IA ou ceux qui agitent les fantasmes ?

Benoît Le Blanc

Faut-il craindre les IA ou ceux qui agitent les fantasmes ?

Should we fear AI or those who play with fantasies?

Benoît Le Blanc[1]

[1] École Nationale Supérieure de Cognitique - IMS UMR-5218 CNRS–UB–Bordeaux-INP - Bordeaux (France)

RÉSUMÉ. La question de la crainte ou de la confiance à l'égard de l'intelligence artificielle revient régulièrement sur le devant de l'actualité. La nouveauté provient de ce que des applications concrètes de l'IA se font jour. D'ailleurs il ne se passe plus un jour sans que les journaux évoquent « les IAs ». Mais faut-il véritablement en avoir peur ? Evoquer cette question passe par trois points : comprendre ce qu'est l'IA, examiner l'actualité du domaine, et se doter de quelques moyens pour réfléchir à comment avancer socialement et politiquement avec l'IA.

ABSTRACT. The question of fear or confidence in artificial intelligence regularly comes to the forefront of the news. The novelty comes from the fact that concrete applications of AI are emerging. Besides, not a day goes by without the newspapers mentioning "AIs". But should we really be afraid of it? Raising this question involves three points: understanding what AI is, examining current events in the field, and equipping ourselves with some means to think about how to move forward socially and politically with AI.

MOTS-CLÉS. IA, IAs, Intelligence Artificielle, éthique, confiance.

KEYWORDS. AI, AIs, Artificiel Intelligence, ethics, trust.

Faut-il craindre « les IA », ces nouvelles entités, sorte de déclinaisons multiformes et omniprésentes, produits de l'Intelligence Artificielle, « l'IA » ? Cette question se pose et se repose régulièrement, depuis des dizaines d'années. Elle revient systématiquement à chaque avancée technologique plus ou moins bluffante de la programmation en IA. Cette question tend toujours les mêmes ressorts émotionnels, ceux de l'ambiguïté, du malaise et, au final, de la peur. Toute réponse raisonnable est alors inaudible ou très vite oubliée dans le concert des cassandres ou des agitateurs et autres amateurs de l'angoisse. Le progrès technologique est-il corrélé aux décharges d'adrénaline, au déficit de sérotonine, à la sécrétion d'hormone du stress, et à l'irruption sur les plateaux de télévision des marchands de terreur ou des instigateurs de l'anxiété ?

Il n'y a évidemment pas de raison objective de craindre les réalisations basées sur l'IA, comme il n'y a pas de raison de craindre les produits des grandes avancées scientifiques et des technologies ; c'est de leur utilisation et donc des acteurs de leur usage dont ils convient de prendre garde. Faire le procès à l'Intelligence Artificielle est injuste. Il s'agit d'une discipline scientifique comme une autre. Condamne-t-on la physique pour les bombes qui tuent des enfants, la chimie pour les suicides des mélancoliques désespérés, la météorologie pour les tempêtes ou la climatologie pour les sécheresses de ces dernières années ? L'IA permet de produire des programmes, souvent intégrés dans d'autres programmes, destinés à une utilisation donnée. C'est de leur emploi malveillant dont il faut se méfier, autant que du fait de celles et de ceux qui en font déjà un usage inconsidéré, parfois mal compris, souvent à leur seul profit.

Dans cette perspective, cet article propose une étude en trois points : une présentation de quelques principes permettant de comprendre ce que sont l'IA et « les IA » ; une actualité de ce qui se fait avec l'IA ; et la formulation de quelques considérations pour avancer socialement et politiquement avec l'IA.

1. Ce qu'est l'intelligence artificielle

Depuis son origine dans les années 1950, l'IA s'est toujours inscrite dans le processus d'informatisation. Les pionniers du domaine sont aussi ceux qui ont construit l'informatique moderne et le calcul automatique [MCC 56] [MIN 56] [NEW 57]. Pas d'IA sans informatique, certes, mais en retour pas de science des ordinateurs sans une réflexion sur ce que peut être et ce que peut faire l'IA [MIN 67] [MCC 69] [BOS 87].

1.1. *L'informatique et la perte du sens*

L'informatisation passe par la numérisation. Cette opération parfois complexe comporte des étapes de réduction, d'approximation et à la fin, de codage. Avant de devenir des données, les informations du monde que l'on veut coder doivent consentir à une réduction de ce monde. En informatique, aborder la réalité se fait en abandonnant toute une partie de celle-là. On y plaque une matrice arbitrairement choisie, là où l'on pense qu'une information pertinente se situe. Au-delà du périmètre de cette sorte de grille, aucune donnée ne prendra forme, sauf à appliquer une autre grille à côté de la première, et ainsi de suite. Le monde est alors réduit au périmètre des tableaux de numérisation que le chercheur a choisi [CLA 19a]. À l'étape suivante, chaque case de la grille se verra attribuée une sorte d'étiquette uniforme, venant ainsi prendre dans une approximation toutes les nuances que la case pouvait comporter. Ces « étiquettes » ou « désignations » ont un caractère certain d'arbitraire, mais participent à une étape nécessaire pour pouvoir obtenir une donnée utilisable par une machine. Vient enfin le processus de codage ; il détermine les seuls états qui seront pris en considération et donc les seules valeurs que l'on pourra attribuer à chacune des cases de la grille que l'on considère. Tout ce qui est hors de la grille ou trop grand pour entrer dans cette grille, tout ce qui est autre que l'étiquette de chaque case, ou bien tout ce qui est plus fin que la différence entre deux valeurs d'étiquettes, n'existe tout simplement pas dans le monde numérique.

Voilà en quelques mots ce qui constitue la base nécessaire à l'informatisation. Tout cela s'applique évidemment à l'IA. L'information doit rentrer dans un format pour générer une donnée, alors que chez l'humain, l'information doit être plongée dans un contexte sémantique pour devenir une connaissance. Il y a là le début d'une explication de la différence entre information et cognition. On est bien dans deux modes de fonctionnement, deux mondes différents : le monde humain dans lequel informations et contextes constituent des connaissances, et le monde des machines où informations et formats ouvrent à la capacité de stockage et de traitement des données par des programmes de manipulation de bits, réduisant en zéros et en uns toute la complexité du monde. Certains auteurs voient pourtant là une sorte de continuum : données, informations, connaissances, savoir [ERM 12]. C'est avec cette vision que l'on peut développer par exemple le domaine de la gestion des connaissances, en œuvrant pour formaliser les connaissances humaines et capitaliser le savoir des experts [ERM 08] [LEB 13]. D'autres diront que les connaissances sont et restent une exclusivité humaine, en conséquence de quoi les machines ne peuvent y accéder et se contenteront de ne gérer que des données [SEA 80]. Pour les premiers, l'intelligence de l'IA peut se construire en capturant l'intelligence humaine ; pour les seconds, il est illusoire de parler d'intelligence dans le cadre de l'IA.

Bien ancrée dans l'informatique, l'IA a l'ambition de se comparer à l'intelligence humaine, au cerveau animal et aux ensembles de neurones biologiques. Mais sait-on véritablement comment toute cette matière vivante fonctionne [FIN 85] ? On cherche, on propose, on conjecture et, à chaque avancée [RES 20], ce sont de nouvelles idées qui viennent fertiliser l'IA et de nouveaux espoirs pour faire évoluer les artefacts, machines et programmes. Amplifiées par les médias, ces nouvelles pistes de recherche sont autant de sirènes qui retentissent dans le discours toujours renouvelé des collapsologues. Comment s'y retrouver sans poser quelques points de repère ?

Tout d'abord l'intelligence humaine est une chose, l'intelligence artificielle en est une autre. Dans leurs principes, leurs constructions et leurs évaluations, ces deux domaines sont très différents. Bien

sûr l'IA va s'inspirer des découvertes sur le cerveau, le corps ou le comportement ; cela présidait même à la définition initiale de l'IA. Mais l'IA moderne garde les limites inhérentes à l'informatique. Même si l'on imagine ce que pourrait être un esprit pur, et que l'on cherche à y tendre en souhaitant des cerveaux informatiques et qui puissent être transférés, stockés sur disques durs ou circulant dans les réseaux, il faut admettre que les seuls cerveaux connus dans la nature sont humides, incarnés et remarquer que le temps les use vite. Rien de tout cela n'est modélisé en IA… Par ailleurs, il est amusant de remarquer que les avancées en IA poussent les neurologues, les physiologistes et les psychologues à reconsidérer leurs points de vue, sur leurs propres modèles et leurs propres recherches en faisant usage du vocabulaire et des métaphores de l'IA. On parle ainsi de mémoire à court terme et à long terme chez l'humain [ATK 68], alors que cette distinction n'a été posée en informatique que pour des questions de taille des zones de stockage consacrées aux registres du processeur. De même, l'administrateur central [BAD 74] n'est qu'une copie symbolique de l'unité de contrôle de l'architecture Von Neumann [VON 49].

1.2. *Les deux voies de l'IA*

Pour faire simple, au cours de son histoire l'IA s'est structurée et continue à avancer selon deux voies différentes : celle des symboles et celle des configurations. Ces deux concepts traduisent beaucoup plus l'état d'esprit des différents chercheurs en IA, qu'une réelle nécessité physique ou matérielle. Cette forme de dualité existe bien entendu dans d'autres champs disciplinaires, par exemple avec l'histoire et la géographie, la psychologie et la sociologie, l'électronique et l'informatique, etc. À chaque fois, il s'agit de deux points de vue portés sur un même concept : celui du territoire, de l'individu social, du calcul numérique, etc. C'est le concept de prise de décision qui constitue le centre des préoccupations de l'IA. Pourtant, en IA, la différence entre les symboles et les configurations est forte, et elle clive résolument le domaine en deux camps. Historiquement, ces deux voies sont nées en même temps, dès le début de l'IA. Elles ont dès lors, chacune tour à tour, servi des intérêts stratégiques et économiques différents. À chaque étape, les avancées technologiques ont constitué de véritables bonds en avant. On peut citer le Perceptron, le système expert Mycin, le programme de jeu d'échecs Deep Blue® d'IBM, le Knowledge Graph® de Google, le programme Alphago® de DeepMind, le projet Debater d'IBM, l'agent conversationnel ChatGPT® d'OpenAI, etc. Ces avancées sont de plus en plus rapides et combinent configurations et symboles dès lors qu'il s'agit de composer des produits commerciaux, comme pour les identifications faciales sur les photos telles que celles proposées par Facebook, les recommandations de produits telles que celles proposées par Netflix ou Amazon, les recherches d'informations telles que celles proposées par MicroSoft ou Google, etc. La recherche en IA est aujourd'hui dirigée par la technologie dans une rude compétition économique.

La voie de l'IA qui utilise les symboles est le « cognitivisme ». L'intelligence de la machine reposerait alors sur la capacité à manipuler ces symboles, un peu comme le langage nous permet d'assembler des mots et in fine nous permet de faire passer une idée [SIM 83]. Les premières réalisations ont consisté à proposer une modélisation du monde à explorer sous la forme d'un graphe et d'y coller des algorithmes astucieux de parcours. Pour développer de tels algorithmes, des langages de programmation tels que LISP [MCC 60] ou Prolog [COL 83] ont été mis au point. L'idée générale consiste à enchaîner des règles et à manipuler les formes de connaissances que l'on aura comprises, on dit extraites, de l'expertise d'un humain. Ce passage d'une connaissance implicite à des éléments explicites manipulables par une machine, est au principe du cognitivisme. L'intelligence reste humaine et se retrouve simulée sur une machine par le jeu des enchaînements de symboles. Les perspectives sont de pouvoir mettre le maximum de symboles dans une machine pour obtenir une IA performante.

La voie de l'IA qui explore les configurations est appelée le connexionnisme. L'intelligence de la machine émergerait alors de la capacité à connecter des configurations. On parle d'une approche subsymbolique car ce sont les configurations qui vont ici jouer le rôle qu'ont les symboles dans le

cognitivisme précédent [SMO 86]. Sans que ces configurations ne puissent réellement correspondre à quelque chose d'identifiable pour un humain, elles vont servir de base aux calculs de la machine. La reconnaissance de motifs dans des images, des sons, ou tout autre type de signaux, ouvre la voie à des opérations de classification et d'enclenchement d'actions dédiées. Pour ce dispositif de reconnaissance de motifs appris, ce qui est codé dans la machine correspond à ce que l'on pense être du processus de caractérisation de l'intelligence, de la reconnaissance immédiate faite par l'œil humain. Les perspectives sont de pouvoir réduire au minimum la taille de ces motifs pour obtenir une autre forme d'IA performante.

Dans un cas comme dans l'autre, les informaticiens partent de l'hypothèse d'un « monde clos ». Ce jargon informatique explique ainsi au naïf que tout ce qui n'est pas codé dans la machine est inintéressant et sera considéré comme faux ou hors contexte, conduisant à des résultats inconnus. Le glissement sémantique va jusqu'à la conviction que tout ce qui n'est pas numérisable n'est pas digne d'intérêt, voire pour certains n'existe pas. Le dispositif d'IA enchaîne-t-il les règles édictées par un expert automobile ? À l'utilisateur du programme de savoir si cet expert est un motoriste, un carrossier, un agent de la circulation ou un assureur… car tout emploi d'un tel dispositif en dehors de son cas d'usage ne sera pas adapté. De la même manière, un dispositif d'IA reconnaissant les vingt-six lettres de l'alphabet sera en échec face à un symbole tel que celui de l'euro, ou bien produira un résultat totalement imprévisible… Et d'ailleurs s'est-on interrogé sur les conséquences des différences biologiques entre droitiers et gauchers pour écrire les lettres manuscrites venant alimenter l'apprentissage de la machine, ou bien des spécificités culturelles des alphabets étrangers ? L'espagnol utilise le tilde, des accents toniques ou les points d'exclamation ou d'interrogation inversés en début de phrase, le polonais des caractères augmentés, et par exemple le russe, le chinois ou le japonais des caractères totalement différents de ceux des claviers européens. Pourquoi la description du monde devrait-elle être à la fois numérique et spécifique à la langue anglaise ? Tout ceci est problématique et la confrontation à une situation inconnue par la machine dans tel ou tel contexte engendrera une réponse imprévisible et probablement hors contexte, mais choisie dans la panoplie des actions ou de réponses prédéterminées.

1.3. *Des pionniers aux usages nouveaux*

L'IA a suivi une évolution que l'on peut retracer à travers les définitions qui ont été adoptées pour décrire le domaine scientifique. On s'accorde sur une naissance en 1950 avec l'article d'Alan Turing [TUR 50]. Il n'y parle ni d'ordinateur, ni d'intelligence artificielle : les « computers » désignaient à l'époque les personnes qui effectuaient des calculs et le terme « IA » n'a été proposé que deux ans après le décès de Turing [LEB 14]. Ce sont pourtant bien les travaux de Turing qui ont permis de déchiffrer certains codes secrets de communication allemands de la seconde guerre mondiale, et en ce sens qui constituent une première réalisation d'IA.

La définition initiale et explicite de l'IA est associée à un cycle de conférences de l'été 1956 [MOO 06]. Les instigateurs de ces conférences avancent que « chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut en principe être décrit si précisément qu'il est possible de construire une machine pour le simuler » [MCC 06]. On voit bien le lien avec le « jeu de l'imitation » proposé par Turing qui concluait qu'à la fin du vingtième siècle, il ne serait plus possible de distinguer pour un humain si les réponses à n'importe quelle question qu'il formulerait sont celles d'un humain ou d'une machine.

Dans les années 1970, des chercheurs comme Marvin Minsky (*op. cit.*) ou Margaret Boden [BOD 77] recourraient à une désignation plus pragmatique et finalement moins audacieuse : « Le but de l'IA est de construire des machines qui réalisent des choses qui requièrent de l'intelligence lorsqu'elles sont faites par des humains ». En 2017, à l'occasion du programme interministériel *#FranceIA*, et à l'époque en responsabilité du domaine de l'IA au sein de la Direction générale de la recherche et de l'innovation du Ministère de l'enseignement supérieur et de la recherche, nous avons produit une définition plus large et plus interdisciplinaire : « L'intelligence artificielle désigne un

ensemble de notions ou d'algorithmes, s'inspirant de la cognition humaine ou du cerveau biologique, et destiné à assister ou suppléer l'individu dans le traitement des informations massives » [FIA 17]. L'idée était alors de mettre en évidence, d'une part que le champ de l'IA dépasse largement celui de l'informatique, et d'autre part que l'économie vient désormais y jouer un rôle majeur. Ces idées ont été reprises par la Commission d'enrichissement de la langue française publiant au Journal officiel du 9 décembre 2018 cette nouvelle définition (la précédente datait de 1989, révisée en 2000 pour autoriser l'acronyme « IA ») : « Champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines » [CEL 18].

2. Ce que l'on fait avec l'intelligence artificielle

Reconnaître des visages sur une photo, transformer une page manuscrite en un fichier texte, traduire ce texte, déterminer sur une image si un véhicule est un vélo, une voiture ou un camion, suivre le parcours de ce véhicule en mouvement sur une vidéo, lire sa plaque d'immatriculation s'il en a une, déterminer un itinéraire, ou encore proposer une destination à visiter, sont autant d'actions qui composent le quotidien de chacun sans qu'il en prenne réellement conscience. Lorsqu'elles sont automatisées, ces tâches utilisent des résultats de la recherche en IA. Pour chacune d'entre elles, il s'agit de mettre en œuvre une forme de prise de décision par le classement d'un signal reçu dans une catégorie apprise. La performance du système dépend donc directement de la quantité de catégories envisagées à l'avance et de la représentativité de ces catégories pour faire face à l'ensemble des signaux qui seront reçus par le système. Tous ces systèmes fonctionnent en deux temps. Le premier est celui du calibrage, de l'initialisation des valeurs de façon à assurer une bonne reconnaissance d'un ensemble déterminé de formes présentées au système. Le second temps est celui de l'utilisation, du traitement des signaux en situation réelle en s'appuyant sur la classification constituée par le système.

À côté de cette « IA du traitement » des données du monde s'est développé depuis 2016 un nouveau domaine : celui de l'« IA générative » [JOV 22]. Il ne s'agit plus seulement de reconnaître une catégorie apprise et de classifier, mais d'utiliser ces catégories pour générer un « signal vraisemblable ». Cette nouvelle approche revisite complètement la manière de travailler de nombre de spécialistes tels que les publicitaires, les cinéastes, mais aussi les auteurs, les mathématiciens ou les informaticiens… Elle ouvre aussi une nouvelle voie dans les capacités de traduction des textes, de programmation, et de création d'applications utilisant des « agents conversationnels ». Derrière cette IA générative se cachent trois concepts : l'apprentissage profond, les réseaux antagonistes et les modèles de langage.

2.1. *L'apprentissage profond*

La méthode algorithmique d'IA la plus utilisée est aujourd'hui celle des « réseaux de neurones formels » [CEL 18, *ibid.*], plus particulièrement celle des réseaux à propagation en couches, et plus précisément celle des modèles d'apprentissage profond. On le voit bien dans cette phrase, les informaticiens ont la fâcheuse tendance à nommer ce qu'ils font à partir d'un sabir qui leur est propre, et donc de s'approprier le vocabulaire courant en détournant les mots de leur sens usuel, pour leur en donner un autre, spécifique à l'informatique.

Y comprendra qui pourra ! Bien évidemment, il n'y a pas de neurones dans une machine, mais il y a des structures de données sous la forme d'automates à seuil de déclenchement. Il n'y a pas de véritable capacité d'apprentissage par la machine, mais il y a des structures de contrôle positionnant des exemples prédéterminés sur les valeurs propres d'une matrice de transition d'un espace vectoriel de très, très, très grande dimension. Il n'y a pas non plus de profondeur dans une machine, mais des simulations de couches dites cachées dont le nombre s'accroit au détriment de la compréhension de

fonctionnement. Quoi qu'il en soit, le modèle des réseaux de neurones formels vient proposer une solution à l'automatisation d'un mécanisme associatif entre un signal d'entrée et une catégorie de sortie, même lorsque cette association ne se fait pas de façon linéaire : c'est-à-dire lorsque l'on ne peut pas séparer par un simple trait, les différentes catégories recherchées dans les signaux [ACK 85]. Et c'est bien évidemment le cas le plus courant ; on parle de problèmes non linéairement séparables.

Les bases de l'apprentissage profond, ou « *deep learning* » (DL), ont été proposées dès les années quatre-vingt, notamment par Y. Le Cun [LEC 19]. Mais le DL nécessite beaucoup de calculs pour pouvoir atteindre une performance rivalisant avec d'autres types de solutions. Ce n'est donc qu'en 2012 qu'un programme de DL a pu dépasser les autres solutions dans le domaine de la reconnaissance d'images [CAR 18]. Le terme « profond » provient du fait que plusieurs couches successives de groupes d'automates viennent traiter et transformer les informations initiales. On parle de couches de neurones dites convolutives car l'opération arithmétique de convolution [STR 83], très utile en traitement du signal, y joue un rôle fondamental. Il s'agit d'appliquer une petite fenêtre de quelques pixels (typiquement 3x3 ou 5x5) sur l'information délivrée par la couche de neurones précédente et d'en tirer une valeur globale pour cette petite fenêtre. L'opération est reproduite de proche en proche, sur toute la surface de l'image, et c'est donc une nouvelle image qui est ainsi fabriquée, en traduction des continuités ou ruptures présentes initialement. Ce processus de convolution est remis en œuvre sur une succession de couches, en étant entrecoupé de couches de traitement plus classique. À la fin du processus, le système a déterminé grâce à cette suite de calcul, une répartition des informations à apprendre dans un espace assez grand pour les loger et les distinguer toutes. Le réseau est alors prêt à être utilisé, en lui fournissant une information nouvelle pour lui, qui va être classée dans l'une des catégories apprises.

Si l'on dispose, par exemple, d'une dizaine de photos de visage de chaque individu dans une population d'un million de personnes (cela fait déjà dix millions de photos), on peut composer un réseau de neurones destiné à reconnaître chacune de ces personnes (il y a un million de catégories en sortie, soit une pour chaque individu). Quand quelqu'un se présente ensuite, par exemple devant une caméra de surveillance (il s'agit donc d'une nouvelle photo, jamais encore vue par le système), il devient alors possible de déterminer qui est cette personne parmi le million d'individus connus. Le système doit être assez robuste pour fonctionner même si la personne porte de nouvelles lunettes ou se coiffe différemment car les informations codées ne sont pas les informations sur le nez, les oreilles, ou sur tout autre signe que nous autres humains jugeons distinctif. Les informations stockées sont propres au fonctionnement de la machine et dépendantes de l'ensemble des visages-catégories. Si on porte un masque représentant le visage de quelqu'un de précis, il n'est même pas certain que le système n'associe pas le mystificateur à cette personne ! Ce qui a du sens pour un humain n'est pas nécessairement le point d'ancrage des repères de la machine, et en retour, ce qui constitue les valeurs d'apprentissage de la machine, n'a aucun sens pour les humains.

2.2. *Les réseaux antagonistes*

En 2016 un chercheur [GOO 16] a proposé d'utiliser un modèle d'IA composé de deux réseaux de neurones, pour produire des visages fictifs. Son idée a été de démarrer avec un réseau de neurones qui fonctionnerait « à l'envers ». Plutôt que de partir d'une image et d'en trouver la catégorie, le premier réseau part de catégories et propose par calcul un nouveau visage. Bien évidemment ce visage fabriqué n'en sera pas un et ne ressemblera à rien du tout, car il ne sera composé que de pixels répartis au hasard. Vient alors un second réseau de neurones préparé en entrée avec un ensemble d'images dont certaines sont de vrais visages, et disposant de seulement deux catégories de sorties : « visage » ou « non-visage ». Ce second réseau permet de juger si la production du premier est crédible ou non. Si le premier réseau produit n'importe quoi, il doit alors revoir ses valeurs d'attribution (appelées poids des neurones) et recommencer. Les deux réseaux sont dits « antagonistes » et on parle alors de « *Generative Adversarial Networks* » (GAN). C'est la

combinaison de ces deux réseaux de neurones, l'un créatif, l'autre examinateur, qui ouvre la voie à l'IA générative, en capacité de produire des « *fake faces* » [PAR 23], mais aussi des « *fake videos* » [ALD 22] et toutes autres images irréelles mais crédibles [ZHA 21].

Si le principe de la convolution fait que les pixels d'une image sont générés par petites fenêtres, de proche en proche, il est donc impossible pour la machine d'établir un véritable lien de symétrie entre les différentes parties de l'image produite. Ceci explique que les visages générés ont souvent des symétries gauche/droite imparfaites. Cela n'est pas gênant pour un visage puisque c'est souvent le cas dans la nature. En revanche les objets comme les chapeaux ou les paires de lunettes ont des asymétries à respecter et les erreurs sont alors troublantes [SHA 23]. Ces programmes d'IA générative doivent alors ajouter des traitements pour redresser ces éléments qui permettraient d'identifier immédiatement une supercherie.

2.3. *Les modèles de langage*

Lorsque la base d'apprentissage du GAN ne concerne plus des images, mais le langage, la machine doit alors disposer d'un très grand modèle de ce langage. On parle de LLM, pour Large Language Models. C'est à partir de 2018 que ces modèles ont été mis au point, dans l'objectif de fabriquer des agents conversationnels. Ils se basent sur des textes connus et y déterminent toute une série de concordances entre les mots : après chaque mot, la probabilité d'apparition de tel ou tel autre mot est de tant, etc. Si x mots avant, il s'agissait de tel mot, alors cette probabilité se modifie, etc. Le modèle du langage est ainsi fabriqué par des probabilités conditionnelles sur les occurrences de mots, de groupes de mots, voir de phrases entières. Les agents conversationnels basés sur ces modèles de langage vont alors générer un texte, mot à mot, non pas en fonction de connaissances particulières, mais en fonction de probabilités conditionnées par toute une série de paramètres. Les résultats sont suffisamment proches d'une génération naturelle pour que l'on obtienne un texte crédible, écrit tout à fait correctement, dans une langue que l'on peut choisir. Si on le souhaite, ce texte généré peut également répondre à des paramètres initiaux et contraintes imposées, tels qu'« en deux lignes », « à la manière de Victor Hugo », ou « dans un ton colérique », etc.

L'IA générative produit des éléments (textes, images, vidéos, et prochainement sons et mélodies bien que la recherche soit en retard dans ce domaine) le plus souvent faux, car il n'est pas possible, au hasard des paramètres, de tomber sur quelque chose de vrai. Pour cela il faudrait que les enchaînements des mots correspondent à des locutions dédiées à quelque chose de précis. Les opérateurs d'IA générative cherchent bien évidemment à renforcer cette véracité de la production, en jouant sur des règles de supervision et de production pilotées et contrôlées par des humains. L'ajout de ces règles « humaines » rend la production « machine » beaucoup plus performante. Quoi qu'il en soit, les domaines très techniques ou faibles en signification (comme la production de codes informatiques, de slogans publicitaires, de scénarios basiques, de résumés de texte) vont très rapidement être envahis par des productions issues d'IA générative.

La plupart des utilisateurs ou usagers de ces systèmes, tels que Dall-E® ou Midjourney®, qualifient leurs productions d'« inventions » ou de « créations ». C'est oublier que les machines n'ont pas de conscience, pas d'intention, pas de jugement [LEV 23]. Cet anthropomorphisme qui prête aux machines des opinions, de la politesse, des idées, est probablement le point le plus critique de cet envahissement par les IA génératives. Il témoigne en effet d'une sorte de crédulité spontanée, porte d'entrée à la manipulation potentielle des esprits. Par exemple, le terme d'« hallucination » a commencé à être employé à partir de mars 2023, pour désigner des productions de ChatGPT® crédibles mais invraisemblables. Il ne s'agit pourtant là ni d'une procédure particulière du programme, ni d'une dérive de ce programme. Dans son fonctionnement normal, ChatGPT® génère des phrases crédibles mais sans intentionnalité, ni besoin de cohérence ; la seule logique est ici statistique et la performance n'est que celle de l'illusion de l'utilisateur. Chaque emploi nouveau d'un terme anthropomorphe vient un peu plus brouiller les pistes.

3. Ce qui va changer avec l'IA

L'intelligence artificielle pénètre de plus en plus la société. Les technologies et les outils actuels ont de plus en plus recours à des algorithmes d'IA. En automatisant certaines reconnaissances de configurations, le but est d'enclencher des actions adaptées à ces reconnaissances. Parmi celles-là, il y a la production de nouveaux contenus, avec une crédibilité et un fort pouvoir de conviction qui font oublier l'imprécision ou l'inexactitude de l'information générée automatiquement. On va ainsi utiliser l'IA générative pour obtenir des textes plausibles, demander quelles sont les principales idées présentes dans d'autres textes, quelles sont les références à associer à telles ou telles idées, obtenir du système des références… et oublier que tout cela a été généré, donc se présente sous la forme de quelque chose de vraisemblable, mais pas nécessairement vraie. La plupart du temps ces références n'en sont pas, elles peuvent être inventées, ou être simplement des références hors contexte, exactes pour d'autres sujets et appliquées de manière statistique au contexte utile.

Un avocat new-yorkais en a subi les conséquences. Dans sa préparation d'un procès en début d'année 2023, il a utilisé ChatGPT pour construire son argumentation. Sans s'en rendre compte, il a ainsi généré une jurisprudence totalement fictive, en s'appuyant sur des références crédibles mais imaginaires. Le juge l'a confondu et cette affaire a été relatée dans la presse mondiale.

3.1. *L'influence omniprésente*

Avocats, publicitaires, journalistes, scénaristes, enseignants, chercheurs, etc. : tous ces métiers dits intellectuels vont vite être, et sont déjà pour certains confrontés à l'usage de l'IA générative et aux potentiels dégâts qui peuvent être provoqués. Il y a dix ans, une étude avait déjà mis en avant le problème de la mutation ou de la disparition des métiers, du fait de l'informatisation et de l'IA devenant opérationnelle. Les auteurs ont essayé de quantifier la part informatisable de chacun des métiers et en ont extrapolé que près de la moitié des métiers actuels disparaîtront ou seront bouleversés par l'informatique [FRE 17]. Il s'agit là d'une situation alarmiste, forçant le trait, et ne prenant ni en compte la temporalité des évolutions dans les professions, ni la création ou la transformation de certains emplois dévolus au meilleur fonctionnement des IA, dans un postulat de « destruction créatrice » [SCH 62].

Pourtant, un processus de mutation est certainement en œuvre. Il s'illustre dans le cas de l'informatisation apparue dans les années 1980 et accélérée dans les deux décennies suivantes. On a pu voir un passage progressif du papier aux fichiers, et l'apparition d'écrans sur à peu près tous les outils. Des entreprises se sont créées en commençant par poser leur système d'information et en structurant ensuite les différents métiers autour de cet axe. Une économie du numérique s'est mise en place et de nouveaux métiers sont apparus dans la fabrication et l'assemblage de machines, le commerce informatique, la programmation, la création et la maintenance des réseaux numériques. De nouveaux usages ont émergé, avec l'informatisation en santé, en sécurité, dans le commerce connecté, pour la gestion des stocks, la pédagogie en distanciel, les réseaux sociaux, etc. À l'heure actuelle, ce sont les données massives qui tirent les innovations en entreprise. Tout est numérisé, et sauf contraintes légales ou réglementaires, la trace papier a été abandonnée. On est passé du rapport écrit au tutoriel : ce n'est plus la documentation qui est recherchée mais plutôt des exemples filmés de situations qui lui sont préférés. Ces vidéos que l'on retrouve maintenant un peu partout, et qui composent les supports de formation, vont bientôt constituer les points d'ancrage des gestes professionnels. Dans ces années 2020, un schéma d'articulation se met en place à partir des données massives, des bases de vidéos et des algorithmes d'IA venant recommander la vidéo présentant le bon geste à accomplir en fonction de la situation reconnue selon les données captées. La notion de métier devient de plus en plus floue et la reconfiguration des gestes va venir structurer l'exercice même du travail.

La taille gigantesque de ces bases de données dépasse l'entendement. On parle de données massives pour désigner des informations en très grand nombre, en très grande variabilité de formes

et en très grand flux de production. Jusqu'à l'ordre de grandeur du milliard, on peut encore se faire une représentation mentale utile des choses. Au-delà, la quantité est vue comme un quasi infini, et toute croissance supplémentaire devient très difficile à concevoir. L'exponentielle n'est pas une fonction facile à manipuler [CLA 19b]. On raconte que l'inventeur du jeu d'échecs avait demandé à son Prince, séduit par ce jeu, une récompense en grains de blé : 1 grain sur la première case, 2 sur la deuxième, 4 sur la troisième, 8 sur la suivante et ainsi de suite en multipliant par deux à chaque case. Avant d'avoir atteint la moitié de l'échiquier, on a déjà plus d'un milliard de grains par case, et bien évidemment la croissance exponentielle continue (*ibid.*). Si tout un chacun peut se faire des repères pour comprendre comment se construit l'augmentation sur la première moitié de l'échiquier, plus personne n'en a pour comprendre la suite de cette croissance. Les potentialités liées aux calculs sur les machines croissent de façon exponentielle depuis déjà des dizaines d'années [MOO 75] et même si la loi de Moore semble dépassée [TUO 02], les nouveautés, notamment logicielles, arrivent aujourd'hui plus vite que quiconque peut les assimiler ou même en embraser l'exhaustivité documentaire.

Si l'on ne peut plus se faire une idée juste des capacités accrues des machines et des potentialités offertes par l'IA, il devient alors du ressort des dirigeants politiques que de se saisir de ce sujet. C'est ce qu'a fait l'administration Obama en 2016 en rendant public un rapport pour préparer le futur de l'IA [NAT 16], concluant sur 23 recommandations. Dans les deux années qui ont suivi, nombre de pays de l'OCDE se sont dotés d'une stratégie en IA, de façon d'une part à consolider ou étendre les bases de sa recherche scientifique, et d'autre part à aider les administrations et soutenir les entreprises dans la constitution de larges bases de données exploitables par les algorithmes d'IA.

3.2. *Expérience de la stratégie française*

En France la stratégie IA a été dévoilée par le Président de la République en 2018. En plus des deux volets sur la recherche et sur l'économie que l'on retrouve présents dans les stratégies des autres pays, un troisième volet a été ajouté, en rapport avec la réflexion sur l'enjeu éthique de l'IA. Comme évoqué plus haut, l'auteur de cet article était au cours de cette période membre de la Direction générale de la recherche et de l'innovation (DGRI) du Ministère de l'enseignement supérieur et de la recherche (MESR), et responsable du domaine scientifique de l'IA,. Il a été amené à discuter avec ses homologues allemands, canadiens et japonais. Tous étaient conscients des enjeux et soucieux d'agir et de mobiliser le niveau politique de leur état sur ce thème. Anticipant le fait que l'IA viendra très vite changer les relations entre les individus et entre individus et société, une organisation internationale a été décidée. L'idée de construire un « GIEC de l'IA » a été rapidement écartée, puisqu'il ne s'agit pas d'étudier un domaine comme peut l'être le climat, c'est-à-dire un sujet dont on ne connaît ni les contours, ni les règles de fonctionnement. Ici, il s'agit d'étudier les possibilités offertes par une IA que les scientifiques maîtrisent, en tant qu'objet technique, au bénéfice des humains. Le problème avec l'IA est d'arriver à donner un cadre aux agissements des individus. La structure adoptée a été celle d'un GPAI (Global Partnership for AI) dont l'animation et la gestion ont été confiées à l'Organisation de coopération et de développement économiques (OCDE).

Afin d'assurer la mise en place de ce groupe de réflexion, l'auteur a été chargé d'en dresser des axes de structuration. Sa proposition reposait sur trois niveaux : (i) celui de l'interaction avec les machines (le niveau de l'algorithme), (ii) celui relatif aux données personnelles (le niveau de l'individu), (iii) celui des différentes mises en action, comme dans le commerce (le niveau de la société). Un quatrième niveau a été ajouté pour traiter des (iv) relations internationales ; la réflexion était élaborée entre plusieurs pays. Les thèmes du mandat de ce GPAI concernaient la collecte et l'équilibre des données, le contrôle des données et le respect de la vie privée, le partage des modèles et des connaissances, la validation, la qualification, l'intelligibilité et la transparence des techniques d'IA, l'acceptabilité et l'appropriation de l'IA, l'avenir du travail, la gouvernance, les lois et la justice face à l'IA, l'IA responsable et les droits de la personne, et enfin l'équité, la responsabilité et

les biens publics. Cette longue liste des thèmes génériques montre bien que des pans entiers de la société sont attendus comme étant significativement impactés par les prises de décisions automatiques rendues possibles par l'IA.

L'État se préoccupe de ces évolutions et mobilise, pour ce qui est des domaines de sa responsabilité, l'effort de recherche ainsi que la formation de ses agents potentiellement concernés. La création du réseau français de quatre Instituts Interdisciplinaires d'Intelligence Artificielle (3IA) et de quarante autres chaires de recherche et d'enseignement en IA en est un exemple sans précédent de développement et de coordination d'un maillage, sur l'ensemble du territoire national. Par ailleurs, des programmes de recherche technologique ont été mis en place, tel que « Confiance.ai » qui vise à faciliter l'intégration de l'« IA de confiance » dans les systèmes critiques développés par de grands industriels nationaux. L'IA est ainsi considérée comme un enjeu de compétitivité scientifique, industrielle, économique et sociale, jugé comme constitutif d'opportunités de création de valeur, porteur de la transformation de la société, et d'enjeux de compétitivité et de souveraineté nationale.

3.3. L'éthique, droit et futur de l'IA

Les rapports du politique, de l'éthique et du droit sont depuis longtemps source de réflexion et de doctrine [AER 08] [SUM 95]. La Commission nationale de l'informatique et des libertés (CNIL) a rendu un rapport en 2017 sur l'éthique des algorithmes et de l'intelligence artificielle [CNI 17], déterminant six grands domaines de risques : (i) l'autonomie humaine face à l'autonomie des machines ; (ii) les biais sources de discrimination et d'exclusion ; (iii) la fragmentation algorithmique et la personnalisation contre les logiques collectives ; (iv) la nécessaire limitation des mégafichiers ; (v) les ruptures de qualité, quantité et pertinence des données de l'IA ; (vi) les dangers pour l'identité humaine. Le rapport propose des recommandations portant sur la formation et la recherche en IA et la sensibilisation à l'éthique, ainsi que sur la compréhensibilité et l'audit des algorithmes pour la maîtrise de l'effet « boîte noire » dans tous les domaines de la société. De grandes organisations internationales ont depuis développé leur propre approche du problème : l'UNESCO avec le programme du « Forum mondial sur l'éthique de l'IA », a adopté en 2021 à l'unanimité des 193 États membres [UNE 22] une « recommandation sur l'éthique de l'intelligence artificielle » basée sur le « respect de droits de l'homme et de la dignité », la transparence et l'équité et la responsabilité humaine dans le contrôle des systèmes d'IA notamment dans la gouvernance des données pour le respect de l'environnement et les écosystèmes, des genres, de l'éducation et de la recherche, de la santé et du bien-être social [AZO 19].

Une autre manière d'aborder le futur de l'IA se dresse à travers deux axes de réflexion. L'un d'eux vient décrire les domaines « proximal » versus « distal » de la décision artificielle, l'autre vient qualifier la manière dont les données sont recueillies, par déclaration consentie ou bien par trace des activités épiées. Plus orientée vers la personne et moins dépendant de la technique, cette approche s'inspire de certains auteurs du domaine des sciences humaines [DUB 02] [CLA 21]. Le premier axe, celui de la prise de décision, peut se voir comme un axe reliant la sociologie à la psychologie ; l'étude de l'action distale va bénéficier des sciences de la société, alors que l'étude de l'action proximale va bénéficier des sciences du comportement. Le second axe, celui des données, peut se voir comme un axe reliant les besoins en cadrage politique ; il y a un besoin d'éducation des personnes pour les avertir de conséquences liées aux données qu'elles déclarent, et il y a un besoin de protection des personnes en obligeant à anonymiser les données relevées automatiquement dans les traces de visite des sites web ou dans les relevés d'usages des applications sur téléphones mobiles.

Une autre source d'inspiration pour penser l'éthique de l'IA est dans les délibérations du Conseil d'État. Lors de la révision des lois sur la bioéthique en 2018, l'auteur pour le compte de la DGRI (*cf. supra*) a participé aux séances de discussion pour traiter de la question de l'incidence de l'IA et des

données numériques sur la bioéthique. Bien que ce sujet de l'IA et des données n'était pas encore mûr, cette question sera certainement au menu d'une prochaine révision de la Loi, qui intervient tous les sept ans. Si l'une des principales avancées de la révision de 2021 fut l'autorisation de la procréation médicalement assistée pour toutes les femmes, et le maintien de l'interdiction de la gestation pour autrui, la ligne de conduite du Conseil d'État a ainsi été réaffirmée : dignité, solidarité, liberté. Ainsi, avant même la question de la liberté de faire ce qui est techniquement possible, se posent d'une part la question de la solidarité et donc de la capacité à proposer cette action à tous et à toutes, et d'autre part, celle de la dignité de la ou des personnes impliquées dans la réalisation de cette action que la technologie rend possible. Ce triptyque pensé par le Conseil d'État a été plusieurs fois repris dans les débats de l'Assemblée Nationale [TOU 21]. La dignité d'abord, c'est-à-dire la primauté de la personne, ainsi que l'inviolabilité et l'intégrité de l'espèce humaine ; la solidarité ensuite, comme l'illustre la notion d'altruisme, ou celles plus spécifiques, dans le domaine de la santé, d'égal accès aux soins ; la liberté, enfin, qui vise à préserver la part de vie privée et donc l'autonomie de l'individu dans ses choix, y compris les plus intimes.

C'est donc sur ce triptyque que repose le modèle de bioéthique depuis plus de vingt-cinq ans et dont on pourrait s'inspirer pour élaborer une éthique de l'IA. La dignité, c'est-à-dire le respect que mérite quelqu'un, se traduit par une protection particulière de la personne face à une technologie et un monde numérique qui vient bouleverser les rapports des uns aux autres. La solidarité, se retrouve dans la conception du don altruiste et dans l'attention portée aux plus vulnérables, ainsi qu'à l'accès gratuit et universel des outils : individus vulnérables car moins conscients de ce qu'est ce monde numérique, moins en capacité d'en utiliser les outils ou d'y agir car se trouvant en déficit de connaissance face à des entreprises dont la place économique se fait en monétisant les données des personnes. La liberté individuelle, s'exprime enfin dans l'équilibre personnel entre éléments consentis et bénéfices tirés de ces prises de décision automatisées grâce aux algorithmes d'IA.

4. Conclusion

L'IA bouleverse le monde, et le monde doit s'adapter à ces modifications majeures et au bouleversement des relations interindividuelles comme socio-économiques. L'avenir des sociétés et des individus pourrait dépendre de ce que deviendront les IA, et de qui s'en servira, voire en aurait la maîtrise. Cela ne peut se faire sans un contrôle politique, de portée internationale, et une réflexion sur l'éthique de l'IA, de ses usages, de son accessibilité, et de ses objectifs. Des deux axes cités, celui concernant la portée des décisions, qui croise celui de la propriété et de l'accès des données, constituent un cadre pour tracer la façon dont pourrait être définie la dignité de la personne dans un monde numérique envahi par l'IA.

Les travaux sur l'éthique de l'IA sont nombreux, beaucoup de voix se font entendre et plusieurs arènes de discussion apparaissent, mais bien peu sur une orientation qui tirerait profit du triptyque du Conseil d'État. Pourtant le temps passe, les progrès techniques s'amoncellent et la réponse politique doit se structurer pour que dignité, solidarité et liberté ne laissent pas la place à des logiques purement technologiques, commerciales ou guerrières, comme c'est encore trop le cas aujourd'hui.

Benoit Le Blanc est professeur d'intelligence artificielle à l'Institut Polytechnique de Bordeaux (Bordeaux INP), directeur de l'École nationale supérieure de cognitique (ENSC) et président de l'Association française pour l'Intelligence Artificielle (AfIA).

5. Bibliographie

[ACK 85] Ackley D.H., Hinton G.E., Sejnowski T.J., "A learning algorithm for Boltzmann machines". *Cognitive Science*, vol.9, *n°1*, pp.147-169, 1985.

[AER 08] Aernoudt R., "Éthique et politique : un couple infernal", *Pyramides*, vol.16, *n°1*, pp.169-190, 2008.

[ALD 22] Aldausari A., Sowmya A., Marcus N., Mohammadi G. "Video Generative Adversarial Networks: A Review". *ACM Computing Surveys*, vol.55, *n°2*, pp 1–25, 2022.

[ATK 68] Atkinson R.C., Shiffrin R.M. "Human Memory: A Proposed System and its Control Processes", *Psychology of Learning and Motivation*, vol.2, pp.89-195, 1968.

[AZO 19] Azoulay, A. "Vers une éthique de l'intelligence artificielle". *Chroniques des Nations Unies*. 1er mars 2019. New-York (NY, USA): Département de la communication globale de l'Organisation des Nations Unies / United Nations Library, 2019.

[BAD 74] Baddeley A.D., Hitch G. "Working Memory", *Psychology of Learning and Motivation*, vol.8, pp.47-89, 1974.

[BOD 77] Boden M. *Artificial intelligence and natural man*. London (UK): Harvester/Basic Books, 1977.

[BOS 87] Boss G. *Les machines à penser : L'homme et l'ordinateur*. Zurich (CH): Éditions du Grand midi, 1987.

[CAR 18] Cardon D., Cointet J., Mazières A. "La revanche des neurones", *Réseaux*, vol.211, *n°5*, pp. 173-220, 2018.

[CEL 18] Commission d'enrichissement de la langue française. "Vocabulaire de l'intelligence artificielle - liste de termes, expressions et définitions adoptés". *Journal officiel de la République Française*, 9 décembre 2018, n°0285, Texte *n° 58*, p.2, 2018.

[CLA 19a] Claverie B. *Introduction à l'épistémologie et à la méthode de recherche*. Paris (FR): L'harmattan, 2019.

[CLA 19b] Claverie B. "Dynamique exponentielle et naturalité de l'intelligence artificielle", *Hermès*, vol.85, pp.189-200. Paris (FR): CNRS Éditions, 2019.

[CLA 21] Claverie B. *Des théories pour la cognition - différences et complémentarité des paradigmes*. Paris (FR): L'Harmattan, 2021.

[CNI 17] Commission nationale de l'informatique et des libertés, "Comment permettre à l'Homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle". Paris (FR): CNIL, 2017.

[COL 83] Colmerauer A., Kanoui H., Van Caneghem M. "Prolog bases théoriques et développements actuels", *Techniques et Science Informatiques*, vol.2, *n°4*, pp.271–311,1983.

[DUB 02] Dubucs J. "Calculer, percevoir et classer". *Archives de Philosophie*, vol.65, *n°2*, pp.335-355, 2002.

[ERM 08] Ermine J.-L. *Management et ingénierie des connaissances: Modèles et méthodes*, Paris (FR): Editions Hermes Science Publications, 2008.

[ERM 12] Ermine J.-L., Moradi M., Brunel, S. "Une chaîne de valeur de la connaissance". *Management international / International Management / Gestiòn Internacional*, vol.16, pp.29–40, 2012.

[FIA 17] FranceIA. *Rapport de synthèse – France Intelligence Artificielle*. Paris (FR): Gouvernement de la République Française, 2017.

[FIN 85] Fincher J., Pinchot R.B. (eds.). *The Brain: Mystery of Matter and Mind*. New-York (NY, USA): Scribner, 1985.

[FRE 17] Frey C.B., Osborne, M.A. "The future of employment: How susceptible are jobs to computerisation?", *Technological Forecasting and Social Change*, vol.114, *n°1*, pp.254-280, 2017.

[GOO 16] Goodfellow I., Bengio Y., Courville A. *Deep Learning.* Cambridge (MA, USA): MIT Press, 2016.

[JOV 22] Jovanović M., Campbell M. "Generative Artificial Intelligence: Trends and Prospects". *Computer*, vol.55, *n°10*, pp.107-112, 2022.

[LEB 13] Le Blanc B., Brunel S. "Les experts inégaux face à la communication de leur savoir". *Hermès*, n°2, pp.208-213, 2013.

[LEB 14] Le Blanc B. "Alan Turing: les machines à calculer et l'intelligence". *Hermès*, n°1, pp.123-126, 2014.

[LEC 19] Le Cun Y. *Quand la machine apprend, La révolution des neurones artificiels et de l'apprentissage profond*. Paris (FR): Éditions Odile Jacob, 2019.

[LEV 23] Leveau-Vallier A. *IA: L'intuition et la création à l'épreuve des algorithmes d'apprentissage profonds*. Paris (FR): Éditions Champ Vallon, 2023.

[MCC 56] McCarthy J. "The inversion of functions defined by Turing machines", in Automata Studies. *Annals of Mathematical Study*, n°34, pp.177-181, 1956.

[MCC 59] McCarthy, J. "Programs with Common Sense ». *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*. London: Her Majesty's Stationery Office, pp.75-91, 1959.

[MCC 60] McCarthy J. "Recursive Functions of Symbolic Expressions and Their Computation by Machine". *Communications of the ACM*, vol.3, *n°4*, pp.1-34, 1960.

[MCC 06] McCarthy J., Minsky M., Rochester N., Shannon C. "A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955". *AI Magazine*, vol 27, *n°4*, pp.12-14, 2006.

[MIN 56] Minsky M.L. "Heuristic Aspects of the Artificial Intelligence Problem". *Lincoln Laboratory Report*. Cambridge (MA, USA): MIT Press, pp.34-55, 1956.

[MIN 67] Minsky M. *Computation : Finite and Infinite Machines*. Upper-Saddle-River (NJ, USA): Prentice Hall, 1967.

[MOO 06] Moor J. "The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years". *AI Magazine*, vol.27, *n°4*, pp.87-91, 2006.

[MOO 75] Moore G.E. "Progress in Digital Integrated Electronics". *International Electron Devices Meeting*, IEEE, Washington (DC, USA), december 1-3, 1975. *Technical Digest*. pp.11-13, 1975.

[NAT 16] National Science and Technology Council & Office of Science and Technology. *Preparing for the future of AI*. October 12, 2016. Washington (DC, USA): Executive Office of the President of the USA, 2016.

[NEW 57] Newell A., Shaw J. C., Simon H.A. "Empirical Explorations of the Logic Theory Machine. À case Study in Heuristic", *Proceedings of the Western Joint Computer Conference*. New York (NY, USA): Institute of Radio Engineers, pp.218-230, 1957.

[PAR 23] Pardete J., Putri Setyaningrum A. "Implementation of Generative Adversarial Network to Generate Fake Face Image". *Jurnal Online Informatika*, vol.8, n°1, pp.44-51, 2023.

[RES 20] Restak R.M. *Mysteries of the Mind*. Washington (DC, USA): National Geographic Society, 2020.

[SCH 62] Schumpeter J.A. *Capitalism, socialism and democracy*. New-York (NY, USA): Harper & Row, 1962.

[SEA 80] Searle J.R "Minds, Brains and programs", *Behavioral and Brain Sciences*, vol.3, *n°3*, pp.417-424. Cambridge (MA, USA): Cambridge University Press, 1980.

[SHA 23] Sharma DK, Singh B, Agarwal S, Garg L, Kim C, Jung K-H. "A Survey of Detection and Mitigation for Fake Images on Social Media Platforms". *Applied Sciences*, vol.13, *n°19*, article 109802023, pp.1-36, 2023.

[SIM 83] Simon, H.A. "Why should machines learn?" In R.S. Michalski, J.G. Carbonell, T.M. Mitchell (eds.) *Machine learning: an artificial intelligence approach*, pp.25-37. Palo Alto (CA, USA): Tioga Publishing Company, 1983.

[SMO 86] Smolensky P. "Neural and conceptual interpretation of PDP models", in D.E. Rumelhart, J.M. McClelland (eds.) *Parallel distributed processing. Explorations in the microstructures of cognition*, vol. 2 : *Psychological and biological models*, pp.390-431. Cambridge (MA, USA): MIT Press, 1986.

[STR 83] Struppa D.C. "The Fundamental Principle for Systems of Convolution Equations". *Memoirs of the American Mathematical Society*, vol.41, *n°273*. Providence (RI, USA): American Mathematical Society Bookstore, 1983.

[SUM 95] Šumič-Riha J., "La politique existe-t-elle sans éthique?" *Acta Philosophica*, vol.16, *n°2*, pp.53-65, 1995.

[TOU 21] Touraine J-L., Dubost C., Berta P., Eliaou J.-F., Romeiro Dias L., Leseul G. "Rapport sur le projet de loi relatif à la bioéthique", *Journal officiel de la République Française*, *n°0131* du 8 juin 2021, texte 56, 2021.

[TUO 02] Tuomi, I. "The Lives and Death of Moore's Law". *First Monday*, vol 7, *n°11*, 2002.

[TUR 50] Turing, A. "Computing machinery and intelligence". *Mind*, vol.69, *n°236*, pp.433–460, 1950.

[UNE 22] Organisation des Nations Unies pour l'éducation, la science et la culture. "Recommandation sur l'éthique de l'intelligence artificielle - adoptée le 23 novembre 2021". Paris (FR): UNESCO, 2022.

[VON 49] Von Neumann, J. "The general and logical theory of automata". *Collected Works*, vol.5, pp.288-328, 1949.

[ZHA 21] Zhang E., Banovic N. "Method for Exploring Generative Adversarial Networks (GANs) via Automatically Generated Image Galleries". CHI'21 - *Conference on Human Factors in Computing Systems*, Yokohama (Japan), May 7-17, 2021. *Proceedings of the 2021 CHI*, article n°76, pp.1-15, 2021.

