

# De l'IA-analphabétisme à la XAI

Amal El Fallah Seghrouchni

digital-ia.org

Directrice du Centre International d'Intelligence Artificielle du Maroc **Ai movement**

Université Mohammed VI Polytechnique

Professeure en Sorbonne Université et Membre de la commission mondiale de l'éthique des sciences et des technologies de l'UNESCO

L'IA s'est imposée comme figure de proue du digital et comme catalyseur de la transformation organisationnelle, du développement économique et de l'instrumentalisation politique. Désormais, il est difficile d'imaginer une vie sans l'IA même sur Mars. *Perseverance* y a atterri avec une panoplie d'algorithmes d'IA dans ses bagages incluant le réordonnement de tâches sous contraintes spatiales, la prise et l'analyse d'images en temps réel, et des techniques de navigation autonome sans intervention humaine. Ce qu'il faut retenir, c'est que, comme de nombreux systèmes critiques, l'intelligence de *Perseverance* n'a acquis les données sensibles qu'une fois sur site !

L'IA joue également un rôle central dans le pacte vert pour l'Europe et dans la relance espérée suite à la pandémie de la Covid-19.

En géopolitique, la lutte engagée par les géants du numérique, les GAFAMI (GAFA + Microsoft + IBM) et les BHATX chinois, pour maîtriser l'IA s'apparente à un retour des empires. Le déploiement à large échelle de l'IA et la manipulation de grandes masses de données permettant un contrôle panoptique posent de nombreux problèmes éthiques et des questions majeures de gouvernance et de souveraineté. Les investissements au niveau mondial se multiplient (18 \$Mds en Amérique du nord ; près de 10 \$Mds en Asie, source CE – 2019) et finissent par convaincre les derniers sceptiques de l'ampleur du raz de marée de l'IA.



## Rien ne se perd, rien ne se crée, tout se transforme (A. Lavoisier)

Des transformations et des mutations profondes sont palpables dans tous les secteurs. Dans la sphère du travail, 47% des métiers seraient remplacés par des robots alors que 60 millions de nouveaux emplois pourraient être créés par l'IA et la robotique à horizon 2025. Dans la finance, les fintechs fleurissent avec les modèles prédictifs ; la santé de plus en plus augmentée par la télémédecine et l'holographie numérique ; la (cyber-) sécurité bénéficie d'algorithmes d'IA pour la détection de fraude ou d'intrusion ; l'agriculture de plus en plus précise grâce aux images satellitaires et aux drones. Les *legaltechs* s'installent durablement pour pallier la submersion en dossiers dans le milieu judiciaire. La défense s'appuie sur une artillerie d'algorithmes et de paradigmes d'IA comme les drones militaires, la planification multi-robots, la patrouille Air-Mer-Sol multi-agents, le développement et la simulation de stratégies collectives (jeux sérieux, adversarial games, neurogaming..), l'apprentissage par renforcement pour le suivi de cibles manœuvrantes dans des

environnements hostiles ; la smart mobilité et les véhicules autonomes dans les villes intelligentes ; les jumeaux numériques (digital twins) révolutionnent l'industrie 4.0 en introduisant des modes de contrôle, d'anticipation et de simulation inédits ouvrant la voie au déploiement d'une autonomie large échelle.

## L'IA impacte autant qu'elle fascine

**L'IA impacte la vie humaine**, y compris l'esprit humain, en amenant des nouvelles façons d'agir sur la pensée, les interactions et la prise de décision humaine. Les questions éthiques se posent avec acuité à telle enseigne que certains pays comme le Chili cherchent à instaurer la notion de « droits cognitifs » pour protéger leurs citoyens des manipulations cognitives (*nudges*, *neuromarketing*).

**Au niveau sociétal**, on assiste à une irrésistible tentation de contrôle à travers la mainmise sur les données personnelles comme l'affaire *Nightingale* de Google Cloud où 50 millions de dossiers médicaux de patients américains ont été utilisés à leur insu ; la surveillance de masse sous certains cieus peu soucieux de la vie privée ; l'emprise des géants de l'IA sur des pans entiers de nos vies privées ou encore les démocraties dévoyées par les réseaux sociaux (ex. Cambridge Analytica). De nouvelles formes d'aliénation se dessinent avec la collecte de données neuronales par phénotypage ou en utilisant *l'affective computing*. Ces données, plus caractéristiques que des empreintes digitales, peuvent permettre de fabriquer des chimères électroniques cérébrales et déclenchent le débat sur la (bio-)éthique.

Cette tentation franchit un pas aujourd'hui en entrant dans une phase de Mormonisme post-mortem. Microsoft a déposé le brevet d'un *chatbot* qui permettrait de communiquer avec des proches décédés. L'IA du *chatbot* récupère les données personnelles pour créer la version numérique (le *digital twin*) d'une personne vivante ou décédée. Le même principe est utilisé à des fins ludiques et parfois pédagogiques comme dans le *Salvador Dali Museum* de St. Petersburg, en Floride, où les visiteurs peuvent dialoguer avec un réplica de Dalí, l'interroger sur l'actualité ou même prendre un selfie.

**L'IA impacte également les écosystèmes** et l'utilisation d'algorithmes comme le *deep learning* (apprentissage profond) a une empreinte énergétique importante. Le fameux modèle de langage GPT-3 est entraîné avec 175 milliards de paramètres ; Une version réduite à 12 milliards de paramètres permet de générer des images à partir de descriptions textuelles en utilisant des données de paires texte-image.

Dans la vie quotidienne, si vous posez une question à Alexa (***son micro toujours ouvert et vous enregistre***), elle va mobiliser un ensemble extraordinairement complexe de couches de traitement de l'information alimentées par des marées de flux de voix humaines. Ces flux sont traduits sous forme de questions textuelles par un *module de reconnaissance automatique de la parole*, puis ***un module de traitement automatique de la langue (NLP) prend le relais pour traiter la question*** et ***un module de fouille de données / textes*** interrogera les bases de données de réponses potentielles (stockées au fur à mesure de l'apprentissage) pour *extraire* la réponse. Une fois la réponse trouvée, Alexa rebrousse chemin au bout duquel, ***un module de dialogue transformant le texte en parole, fournit la réponse orale à la question, et Alexa continue d'interagir avec vous !***

Pour chaque question/ réponse, l'échelle des ressources requises est de nombreuses magnitudes supérieures à l'énergie et au travail qu'il faudrait à un humain pour regarder sa montre si la question était « Alexa, il est quelle heure stp ? ». Une question plus complexe du type « Alexa, mettez-moi ma musique préférée » ! nécessiterait de construire et de maintenir un modèle de vos préférences musicales qui peuvent varier selon le contexte : votre humeur décelée grâce à l'intonation de la voix, l'heure ou encore l'ambiance générale si vous avez quelques objets connectés autour. Ce scénario pose de nombreuses questions : a-t-on envie de livrer notre « intérieur » à Alexa ? si on accepte aujourd'hui l'argument « on n'a rien à cacher », il faudra accepter que demain « on n'ait plus où cacher ». Une autre question est « faut-il mobiliser autant d'énergie pour connaître l'heure ? comment recourir à l'IA de façon écologique et responsable ? ».

## Bienvenue dans l'ère cognitive

C'est dans ce contexte mondial reconfiguré par l'émergence de l'économie numérique que la société se retrouve propulsée dans l'ère cognitive qualifiée de la société 5.0 et où la connaissance occupe une place de choix.

La data est le carburant des algorithmes de l'IA. Les grandes masses de données transcendent l'individu car elles ne concernent plus un individu en particulier ou un petit groupe d'individus, mais des groupes vastes aux contours mal définis. Collectées à partir de sources hétérogènes (objets connectés, réseaux sociaux, *chatbots*, transactions en ligne, etc.) elles sont utilisées, croisées et fusionnées pour créer des « profils » d'individus spécifiques précis et informatifs. Si la performance algorithmique est avérée, la conception de systèmes intelligents, requiert une vision « full stack of AI » pour alimenter une approche intégrative et hybride, où différents paradigmes de l'IA vont concourir à la réalisation de fonctionnalités différentes mais complémentaires au sein du système. Alexa utilise de l'IA symbolique pour le NLP, la modélisation de l'utilisateur pour inférer ses préférences, l'IA statistique pour trouver les réponses, les méthodes probabilistes et l'apprentissage automatique pour améliorer ses réponses, etc.

- L'IA statistique est incarnée aujourd'hui par l'apprentissage automatique ou le Machine Learning. Il s'agit de mécanismes d'apprentissage basés sur des modèles intrinsèques de systèmes de neurones biologiques, des algorithmes de classification et des modèles prédictifs qui manipulent des données et utilisent très souvent l'apprentissage statistique comme l'apprentissage profond. Ces apprentissages s'avèrent d'une efficacité redoutable mais souffrent de leur opacité. Qualifiés de boîtes noires, leurs résultats sont inexplicables et par conséquent non certifiables et difficilement applicables pour des systèmes critiques. Ces difficultés ont impulsé le nouveau champ de la recherche scientifique XAI (eXplainable Artificial Intelligence) dont l'objectif est de décrire son objet d'étude, sa logique et sa prise de décision de manière intelligible.
- L'IA symbolique, centralisée ou distribuée, est la seule explicable aujourd'hui. Elle se base sur une représentation des connaissances pour élaborer des raisonnements logiques basés sur des symboles abstraits. Elle utilise un langage formel explicite pour l'expression des connaissances (ex. des ontologies) et des procédures d'inférence formelle pour opérer des déductions. Il est important de noter que rien n'est « caché » dans le code : c'est une approche entièrement déclarative et plus facile à tracer et à expliquer. Elle est de ce fait plus intuitive et plus compréhensible par l'Humain car elle implique plusieurs capacités cognitives : le raisonnement pour résoudre des problèmes ; la connaissance pour aider à représenter et à comprendre le monde ; la planification pour établir et atteindre des objectifs ; l'apprentissage symbolique ; la communication pour comprendre le langage et communiquer ; et la perception pour transformer des entrées sensorielles brutes (images, sons...) en informations utilisables.

Au vu de la complexité des systèmes d'IA aujourd'hui, une approche hybride et intégrative permet d'articuler, d'expliquer, de faire collaborer plusieurs agents et d'orchestrer différentes briques de l'IA au sein de systèmes intelligents avec de réels impacts économiques et sociétaux. On parle d'IA de confiance.

## **Une XAI c'est bien, une XAI éthique et responsable c'est mieux**

En tant que science et technologie, l'IA interpelle l'humain dans ce qu'il a de plus intime. Elle soulève de nombreuses questions éthiques comme la subordination algorithmique qui remet à l'ordre du jour les questions de partage d'autorité entre l'humain et la machine active voire pro-active (automatisation versus autonomie), ou les biais algorithmiques et de données qu'ils soient cognitifs, économiques ou statistiques.

De nombreux efforts au niveau mondial sont déployés pour délimiter les contours éthiques des technologies de l'IA. L'éthique de l'IA peut être vue comme une base dynamique pour l'évaluation et l'orientation normatives des technologies de l'IA, en respectant la dignité humaine, le bien-être et la prévention des dommages des repères et en s'appuyant sur l'éthique des sciences et des technologies. Plusieurs cadres ont été établis par des organisations intergouvernementales, par le secteur privé, par des organisations professionnelles, des ONG et par de nombreux laboratoires de recherche en éthique. Ces efforts s'accordent tous pour adopter une vision éthique de l'IA et contribuer aux efforts globaux pour apporter des réponses

responsables aux conséquences connues et inconnues des interactions des systèmes d'IA avec les êtres humains, leur société et leur environnement.

**L'IA éthique et responsable relève de l'enjeu démocratique aujourd'hui.**