, 2010, 53 p. 506-14.

MARY, A., DAYAN, J., LEONE, G. *et al.*, « Resilience after trauma : The role of memory suppression », *Science*, 2020, 367, 6479.

VIARD, A., PIOLINO, P., DESGRANGES, B. *et al.*, « Hippocampal activation for autobiographical memories over the entire lifetime in healthy aged subjects : an fMRI study », *Cereb Cortex*, 2007, 17 p. 2453-67.

VILLAIN, N., FOUQUET, M., BARON, J.-C. *et al.*, « Sequential relationships between grey matter and white matter atrophy and brain metabolic abnormalities in early Alzheimer's disease », *Brain*, 2010, 133, 11 p. 3301-14

Table des matières

Hommage à Henry Delisle, président de l'Académie (2022)..... 5

Joël Bruneau, maire de Caen, Caen ville d'inventeurs et de savants..... 9

Intelligence artificielle, biotechnologies neurosciences : vers quel humanisme ? Présentation des travaux (Edgar LEBLANC)..... 19

Printemps

Alain BRETTO, Le calcul et le raisonnement humain..... 29

Jean RICHERT, Une brève histoire d'intelligence, de conscience et d'éthique..... 47

Michel CHEIN, L'intelligence artificielle va-t-elle contribuer à transformer *homo sapiens* en *homo numericus* ?..... 77

Épanouissement

Jean-Claude MARCHAL, Transhumanisme, humanisme, philosophie des Lumières, définitions et relations..... 109

Patrick PENEL, Une IA conquérante et un rêve non accompli..... 131

Jean-Michel DERLON, Francis EUSTACHE, CYCERON entre neurosciences et intelligence 163

Laurent BELLAMY, Didier LAFORGE, Edgar LEBLANC, Iliès Zaoui, Conscience Robotics le chemin d'une réussite..... 181

Frédéric THIBAULT-STARZYK, Intelligence artificielle à l'Université d'Oxford : l'intégration des disciplines..... 185

Ombres et lumières

- Marieke WOLF-FÉDIDA, La communication de l'intelligence artificielle comparée à l'intelligence humaine..... 197
- † Jean GUGLIELMI, Apprendre et penser resteront-ils l'apanage du vivant ?..... 231
- Jacques BOUINEAU, Intelligence artificielle et humanisme naturel..... 245
- Thierry BRETON, La politique de la Commission et de l'Union européenne en matière d'intelligence artificielle et de souveraineté numérique..... 283

Questionnements imagination

- Jacques BOUCHARLAT, Folie artificielle un brin de folie.... 293
- Richard FORESTIER, L'art au croisement des technologies et de l'humanité..... 305
- Bernard DEVAUCHELLE, Synesthésie..... 327

Sagesse

- Claude KIRCHNER, Éthique de la conversion numérique....339
- Éric SURAUD, La révolution numérique annonce-t-elle une renaissance de l'homme ?..... 351
- Jean TOUZET, Intelligence artificielle transhumanisme et justice pénale..... 379
- Francis EUSTACHE, Contribution de l'imagerie cérébrale à une nouvelle science de la mémoire..... 431

Académie des Sciences, Arts et Belles-Lettres de Caen

Mémoires de l'Académie Tome LVII

Secrétariat de rédaction : Chantal Adigard, Edgar Leblanc

Mise en page : Jade Pacary

©septembre 2022

Achévé d'imprimer sur les presses de
l'Imprimerie Moderne de Bayeux - F14400 Bayeux

Dépôt légal : 72257 - Octobre 2022

Imprimé en France