

RÉALITÉS INDUSTRIELLES

« Se défier du ton d'assurance qu'il est si facile de prendre et si dangereux d'écouter »
Charles Coquebert, Journal des mines n°1, Vendémiaire An III (septembre 1794)



L'intelligence artificielle au quotidien



Notre site



MAI 2025

Publiées avec le soutien
de l'Institut Mines Télécom



RÉALITÉS INDUSTRIELLES

ISSN 2271-7978 (en ligne)

ISSN 1148-7941 (imprimé)

Série trimestrielle - Mai 2025

Rédaction

Conseil général de l'Économie (CGE)
Ministère de l'Économie, des Finances
et de la Souveraineté industrielle et numérique
120, rue de Bercy - Télédock 797
75572 Paris Cedex 12
Tél. : 01 53 18 52 68
<http://www.annales-des-mines.org>

Grégoire Postel-Vinay

Rédacteur en chef

Alexia Kappelmann

Secrétaire générale

Daniel Boula

Secrétaire général adjoint

Magali Gimon

Assistante de rédaction et Maquettiste

Nuria Gorris

Webmestre et Maquettiste

Publication

Photo de couverture

générée par IA

Iconographie

Alexia Kappelmann

Mise en page

Nuria Gorris

Impression

Dupliprint Mayenne

Membres du Comité de rédaction

Serge Catoire

Président du Comité de rédaction

Godefroy Beauvallet

Pierre Couveinhes

Jean-Pierre Dardayrol

Jean-Marc Grognet

Hervé Mariton

Grégoire Postel-Vinay

Françoise Roure

Rémy Steiner

Christian Stoffaës

Claude Trink

Benjamin Vignard

La mention au regard de certaines illustrations du sigle « D. R. » correspond à des documents ou photographies pour lesquels nos recherches d'ayants droit ou d'héritiers se sont avérées infructueuses.

Le contenu des articles n'engage que la seule responsabilité de leurs auteurs.

L'intelligence artificielle au quotidien

4

Introduction

Serge CATOIRE et Arnaud MAZIER

Cadre général

7

L'IA générative : une révolution technologique entre promesses et défis

Éric MOULINES

11

Le programme d'intelligence artificielle du CEA, ses applications

François TERRIER

16

Les enjeux d'une IA responsable

Brigitte D'ANDREA NOVEL, David RYCKELYNCK et Georges-André SILBER

19

La stratégie chinoise en matière d'IA

Paul JOLIE

L'IA dans les administrations, la formation et le droit

23

Réussir l'après-Paris 2025 : quelles ambitions pour la France à la suite du Sommet pour l'action sur l'IA ?

Arno AMABILE

27

L'IA générative dans l'administration française : le cas d'« Albert »

Ulrich TAN

30

Comment le système éducatif français se saisit-il de l'intelligence artificielle générative ?

Émilie-Pauline GALLIÉ et Erwan PAITEL

34

Droit et intelligence artificielle : interactions et transformations

Yannick MENECEUR

39

L'office de l'avocat à l'heure du numérique et de l'IA : entretien avec M^e Christiane FÉRAL-SCHUHL

Propos recueillis par Arnaud MAZIER

43

L'utilisation de l'IA pour améliorer la qualité et la réactivité dans les travaux de la DGFIP : avancer et innover par vents budgétaires contraires

Esther MAC NAMARA et Thomas BINDER

Les applications aujourd'hui dans les entreprises

47

L'IA pour la médecine de précision en oncologie

Thomas WALTER

51

Les applications de l'IA pour l'aide à la radiologie

Jérôme KNOPLIOCH et Nicolas GOGIN

56

L'utilisation de l'IA pour mieux tirer parti d'une base documentaire technique au bénéfice des industriels de la mécanique

Fanny LAMBERT

60

Le potentiel de l'intelligence artificielle pour une entreprise technologique industrielle

François MARION

63

Construction : quand l'IA change la donne

Marie-Luce GODINOT

67

GEN AI : le nouveau choc d'autonomie

Pierre DULON

69

L'intelligence artificielle en pratique pour faire face aux défis d'aujourd'hui et de demain

Emmanuel COX

73

IA de confiance condition nécessaire pour le déploiement de l'IA dans les systèmes critiques

Juliette MATTIOLI et Christophe MEYER

76

Blaxtair : le défi du véhicule industriel intelligent pour une industrie plus sûre et plus efficace

Sabri BAYOUDH

81

Mobilisation de l'intelligence artificielle pour les transports publics : un intérêt opérationnel évident, qui doit rester tourné vers les services aux voyageurs

Hélène BRISSET

Hors dossier

86

Qu'est-ce que la "deep tech" ? Définitions et enjeux stratégiques

Benjamin CABANES

90

Traduction des résumés

94

Biographies des auteurs

Ce numéro a été coordonné
par Serge CATOIRE et Arnaud MAZIER

Introduction

Par Serge CATOIRE

Ingénieur général des Mines honoraire, directeur industrie à l'UIMM

Et Arnaud MAZIER

Ingénieur général des Mines

L'émergence récente et soudaine des IA génératives n'a pas relégué les IA déterministes au passé. Par contre, elle a dévoilé un champ inattendu, inscrivant de surcroît l'IA dans le débat public et l'ancrant dans les pratiques quotidiennes. Ces nouvelles technologies ouvrent de nouveaux champs du possible. Les articles réunis dans ce numéro décrivent, à travers divers exemples, un état actuel de l'utilisation effective de l'intelligence artificielle, qu'elle soit déterministe ou générative. Appliqués à de nombreux domaines, ces exemples peuvent inspirer et susciter la réflexion, grâce à leur mise en contexte et aux perspectives qu'ils présentent.

L'IA générative a ouvert récemment des possibilités radicalement nouvelles, et connu une percée rapide : ChatGPT a ainsi atteint 100 millions d'utilisateurs en deux mois. Éric Moulines, professeur à l'École polytechnique et membre de l'Académie des Sciences, replace ces percées dans l'histoire de la discipline ; il illustre des cas d'usage dans le raisonnement mathématique et le génie des procédés, et souligne qu'elle a démontré son potentiel pour créer, assister et innover. Il faut gérer prudemment cet outil puissant mais imparfait, et l'améliorer encore pour en tirer le meilleur parti. Des modèles encore plus performants apparaissent régulièrement et conduisent à une évolution de la relation homme-machine, promettant d'augmenter l'ingéniosité humaine, à condition d'en connaître les limites et de l'orienter vers des objectifs bénéfiques et responsables.

Dès 2017, le CEA a lancé des programmes pour la maîtrise d'une IA « de confiance », en particulier pour des systèmes critiques et embarqués, qui nécessitent des solutions de calcul frugales et à faible coût. François Terrier, directeur de recherches au CEA-LIST, détaille les nombreux cas d'application pratiques étudiés avec une dizaine de grands acteurs industriels dans le cadre du Grand Défi pour la « fiabilité des IA pour les systèmes critiques ». Le CEA déploie aussi l'IA pour ses propres travaux scientifiques et a mis en place pour les PME la plateforme PRISM, nœud français du projet européen de plateforme d'expérimentation et de test de l'IA pour le *manufacturing*.

Avant même l'arrivée de l'IA générative, le numérique représentait en France un dixième de la consommation électrique. Brigitte d'Andréa-Novel, directrice de l'Institut des Transformations Numériques de l'École des Mines de Paris, et ses collègues décrivent les pratiques qu'explorent des chercheurs de Mines Paris - PSL pour un numérique plus efficace utilisant des données de manière raisonnée et intégrant des connaissances *a priori*. Les réseaux neuronaux informés par la physique (PINN) et l'apprentissage fédéré sont des approches prometteuses. Renforcer la recherche académique publique permettra de continuer à trouver des méthodes d'apprentissage automatique plus efficaces et moins gourmandes en énergie.

L'arrivée de Deepseek en janvier 2025 n'aurait dû surprendre personne. Paul Jolie, Ingénieur général des Mines au Conseil général de l'Économie, souligne qu'à partir des années 1970, la Chine a intégré l'intelligence artificielle (IA) dans ses plans nationaux de recherche et a poursuivi ses efforts avec constance. La Chine se classe actuellement au premier rang mondial en termes de publications scientifiques et de brevets en IA, et développe activement des applications dans ce domaine au point de compter, en 2023, 4 482 entreprises spécialisées dans l'IA et 52 licornes. Elle aspire à faire de son système normatif et réglementaire une référence mondiale, notamment pour les pays non alignés.

Arno Amabile, conseiller de l'envoyée spéciale pour le Sommet pour l'action sur l'IA, partage les enseignements et les suites de cet événement mondial visant à faire de la France « une puissance de l'IA ». Dans une stratégie de « *fast follower* », la France peut s'appuyer sur plusieurs atouts : une expertise en IA frugale et embarquée, une maîtrise des systèmes de confiance, et une proximité avec l'industrie européenne, riche en données et en clients potentiels. Pour tirer le meilleur parti de l'IA, les initiatives du Sommet mettent l'accent sur le déploiement et la diffusion de cette technologie.

Lancé mi-2023, « Albert » vise à assurer la confidentialité des données et maîtriser la puissance de calcul, et est une gamme de produits et de services de la direction interministérielle du Numérique de l'État. Ulrich Tan, qui y dirige l'Étalab et le Datalab, en décrit les trois piliers et leurs résultats : 1) co-construction sur mesure : 44 « France services », dans 6 départements disposent aujourd'hui d'un assistant virtuel pour chercher une solution aux problèmes concrets des usagers ; 2) « IA as a service » : une API permet aux administrations d'accélérer leurs déploiements d'IA ; 3) commun numérique ouvrant à une communauté le plus grand jeu de données administratives ouvertes au monde.

Erwan Paitel, rapporteur de la Commission sur l'intelligence artificielle : « IA : notre ambition pour la France », et Émilie-Pauline Gallié, de l'Inspection générale de l'Éducation, du Sport et de la Recherche (IGESR), décrivent les actions menées au sein de l'Éducation nationale pour intégrer l'IA en tant qu'assistant à l'enseignant et outil d'aide aux apprentissages. L'objectif est de renforcer l'interaction humaine dans la classe pour améliorer l'éducation et les apprentissages des élèves. Des défis techniques, pédagogiques et environnementaux sont abordés pour rendre l'IA générative accessible à une échelle inédite formée de 15 millions d'apprenants et d'un million d'enseignants.

Les applications de l'intelligence artificielle dans les domaines du droit et de la justice sont devenues courantes. Dans ce contexte, elles peuvent être analysées sous deux angles : leur encadrement juridique en tant qu'objets de droit et leur influence sur le droit en tant que nouveaux acteurs. L'article de Yannick Meneceur, Inspecteur de la justice au ministère de la Justice, explore les interactions et transformations actuelles, en décrivant les nouveaux cadres juridiques européens et les enjeux sociétaux liés à l'utilisation de l'IA dans les systèmes judiciaires, au-delà de la traditionnelle balance entre bénéfices et risques.

Christine Féral-Schul, qui a présidé le Conseil national des Barreaux et le Barreau de Paris, aborde les enjeux et les risques du numérique pour la profession d'avocat. Pour les avocats, l'intelligence artificielle est devenue incontournable. Ne pas l'utiliser pour les recherches et l'aide à la réflexion pourrait être critiqué. Cependant, le numérique présente des risques, notamment de fuite de données clients ou d'approximation pouvant amener les avocats à négliger les principes de loyauté et de prudence. La formation des avocats doit évoluer pour intégrer ces technologies tout en préservant l'approche humaine et critique du droit.

Depuis 2014, la DGFIP utilise le *data mining*, puis le *machine learning* afin d'optimiser la collecte fiscale et renforcer la lutte contre la fraude. Esther Mac Namara, qui a conduit la transformation numérique de la DGFIP de 2023 à 2025, et Thomas Binder, responsable du département IA de la Délégation à la Transformation numérique de la DGFIP, présentent plusieurs projets internes démontrant la valeur ajoutée de cette technologie depuis l'émergence de l'IA générative en 2023, notamment l'outil LLaMandement, facilitant le résumé et l'affectation automatique des amendements parlementaires aux services concernés. Pour surmonter les contraintes matérielles et organisationnelles identifiées, la DGFIP a établi, en 2024, sa feuille de route stratégique en vue d'un déploiement plus étendu de l'intelligence artificielle.

Appliquée à l'oncologie, l'IA contribue à mieux identifier la présence et la nature des cancers, pour orienter leur traitement. Thomas Walter, directeur du centre de Bioinformatique (CBIO) au sein de l'unité jointe entre l'Institut Curie, Mines Paris et l'Inserm, donne des exemples de cas d'usage : quantification automatique pour faciliter le travail du pathologiste, identification des métastases ganglionnaires, détection du sous-type moléculaire d'un cancer pour anticiper sa réponse au traitement. Les travaux du CBIO ont déjà permis d'améliorer significativement la prédiction de réponse au traitement, donc le choix du traitement, et promettent de nouveaux progrès.

Parmi les 1016 dispositifs médicaux « intelligents » répertoriés par la FDA, 777 concernent la radiologie. Les applications couvrent l'ensemble du processus, depuis la reconstruction des images jusqu'à l'assistance au diagnostic pour les spécialistes cliniques en sous-effectif, dont le volume d'activité a été multiplié par dix au cours des quinze dernières années, en passant par la communication avec les patients. L'article de Nicolas Gogin, directeur en charge des développements en intelligence artificielle pour la visualisation avancée, et Jérôme Knoploch, responsable des architectures pour la visualisation avancée à GE Healthcare, décrit les méthodes utilisées ainsi que les résultats obtenus et leurs perspectives futures.

Afin d'améliorer l'exploitation des documents techniques pour les industriels de la mécanique, le Cetim a fait développer un *chatbot* interne capable d'exploiter de manière sécurisée et contrôlée des milliers de documents produits en interne. Le projet, désormais finalisé, utilise les documents de veille du Cetim pour générer des résumés, des synthèses et des comparatifs, facilitant ainsi la collecte et l'analyse rapide d'informations scientifiques et techniques. Fanny Lambert, responsable au Cetim de la veille sur les IA génératives, partage avec le lecteur les enseignements du projet quant aux méthodes, aux coûts, aux infrastructures IT nécessaires, et aux compétences requises pour l'intégration d'une IA en interne.

Dans les systèmes d'aide à la conduite tels que le freinage d'urgence ou la navigation assistée, Valeo s'appuie sur l'intelligence artificielle depuis treize ans et a ouvert dès 2016 Valeo.ai – à l'époque, le premier centre de recherche mondial en IA dédié aux applications automobiles. François Marion, directeur de la communication et membre du comité exécutif du groupe Valeo, décrit l'ampleur des applications actuelles dans l'entreprise : maintenant prédictive, analyse de spécifications techniques, accélération de la validation de logiciels et de la production de codes informatiques, amélioration des contrôles qualité visuels et de la consommation énergétique... Le potentiel de transformation concerne également la finance, les RH et la communication.

Marie-Luce Godinot, directrice générale adjointe du groupe Bouygues en charge de l'innovation, du développement durable et des systèmes d'information, illustre l'utilisation de l'IA générative pour analyser et répondre aux appels d'offre, optimiser la conception, déployer la vision artificielle et le *machine learning* sur les chantiers pour améliorer la sécurité, réduire les nuisances et gérer les déchets. L'IA est aussi utilisée en maintenance pour les infrastructures énergétiques et la maintenance préventive. Les opportunités continuent de croître à mesure que les solutions mûrissent.

Il y a quarante ans, l'introduction des ordinateurs de bureau et des capacités de calcul décentralisées a significativement transformé les métiers bancaires et financiers, tout en stimulant la croissance des activités liées aux fusions-acquisitions et aux marchés financiers. Pierre Dulong, directeur général adjoint de CACIB, en charge de l'Informatique et des Opérations bancaires, souligne que l'intelligence artificielle générative permet désormais à chacun d'analyser et de naviguer dans l'immense quantité de données disponibles, initiant ainsi une nouvelle révolution de décentralisation. Dans ce contexte, Crédit Agricole Corporate and Investment Bank (CA-CIB), à travers son « projet Humain », relève les défis humains et managériaux associés à cette évolution technologique.

SNCF Réseau, responsable des 28 000 km de voies ferrées françaises et employant 50 000 personnes, doit répondre à une demande croissante tout en modernisant ses infrastructures. Emmanuel Cox, responsable de la Transformation numérique de SNCF Réseau, rappelle que depuis plus de dix ans, l'entreprise utilise l'IA pour la maintenance prédictive, l'analyse des données, la gestion du trafic et l'anticipation des aléas climatiques. Elle développe également des projets d'IA générative pour la gestion documentaire. Un cadre éthique encadre l'utilisation de l'IA, avec un accompagnement des salariés et des partenariats académiques et industriels.

Thales élabore des solutions destinées à des secteurs critiques tels que la défense et la cybersécurité, en mettant l'accent sur une intelligence artificielle de confiance. Juliette Mattioli, présidente du *hub* « *Data sciences & IA* » du pôle de Systematic, et Christophe Meyer, directeur technique du grand laboratoire IA de Thales cortAIx Labs France, expliquent comment Thales garantit la validité, la transparence, la sécurité, la fiabilité et la responsabilité de ses systèmes en combinant des approches symboliques et connexionnistes (IA hybride). L'entreprise veille à ce que ses systèmes soient robustes contre les cyberattaques et transparents pour les utilisateurs. Thales s'engage dans des initiatives visant à développer des méthodologies d'ingénierie de l'IA de confiance, tout en optimisant les ressources utilisées et en réduisant l'empreinte carbone de ses activités.

Selon les données des États-Unis, 11 % des chariots élévateurs sont impliqués dans un accident corporel, conduisant à plus de 80 décès et jusqu'à 62 000 blessures par an. Sabri Bayouh, directeur de l'innovation d'Arcure-Blaxtair, décrit la technologie d'IA embarquée pour la sécurité des véhicules industriels développée par Arcure à partir de travaux commencés en 2009, s'appuyant sur la vision par ordinateur et l'apprentissage profond pour analyser les environnements complexes et prévenir les accidents. Les véhicules ainsi « intelligents » collectent des données pour optimiser la gestion des sites industriels. L'intégration de l'IoT et de la 5G permet une digitalisation accrue, ouvrant la voie à des jumeaux numériques et à une meilleure efficacité opérationnelle.

Le réseau de transport collectif de l'Île-de-France est le deuxième plus fréquenté au monde. Hélène Brisset, directrice du Numérique d'Île-de-France Mobilités, explique notamment comment, à l'occasion des Jeux Olympiques et Paralympiques, IdF Mobilités a déployé l'application Transport Public Paris 2024, conçue pour la recherche d'itinéraires vers les sites olympiques et l'achat de titres de transport, intégrant de manière native des modules de traduction automatique par IA. Dans ce même cadre, une expérimentation de vidéoprotection intelligente a été réalisée afin de détecter les situations à risque, renforçant ainsi la sûreté. IdF Mobilités continue de développer et d'appliquer des solutions basées sur l'IA pour améliorer les services offerts aux voyageurs.

L'IA générative : une révolution technologique entre promesses et défis

Par **Éric MOULINES**

Professeur de statistique au Centre de Mathématiques appliquées (CMAP) de l'École Polytechnique

L'intelligence artificielle générative (IA générative) désigne les modèles capables de produire des contenus originaux comme du texte, des images ou des sons à partir de données d'entraînement. Elle marque une rupture par rapport à l'IA classique grâce à sa polyvalence et sa capacité à réaliser des tâches variées sans apprentissage spécifique préalable. Popularisée récemment par des modèles comme GPT-4 et DALL-E, l'IA générative transforme déjà de nombreux domaines tels que le raisonnement mathématique, l'ingénierie ou la création artistique. Toutefois, cette technologie soulève aussi des enjeux importants en matière d'éthique, de biais, d'impact environnemental et de fiabilité. Bien que puissante, elle nécessite donc une utilisation raisonnée, intégrant étroitement expertise humaine et bonnes pratiques, pour réaliser pleinement son potentiel tout en minimisant les risques.

Introduction

L'intelligence artificielle générative (IA générative) désigne un ensemble de modèles d'IA capables de produire un contenu original – textes, images, sons ou autres données – à partir de données d'apprentissage. Contrairement aux approches prédictives ou discriminatives, une IA générative peut créer des données inédites imitant les caractéristiques des données d'entraînement. Par exemple, un modèle peut rédiger un paragraphe original en réponse à une question ou concevoir l'image d'un objet imaginaire à partir d'une description textuelle (« prompt »).

Cette capacité de génération automatique ouvre des possibilités radicalement nouvelles dans de nombreux domaines, ce qui explique l'essor spectaculaire de l'IA générative au cours des dernières années. Des outils comme les grands modèles de langage (par exemple GPT-3.5/ChatGPT) ou de création d'images (par exemple DALL-E) sont rapidement passés de prototypes de recherche à des applications diffusées à grande échelle (Hu, 2023¹) ("DALL E: Creating images from text", OpenAI²). Lancé fin 2022, le *chatbot* ChatGPT a ainsi atteint 100 millions d'utilisateurs en deux mois, un record historique ("Fact of the week: OpenAI's ChatGPT user base has grown faster..."³).

¹ <https://www.scirp.org/reference/references-papers?referenceid=3561402#:~:text=Hu%2C%20K.growing>

² <https://openai.com/index/dall-e/#:~:text=DALL%2C%B7E%3A%20Creating%20images%20from%20text,concepts%20expressible%20in%20natural%20language>

³ <https://itif.org/publications/2023/02/13/openai-chatgpt-user-base-has-grown-faster-than-tiktoks-or-instagram/#:~:text=Fact%20of%20the%20Week%3A%20OpenAI%27s.note%2C%E2%80%9D%20Reuters%2C%20February%202%2C%202023>

L'adoption de l'IA générative par les acteurs économiques a été exceptionnellement rapide : 70 % des dirigeants disent, en 2024, explorer l'IA générative⁴, et 85 % des responsables informatiques estiment en 2023 qu'elle jouera un rôle significatif, voire majeur dans leur organisation⁵. L'IA générative s'impose donc comme une avancée à la fois scientifique et applicative.

Nous présenterons tout d'abord un bref historique de l'IA générative et de ses évolutions récentes. Nous analyserons ensuite en quoi l'IA générative s'inscrit dans la continuité de l'IA classique tout en constituant une rupture. Puis nous illustrerons deux cas d'usage : le raisonnement mathématique et le génie des procédés. Enfin, nous discuterons des atouts de l'IA générative et de ses limites (biais, éthique, consommation énergétique, fiabilité), avant de conclure sur les perspectives.

Historique et évolutions récentes de l'IA générative

L'idée de faire produire du contenu par une machine est ancienne : dès les années 1960, le programme ELIZA simulait un dialogue humain, créant l'illusion d'une conversation. Par la suite, des approches fondées sur des règles puis des modèles probabilistes – notamment les modèles markoviens – ont été explorées pour générer automatiquement du texte, puis de la parole ou de la musique.

Bien que ces méthodes aient produit des résultats jugés spectaculaires par les initiés, la qualité restait insuffisante

⁴ https://www.forrester.com/blogs/previsions-tendances-2024-intelligence-artificielle-fr/?utm_source=chatgpt.com

⁵ https://www.salesforce.com/fr/company/news-press/press-releases/2023/08/230822/?utm_source=chatgpt.com

pour des applications grand public, à de rares exceptions près. Un saut qualitatif s'est opéré dans les années 2010 avec l'apprentissage profond (*deep learning*), caractérisé par un accroissement très important du nombre de paramètres et de la complexité des modèles. En 2014, les réseaux antagonistes génératifs (GAN) ont fait franchir un cap en permettant la génération d'images photoréalistes grâce à deux réseaux neuronaux en compétition. En 2017, l'architecture Transformer a révolutionné le traitement du langage en modélisant efficacement de longues séquences grâce au mécanisme d'attention ("Zero-shot text-to-image generation"⁶). S'appuyant sur ces avancées, Google et OpenAI ont développé de grands modèles de langage pré-entraînés à partir de 2018 : GPT (Generative Pre-trained Transformer) et BERT.

Au tournant des années 2020 apparaissent des modèles de taille gigantesque. En 2020, GPT-3 (175 milliards de paramètres) démontre qu'un modèle entraîné sur des corpus massifs peut accomplir des tâches variées sans entraînement spécifique (*zero-shot*). En 2021, OpenAI présente DALL-E, capable de générer des images à partir de descriptions textuelles ("DALL·E: Creating images from text", OpenAI⁷) ("Zero-shot text-to-image generation"⁸). L'année 2022 marque l'essor des modèles à diffusion (*diffusion models*) tels que Stable Diffusion, offrant au grand public la génération d'images de haute qualité à partir de texte. Enfin, fin 2022, ChatGPT popularise l'IA générative textuelle en mode conversationnel, suivi par des modèles multimodaux toujours plus puissants (par exemple GPT-4 en 2023). En l'espace de quelques années, l'IA générative est passée du laboratoire à une technologie grand public : les progrès ont été extrêmement rapides, portés par une communauté de chercheurs très active dans le monde académique, mais surtout industriel, où la concurrence s'est exacerbée pour contrôler cette nouvelle frontière.

Continuités et ruptures avec l'IA classique

L'IA générative s'inscrit en continuité avec les approches antérieures de l'IA. Elle repose sur les fondamentaux de l'apprentissage profond – apprentissage de représentations (encodages vectoriels denses), mécanismes d'attention, architectures « profondes » avec connexions résiduelles – mais poussés à un niveau de complexité supérieur. Ces modèles s'apparentent à des « modèles de base » (*foundation models*) généralistes pouvant servir à divers usages.

L'IA générative introduit toutefois une véritable rupture par son efficacité et sa polyvalence inédites. D'une part, la puissance de calcul et les volumes de données mobilisés lui donnent des performances qualitatives jamais vues : les textes ou images qu'elle produit peuvent tromper un œil non averti sur leur origine purement synthétique. D'autre part, ces modèles s'adaptent à de multiples

tâches avec un réapprentissage minimal – voire sans réapprentissage – (écriture de code, résumé d'article, résolution de problèmes mathématiques, etc.), là où l'IA classique nécessitait un modèle spécifique par problème. C'est donc un changement d'échelle et de paradigme dans l'automatisation : on dispose désormais de modèles généralistes servant de base à de nombreux usages. En résumé, si les principes sous-jacents préexistaient, leur mise en œuvre à grande échelle a provoqué un saut qualitatif perçu comme une rupture dans la pratique.

Applications pratiques de l'IA générative

Les capacités de l'IA générative s'appliquent à un vaste éventail de situations concrètes. En voici deux illustrations dans des domaines très différents.

Raisonnement mathématique

Des modèles de langage avancés peuvent aider à résoudre des problèmes complexes nécessitant habituellement un raisonnement logique approfondi. Par exemple, le modèle *open source* DeepSeek-Math 7B est parvenu à résoudre plus de la moitié des problèmes du jeu de données MATH, avec un score de 51,7 % en top 1, approchant le niveau de modèles bien plus grands (GitHub – "deepseek-ai/DeepSeek-Math: DeepSeekMath: Pushing the limits of mathematical reasoning in open language models"⁹). De son côté, AlphaCode de DeepMind a montré qu'un modèle génératif pouvait atteindre un niveau compétitif en programmation, se classant autour du 54^e centile des participants à des concours de code ("Competitive programming with AlphaCode", Google DeepMind¹⁰). En 2024, deux nouvelles IA – AlphaProof, un système de raisonnement formel, et AlphaGeometry 2, un modèle spécialisé en géométrie – ont résolu quatre des six problèmes de l'International Mathematical Olympiad (IMO), obtenant un score de 28/42 équivalent à une médaille d'argent ("AI achieves silver-medal standard solving International Mathematical Olympiad problems : r/MachineLearning"¹¹). AlphaProof a produit des preuves formelles (en *lean*) pour des problèmes d'algèbre et de théorie des nombres, tandis qu'AlphaGeometry a résolu un problème de géométrie plane, ses solutions faisant preuve d'une créativité jugée « surhumaine » par des experts (Pushmeet Kohli sur X¹²) – performance médaillée d'or dans la résolution de la géométrie olympique avec AlphaGeometry¹³. Ces systèmes ont toutefois nécessité de plusieurs minutes

⁹ <https://github.com/deepseek-ai/DeepSeek-Math#:~:text=DeepSeekMath%20is%20initialized%20with%20DeepSeek,RL%20models%20to%20the%20public>

¹⁰ <https://deepmind.google/discover/blog/competitive-programming-with-alpha-code/#:~:text=As%20part%20of%20DeepMind%E2%80%99s%20mission,coding%2C%20and%20natural%20language%20understanding>

¹¹ https://www.reddit.com/r/MachineLearning/comments/1ebyx03/n_ai_achieves_silvermedal_standard_solving/#:~:text=level%2F%20deepmind

¹² <https://x.com/pushmeet/status/1816502058758869014#:~:text=Pushmeet%20Kohli%20on%20X%20We%27re,combines%20AlphaProof%2C%20a%20new>

¹³ <https://www.chatpaper.ai/fr/paper/a2594ad9-b0ae-4186-8e3a-d87ad6033928#:~:text=qui%20combine%20plusieurs%20arbres%20de,qui%20r%C3%A9sout%20de%20mani%C3%A8re%20fiable>

⁶ <https://arxiv.org/abs/2102.12092#:~:text=Title%3AZero>

⁷ <https://openai.com/index/dall-e/#:~:text=DALL%C2%B7E%3A%20Creating%20images%20from%20text,concepts%20expressible%20in%20natural%20language>

⁸ <https://arxiv.org/abs/2102.12092#:~:text=Title%3AZero>

à plusieurs jours de calcul pour trouver certaines solutions, et leur succès a requis une traduction manuelle des énoncés en langage formel ("AI achieves silver-medal standard solving International Mathematical Olympiad problems", LessWrong¹⁴). Néanmoins, ils constituent une avancée majeure vers des IA capables de raisonnement symbolique de haut niveau.

Optimisation des procédés industriels

Les modèles d'IA générative, en capturant des structures complexes au sein de grands jeux de données (à l'image des capacités linguistiques de modèles comme ChatGPT), se révèlent efficaces pour relever des défis majeurs en ingénierie des procédés, notamment la gestion de phénomènes non linéaires et incertains propres aux systèmes chimiques. Les modèles génératifs ont ainsi émergé comme de puissants outils de conception moléculaire, permettant la génération automatique des molécules candidates et la planification de leurs voies de synthèse. Parmi les avancées notables, on peut citer la création de structures métallo-organiques optimisées pour la séparation du gaz naturel dans des milieux poreux ("Inversely designing new materials for natural gas separation"¹⁵), l'optimisation de l'usage du lithium et d'autres minerais critiques dans le développement de nouvelles batteries (par exemple les batteries lithium-soufre), ou encore la conception de médicaments innovants à activité antibiotique ciblée ("Generative artificial intelligence model can easily design billions of..."¹⁶). Ces modèles ont progressivement évolué, passant de représentations moléculaires simples basées sur des chaînes de caractères (telles que le format *smiles*) vers des approches par graphes moléculaires. En maintenance prédictive, les GAN ont été utilisés pour générer des données synthétiques, simulant des scénarios rares ou difficiles à observer expérimentalement, avec une précision atteignant 99 % pour la détection de défauts sur des moteurs à induction ("Classification framework of the bearing faults of an induction motor..."¹⁷) ("Intelligent fault diagnosis of rotating machine via expansive dual..."¹⁸). Des modèles autorégressifs ont permis de modéliser précisément la dynamique complexe d'une colonne de distillation discontinue, avec des écarts inférieurs à 10 % par rapport aux données réelles de test (KEEN, 2023¹⁹). Enfin, les grands modèles de langage (LLM) font l'objet de recherches pour leurs capacités exceptionnelles, de la prédiction des propriétés de matériaux à l'extraction d'informations à partir de données non structurées

¹⁴ <https://www.lesswrong.com/posts/TyCdgpCfX7sfiobsH/ai-achieves-silver-medal-standard-solving-international#:~:text=official>

¹⁵ <https://www.advancedsciencenews.com/inversely-designing-new-materials-for-natural-gas-separation/#:~:text=Reference%3A%20Yao%2C%20Z,Generative%20Models%2C%20Nature%20Machine>

¹⁶ <https://www.news-medical.net/news/20240322/Generative-artificial-intelligence-model-can-easily-design-billions-of-novel-antibiotics.aspx#:~:text=...%20www.news.Nature>

¹⁷ <https://ouci.dntb.gov.ua/en/works/980rNk29/#:~:text=The%20experimental%20results%20demonstrate%20that,classification%20accuracy.%20The>

¹⁸ <https://dl.acm.org/doi/10.1016/j.eswa.2024.125398#:~:text=Intelligent%20fault%20diagnosis%20of%20rotating,2019%29%2085%E2%80%93>

¹⁹ <https://keen-plattform.de/Ergebnisse/Ver%C3%B6ffentlichungen/2023.html#:~:text=Brand,202200228>

("Leveraging large language models for predictive chemistry", *Nature Machine Intelligence*²⁰). Un exemple remarquable est « nach0 », un modèle multi-domaines combinant capacités linguistiques et chimiques : en exploitant un LLM, nach0 peut générer des molécules, planifier leur synthèse et prédire leurs propriétés ("nach0: multimodal natural and chemical languages foundation model"²¹).

Potentiel et limites de l'IA générative

L'IA générative offre d'importants atouts mais soulève aussi des défis qu'il faut garder à l'esprit. Du côté des atouts, elle permet d'automatiser la création de contenu et la recherche de solutions, ce qui peut accélérer de nombreux travaux d'ingénierie, de conception ou d'analyse. Elle agit comme un assistant créatif capable de générer des idées ou des formes nouvelles (textes, images, molécules, preuves mathématiques) à une vitesse et une échelle inatteignables par un être humain. Bien utilisée, elle peut non seulement augmenter la productivité, mais aussi transformer en profondeur de nombreux secteurs (ingénierie, droit, marketing) et même la recherche scientifique.

En revanche, plusieurs limites et risques accompagnent cette technologie.

Biais des modèles

En étant entraînés sur des données humaines imparfaites, les modèles génératifs peuvent reproduire, voire amplifier des biais (stéréotypes, discrimination) présents dans leurs corpus. Cela peut conduire à des contenus injustes ou offensants, posant un problème éthique et de fiabilité dans certains usages (recrutement, assistance juridique, etc.).

Usages malveillants

La même puissance générative peut servir à des fins malveillantes. On a vu émerger des *deepfakes* et d'autres formes de désinformation. Par exemple, un faux discours du président ukrainien Zelensky appelant à la reddition a circulé en 2022, première tentative documentée de manipuler l'opinion publique *via* un *deepfake* vidéo ("Generative AI: Fact or fake?"²²). La génération massive de faux contenus (avis, commentaires, propagande ciblée) est une menace réelle, soulevant la question d'un encadrement éthique et juridique de l'IA générative pour prévenir ces dérives.

Consommation énergétique

Les modèles les plus puissants requièrent d'immenses ressources de calcul pour leur entraînement et leur fonctionnement. Leur empreinte carbone est très élevée :

²⁰ <https://www.nature.com/articles/s42256-023-00788-1#:~:text=Nature%20www.a%20large%20language%20model%20hackathon>

²¹ <https://pubs.rsc.org/en/content/articlelanding/2024/sc/d4sc00966e#:~:text=model%20pubs,question%20answering%2C%20named%20entity>

²² <https://dialog.vde.com/en/vde-dialog-editions/2022-03-artificial-intelligence/ki-generative-ki#:~:text=%E2%80%9C%20advise%20you%20to%20lay,recent%20years%2C%20and%20the%20possible>

entraîner un modèle de langage géant peut émettre plusieurs centaines de tonnes de CO₂. Une étude de 2019 a estimé que l'empreinte carbone de certains entraînements de réseaux neuronaux était environ cinq fois supérieure à celle de la durée de vie d'une voiture moyenne ("An AI's carbon footprint is 5 times bigger than a car's"²³). Cet aspect peu durable appelle au développement de techniques d'IA plus économes, et à mesurer l'impact environnemental des projets impliquant ces modèles.

Fiabilité et « hallucinations »

Bien qu'impressionnante, l'IA générative n'est pas infaillible. Elle peut fournir des informations fausses ou incohérentes avec le même aplomb que des réponses correctes. Ces « hallucinations » incluent par exemple l'invention de références ou de faits imaginaires ("8 times AI hallucinations or factual errors caused serious problems", Originality.AI²⁴). Un cas notable est celui d'un avocat ayant soumis en justice un document rédigé par ChatGPT qui contenait de fausses jurisprudences fabriquées par le modèle. Il est donc risqué de faire confiance aveuglément aux sorties d'un modèle génératif sans vérification humaine, surtout dans les domaines critiques. La fiabilité limitée de ces systèmes constitue aujourd'hui l'une des principales barrières à une adoption totale, d'où la nécessité de les utiliser en tandem avec l'expertise humaine et des outils de validation.

Conclusion

L'IA générative représente un progrès majeur de l'IA contemporaine, combinant une continuité dans la trajectoire technologique (amélioration des modèles d'apprentissage existants) et une rupture dans les usages et les performances. Elle a démontré son potentiel pour créer, assister et innover dans de multiples secteurs, du génie industriel jusqu'à la production de contenu. Dans le même temps, elle nous confronte à de nouveaux défis en termes de fiabilité, d'éthique et de durabilité, qu'il faudra relever pour en tirer le meilleur parti.

Il est important de considérer l'IA générative comme un outil puissant mais imparfait. Bien utilisée, en gardant un esprit critique et en encadrant ses dérives possibles, elle peut devenir un formidable levier d'efficacité et de créativité. L'avenir verra sans doute apparaître des modèles encore plus performants et spécialisés, ainsi qu'un renforcement des bonnes pratiques (audit des biais, réduction de l'empreinte carbone, etc.). L'IA générative n'est pas une mode passagère mais une évolution de fond de la relation homme-machine. Elle offre la promesse d'augmenter l'ingéniosité humaine, à condition d'en connaître les limites et de l'orienter vers des objectifs bénéfiques et responsables.

²³ <https://www.popularmechanics.com/technology/infrastructure/a27793543/artificial-intelligence-carbon-footprint/#:~:text=roles%20as%20the%20fields%20look,side%20to%20AI%3A%20energy%20consumption>

²⁴ <https://originality.ai/blog/ai-hallucination-factual-error-problems#:~:text=Image>

Bibliographie

- ALSHEHRI A. S. & YOU F. (2022), "Deep learning to catalyze inverse molecular design", *Chemical Engineering Journal*, 444, 136669.
- BENDER E. M., GEBRU T., MCMILLAN-MAJOR A. & SHMITCHELL S. (2021), "On the dangers of stochastic parrots: Can language models be too big?", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610-623, ACM.
- BOMMASANI R. *et al.* (2021), "On the opportunities and risks of foundation models", arXiv preprint arXiv:2108.07258.
- BROWN T. B. *et al.* (2020), "Language models are few-shot learners", *Advances in Neural Information Processing Systems*, 33, pp. 1877-1901.
- DEEPMIND (AlphaProof & AlphaGeometry teams) (2024), AI achieves silver-medal standard solving International Mathematical Olympiad problems, DeepMind, [Blog post, July 25].
- GILLHAM J. (2024), 8 times AI hallucinations or factual errors caused serious problems. Originality. AI Blog.
- GOODFELLOW I. *et al.* (2014), "Generative adversarial nets", *Advances in Neural Information Processing Systems*, 27, pp. 2672-2680.
- HOTTELET U. (2022), Generative AI: Fact or fake ? VDE Dialog, 3/2022.
- HU K. (2023), ChatGPT sets record for fastest-growing user base – analyst note, Reuters.
- JABLONKA K. M., SCHWALLER P., ORTEGA-GUERRERO A. & SMIT B. (2023), "Leveraging large language models for predictive chemistry", *Nature Machine Intelligence*, 5(3), pp. 234-242.
- LI Y. *et al.* (2022), "Competition-level code generation with AlphaCode", *Science*, 378(6624), pp. 1092-1097.
- LIVNE M. *et al.* (2024), "nach0: Multimodal natural and chemical languages foundation model", *Chemical Science*, <https://doi.org/10.1039/D4SC00966E>
- RAMESH A. *et al.* (2021), "Zero-shot text-to-image generation", arXiv preprint arXiv:2102.12092.
- ROMBACH R., BLATTMANN A., LORENZ D., ESSER P. & OMMER B. (2022), "High-resolution image synthesis with latent diffusion models", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684-10695.
- SHAO S., WANG P. & YAN R. (2019), "Generative adversarial networks for data augmentation in machine fault diagnosis", *Computers in Industry*, 106, pp. 85-93.
- SHAO Z. *et al.* (2024), "DeepSeekMath: Pushing the limits of mathematical reasoning in open language models", arXiv preprint arXiv:2402.03300.
- SWANSON K., LIU G., CATAUTAN D. B., ARNOLD A., ZOU J. & STOKES J. M. (2024), "Generative AI for designing and validating easily synthesizable and structurally novel antibiotics", *Nature Machine Intelligence*, 6(3), pp. 338-348.
- VASWANI A. *et al.* (2017), "Attention is all you need", *Advances in Neural Information Processing Systems*, 30, pp. 5998-6008.
- YAO Z. *et al.* (2021), "Inverse design of nanoporous crystalline reticular materials with deep generative models", *Nature Machine Intelligence*, 3(1), pp. 76-86.

Le programme d'intelligence artificielle du CEA, ses applications

Par François TERRIER¹

Directeur de recherches au CEA-LIST

Nouvelle électricité du futur, l'IA ouvre un champ très large d'applications. Leur déploiement soulève toutefois des questions de confiance et de frugalité qui pourraient bloquer sa diffusion industrielle. L'Europe a réagi en mettant en place un cadre réglementaire centré sur les risques, qui doit assurer un développement serein des applications. Il amène des enjeux de qualification particulièrement complexes, c'est l'objet d'un des programmes d'action du CEA en IA. Conjointement, le déploiement dans des produits et systèmes industriels pose la question de l'embarquabilité et des besoins ressources, qui pourraient être incompatibles avec les usages visés et le coût de ce type de solutions. C'est pourquoi un deuxième programme du CEA est centré sur les solutions pour l'IA embarquées. Parallèlement, l'IA offre également un fort potentiel d'usage dans les activités scientifiques elles-mêmes, ce qui amène son intégration aux recherches de pointe du CEA. Ces différentes facettes des enjeux de l'IA sont illustrées sur quelques cas d'usage bien représentatifs des enjeux et solutions proposées.

Préparer le déploiement industriel

Pour de nombreux acteurs, l'IA est une technologie transformante dans la durée pour l'ensemble des activités : « L'IA est la nouvelle électricité » (Ng, 2017). Ceci résulte d'avancées multiples : collection de masses de données, hautes performances d'apprentissage, disponibilité de technologies de calcul très efficaces. Elles ont permis d'aborder un champ très large d'applications avec des performances suffisantes pour envisager un usage industriel.

Elles ont toutefois été obtenues au prix d'une augmentation énorme en complexité dans des processus statistiques globaux mal interprétables et d'une consommation de ressource à la croissance peu soutenable. Leur déploiement a ainsi immédiatement soulevé les questions clés de confiance et de frugalité, qui auraient pu devenir de nouveaux murs bloquant la diffusion de ces technologies.

La réglementation : une opportunité à ne pas manquer

Pas de sûreté de fonctionnement sans référentiel de réglementation. L'adoption de nouvelles technologies ne peut se faire que si les utilisateurs ont confiance dans leur emploi et leur impact sur leurs activités et le fonctionnement de la société. Cela impose des règles de qualité et d'assurance de conformité et donc l'ajustement des réglementations. Aussi, les pays européens,

via la Commission européenne, ont lancé en 2017 des groupes de travail sur les usages de l'IA sous tous ses aspects, et en particulier éthiques, vis à-vis des valeurs sociétales européennes (cf. "High-level expert group on artificial intelligence"²). Deux réglementations sont déjà entrées en vigueur : l'AI Act³ et, en lien fort, le Data Act⁴. La première vise à donner un cadre de maîtrise des risques liés aux usages de l'IA, la seconde couvre le sujet plus vaste de la gestion des données, avec l'objectif de faciliter le développement d'un marché de la donnée équilibrant la répartition de la valeur entre producteurs de données et exploitants, tout en assurant un niveau de protection adéquat.

Une stratégie CEA sur la confiance et l'embarqué

À l'apparition de ces nouveaux enjeux de déploiement opérationnel de l'IA, le CEA a lancé dès 2017 la construction de deux programmes ciblés sur ces deux sujets clés : l'IA de confiance et les solutions de calcul pour l'IA. Depuis, il a poursuivi la mobilisation de ses forces à la fois sur la recherche sur les fondements de l'IA⁵ et sur le développement des usages industriels.

² <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>

³ <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> ; <https://artificialintelligenceact.eu/fr/>

⁴ <https://digital-strategy.ec.europa.eu/fr/policies/data-act>

⁵ Programme national de recherche prioritaire en IA piloté par le CEA, le CNRS et Inria, <https://www.pepr-ia.fr/>

¹ Remerciements à mes collègues de différentes équipes du CEA pour les travaux cités : Z. Chihani, O. Bichler, O. Ferret, J. Bobin.

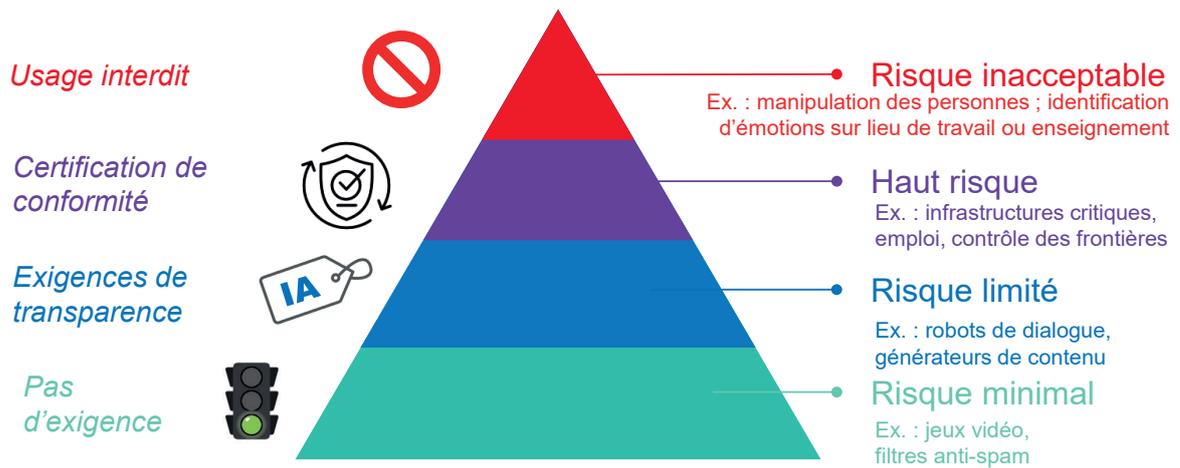


Figure 1 : L'AI Act : une approche basée sur les risques (Source : EU Digital sStrategy).

La confiance, clé de l'industrialisation

La complexité des technologies de l'IA, et tout particulièrement de celles à base d'apprentissage machine, la multiplicité des usages envisagés, leur nouveauté et la rapidité de l'évolution technologique, rendent le sujet particulièrement ardu. L'AI Act a été construit pour fixer les limites de questionnement légitime, notamment en adoptant une approche basée sur les risques et en centrant les exigences de conformité fortes sur les systèmes à hauts risques. Il reste toutefois à construire les bases d'ingénierie et théoriques pour la maîtrise de ces systèmes. C'est l'objectif du programme IA de confiance du CEA en abordant l'ensemble du cycle : spécification, conception, vérification, validation et suivi en exploitation des systèmes à base d'IA.

Le point dur de l'apprentissage machine, en particulier profond, arrive avec, d'une part, une technologie intrinsèquement opaque dont les éléments internes sont difficiles à relier rationnellement aux objectifs et résultats, et, d'autre part, des processus de développement empiriques, avec souvent aucune formalisation de l'objectif, de la fonction à réaliser.

Un programme national pour construire la confiance

Un grand programme national a été mis en place pour explorer différents moyens d'augmenter la confiance dans les IA, avec une dizaine de grands acteurs industriels dans le cadre du Grand Défi pour la « fiabilité des IA pour les systèmes critiques » : le programme Confiance.ai⁶. Il a réalisé l'analyse de l'ensemble du processus d'ingénierie à la fois des applications d'IA et de leur intégration dans des systèmes de différentes criticités. Pour cela, il a conduit des évaluations de briques technologiques (académiques et du marché) sur une douzaine de cas d'usage industriels. On peut citer parmi ces cas : l'optimisation de la chaîne d'approvisionnement par vision (Air Liquide), l'assistance au contrôle de qualité de soudures dans l'automobile (Renault), l'optimisation du trafic aérien (Thales), la détection de la piste d'atterrissage (Airbus) et la perception pour l'aide à la conduite (Valeo). Ces travaux

seront par ailleurs valorisés *via* la création de la start-up SafenAI⁷ dont l'ambition est d'offrir un support outil à la gestion des processus d'ingénierie et à la production des éléments de qualification et de conformité.

Jusqu'à la preuve de fonctionnement pour des IA critiques

Si les systèmes critiques usuels s'appuient sur une formalisation des spécifications et des propriétés de sûreté en exploitant couramment des méthodes formelles (mathématiques), ces approches semblaient hors de portée des systèmes à base d'IA, en particulier avec l'apprentissage profond. La communauté a toutefois beaucoup progressé sur le sujet, notamment sur la formalisation des domaines de fonctionnement – *operational design domain* (ODD) – (Adedjouma *et al.*, 2024) et la vérification de sûreté (Chihani, 2021).

Une application particulièrement emblématique du sujet et de l'apport de l'IA est le contrôle d'anticollision de drones aéronaves. Le système bien maîtrisé avec des technologies classiques s'appuie sur un cas d'usage ouvert, appelé ACAS XU (Owen *et al.*, 2019), qui facilite la recherche par de multiples acteurs.

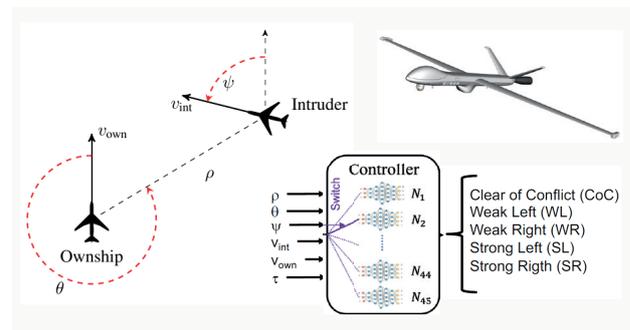


Figure 2 : La problématique de l'anticollision de drones (Crédits Onera).

L'objectif est de passer d'une implémentation classique certifiable, mais qui demande plus de ~2 GB

⁶ <https://www.confiance.ai/>

⁷ <https://safenai.io/>

de mémoire embarquée, à une implémentation par réseaux de neurones dont l'empreinte mémoire pourrait être plusieurs ordres de grandeur plus petite, mais qui amène alors la question de sa certification. L'idée est que le réseau de neurones va naturellement faire une approximation statistique de la fonction originale complexe et permettra d'avoir un modèle approché plus compact, on parle de *surrogate model*.

On est dans un cas favorable où le domaine de fonctionnement et les propriétés de sûreté sont formellement définis. Des travaux réalisés dans le cadre des projets Confiance.ai (pour la certification) et DeepGreen⁸ (pour l'optimisation embarquée) ont permis de développer une implémentation à base d'apprentissage profond, puis une optimisation pour le déploiement embarqué particulièrement compact qui permet de gagner un facteur 1 000. La qualification du domaine de fonctionnement sûr de cette implémentation peut être réalisée à l'aide de techniques d'analyse du modèle d'IA et de l'algorithme d'inférence par interprétation abstraite. Cette analyse sans exécuter le modèle permet de définir des espaces géométriques des valeurs d'entrées dans lesquels des propriétés sont vérifiées. Cela permet de segmenter de manière mathématiquement prouvée le domaine de fonctionnement en espaces de fiabilité, de défaillance certaine, et incertains. La preuve a été réalisée à l'aide d'un outil dédié⁹ développé au CEA (deuxième place de la compétition internationale de vérification des réseaux de neurones¹⁰).

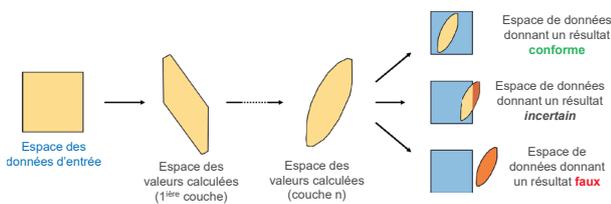


Figure 3 : Preuve de sûreté par domaine de fonctionnement (Source : CEA List).

Ces méthodes formelles ont ainsi permis de définir un fonctionnement prouvé sûr dans une enveloppe couvrant plus de 73 % du domaine opérationnel. Ces résultats permettent déjà d'envisager des exploitations avec des approches hybrides : IA pour la partie prouvée, et classique pour le reste (Gabreau *et al.*, 2024).

L'embarqué : une question essentielle au déploiement industriel

Le déploiement massif de l'IA dans des systèmes industriels se heurte aux capacités de calcul nécessaires pour sa mise en œuvre, à la consommation énergétique et aux coûts associés, et ce même si on se limite à leur exploitation (phase d'inférence, après finalisation de l'apprentissage). Pour pallier ce problème et permettre un déploiement concurrentiel, il est nécessaire de disposer de solutions de calcul frugales et à faible coût. Deux axes sont possibles : l'optimisation drastique des

modèles appris ; et le développement de nouveaux composants de calcul très faible consommation.

Sur le premier axe, le CEA s'appuie sur le projet national de développement d'une plateforme ouverte et souveraine pour le déploiement de l'IA dans l'embarqué, DeepGreen. L'objectif est de redonner à l'industrie une liberté de choix des cibles matérielles de déploiement, en fournissant un environnement qui ne sera plus tributaire des politiques de grands acteurs privés vis-à-vis des chaînes de développement et des composants ciblés. Le projet fournit pour cela la plateforme *open source* AIDGE¹¹, qui offre les meilleures techniques de compression et de compilation des modèles et assure leur déploiement sur une diversité de cibles matérielles : des composants sur étagères aux composants émergeant des nouveaux acteurs. Elle repose sur l'exploitation du standard de représentation d'un modèle à base de réseaux de neurones (ONNX) et d'une architecture modulaire permettant d'intégrer de nouveaux opérateurs (selon les algorithmes et cibles utilisés) et d'intervenir en analyse, optimisation, compression et compilation à tous les niveaux du processus. Il s'appuie sur un consortium de plus de 20 partenaires industriels et académiques avec 12 cas d'usage de référence. La section précédente a déjà évoqué son usage pour le système d'anticollision de drones, on peut aussi mentionner, dans un domaine très complémentaire, celui porté par EDF pour la géolocalisation des personnes dans des bâtiments de production énergétique. Cette application associe sur un même équipement, à base de composants du marché bas coût porté par une personne (contrainte de volume et d'énergie), des fonctions de localisation par odométrie et par vision avec un traitement des données par plusieurs IA.

Parallèlement, sur le deuxième axe, différents travaux du CEA portent sur le développement de nouvelles solutions de calcul pour l'IA. Ils vont de la conception d'architectures de calcul à des travaux fondamentaux sur les technologies microélectroniques visant à exploiter les propriétés du silicium pour définir des modèles de calcul adaptés à l'IA et intrinsèquement porter la physique du silicium (D'Agostino *et al.*, 2024). Côté architectures et composants basés sur les technologies existantes, un résultat marquant est celui du composant NeuroCorgi, qui intègre une implantation matérielle d'un modèle de référence en vision, MobileNetV1, tout en conservant dans une structure programmable les dernières couches spécialisées selon les usages (Miro Panades *et al.*, 2024). Le modèle implémenté a d'abord été optimisé et déployé par la plateforme AIDGE évoquée ci-avant. Ce composant permet un gain de consommation d'un facteur 1 000 vis-à-vis des composants du marché.

L'opportunité des modèles de fondation et de l'IA générative

Au-delà des usages potentiels de l'IA générative pour les services supports qui se retrouvent dans toutes les structures privées ou publiques, les modèles de fondation et

⁸ <https://deepgreen.ai/>

⁹ <https://pyrat-analyzer.com/>

¹⁰ <https://sites.google.com/view/vnn2024>

¹¹ <https://projects.eclipse.org/projects/technology.aidge>

l'IA générative sont des sujets qui ont aussi amené à élargir les champs de recherche existants. Cela intervient de manière évidente pour le traitement automatique des langues, et la vision, mais aussi de manière croissante sur d'autres activités de recherche au cœur du CEA comme les matériaux, la robotique, l'ingénierie logicielle ou même l'astrophysique. Dans ces usages, une attention particulière est mise sur la souveraineté et l'impact sociétal et en ressources.

Un exemple particulièrement représentatif de ce type de travaux est celui proposé par le challenge du ministère de la Défense, EvallLM 2024¹² : évaluer les performances de très grands modèles de langue (LLM) pour la recherche d'informations d'intérêt dans un domaine métier spécifique (veille dans le domaine militaire à partir de documents en langue française). La difficulté réside dans le paramétrage des requêtes envoyées au modèle généraliste avec les informations de contexte (sujets d'intérêt) et la spécialisation avec un très faible nombre d'exemple annotés, mal équilibrés sur les différents sujets (on parle de *few-shot learning*). Le temps de préparation donné au participant était très court et la période de test pour la mesure de performance limitée à trois jours. En pratique, seules des technologies existantes et déjà matures pouvaient contribuer.

L'approche retenue par les équipes CEA a consisté à se baser sur un « petit » modèle de langue (GLiNER de 460 millions de paramètres), de le spécialiser pour le contexte, et de compléter le faible jeu de données de spécialisation par la synthèse automatique de données complémentaires (Armingaud *et al.*, 2024). Parmi les différentes propositions, certaines étaient basées sur les plus grands LLM du marché (par exemple GPT-4o, 200 milliards de paramètres). L'évaluation a placé la solution CEA en première position. Ce résultat démontre, d'une part, l'intérêt de modèles de langue frugaux avec des démarches de spécialisation, et, d'autre part, la capacité à les réaliser très rapidement tout en assurant une efficacité au moins équivalente aux plus grands modèles généralistes.

Dans cette logique d'intégration des principes des modèles de fondation et de l'IA générative pour des domaines spécifiques et leur spécialisation, un partenariat a été mis en place avec le laboratoire d'IA de Thales, CortexAI, visant des solutions avancées pour la défense¹³.

Les applications aux activités scientifiques du CEA

L'IA est entrée comme technologie d'accélération et d'innovation dans l'ensemble des grands programmes de recherche du CEA : tout particulièrement pour la santé, l'énergie nucléaire et renouvelable, la conception de nouveaux matériaux et l'analyse de masses de données scientifiques issues des grands instruments dans les domaines de la physique fondamentale, du climat et de l'astrophysique. Un cas d'usage dans ce

domaine est particulièrement emblématique par son ampleur et l'ensemble des enjeux technologiques liés à l'usage de l'IA. Il s'agit du SKA¹⁴, "Square Kilometer Array", qui vise à la construction du plus grand observatoire radio-astronomique du monde, et qui sera le support d'un ensemble très large de problématiques de l'astrophysique.



Figure 4 : Un ensemble de questions scientifiques très larges : des systèmes planétaires, aux origines et à l'évolution de l'univers (Crédit SKAO).

Il regroupe des enjeux d'IA aussi bien matériels que logiciels, allant du filtrage très haute vitesse de masses de données embarqué sur les capteurs, à l'analyse de masses de données distribuées sur une fédération de centres de stockage et de calcul répartis sur la planète. Le démarrage est visé en 2028, pour une durée de 50 ans.

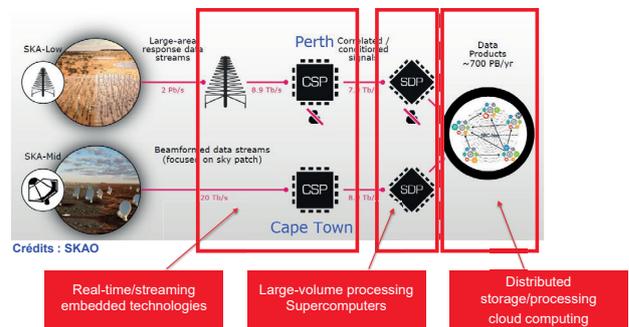


Figure 5 : Un observatoire astronomique à l'heure du big data. Le SKA produira près de 300 Pbytes/an de données réduites (Crédits SKAO).

Dans le cadre du SKA, les utilisations de l'IA sont multiples. La première consiste en la réduction des données au plus proche des antennes du radio-interféromètre avec des contraintes d'embarquabilité particulièrement dures : traitement en temps réel et en continu de données massives pour une réduction des données d'un facteur 1 000 (de 2 Pb/s à 8 Tb/s) ; et une exigence de frugalité et de réduction de l'impact énergétique. Compte tenu du volume et de la complexité des données, l'IA jouera aussi un rôle central dans leur traitement en vue de leur exploitation scientifique. Les technologies mises en œuvre seront basées sur l'apprentissage machine, du *clustering* de données à

¹² <https://evallm2024.sciencesconf.org/>

¹³ https://www.thalesgroup.com/fr/monde/defence-and-security/press_release/thales-et-le-cea-signent-partenariat-lintelligence

¹⁴ <https://www.skao.int/en>

l'apprentissage profond pouvant intégrer les techniques les plus avancées d'IA génératives. Elles s'appuient sur le développement de modèles spécialisés, et sur la construction et l'usage de grands modèles de fondation pour les sciences.

Du point de vue matériel, il sera fait appel à des solutions dédiées (FPGA¹⁵) et de composants d'accélération d'IA avancées pour répondre à l'énorme hétérogénéité des besoins. Tous ces usages sont confrontés à un ensemble de contraintes de qualité qui recoupent les enjeux de confiance et de sûreté évoqués précédemment et qui sont essentiels à l'exploitation de données scientifiques : solutions rapides, précises et capables de traiter de très grands volumes de données ; solutions interprétables avec quantification de l'incertitude et un contrôle des biais ; solutions d'IA de haute précision basées sur la physique.

Observatoire astronomique de référence pour les 50 prochaines années, le SKA est un projet de très grande envergure tant pour les questions scientifiques fondamentales auxquelles il apportera des réponses, que pour les défis technologiques qu'il présente. L'IA y jouera un rôle de premier plan, à tous les niveaux de la chaîne de traitement, du traitement en ligne à l'exploitation scientifique, et à toutes les échelles, des modèles frugaux aux grands modèles de fondation.

L'IA au service de l'industrie

Au-delà des partenariats industriels de développements spécifiques avec des grands groupes ou des PME (comme Thales, Valeo, InGroup, Orano, Siemens, Arcure, AIHeard), la démarche du CEA capitalise également les résultats génériques sur des briques outils et des plateformes ouvertes (par exemple : l'annotation de données vidéo¹⁶ ; l'intégration d'outils de validation formelle¹⁷ ; le raisonnement symbolique¹⁸). Toutefois, l'accès à ces technologies demeure complexe pour la plupart des PME, c'est pourquoi le CEA a mis en place la plateforme PRISM, nœud français du projet européen de plateforme d'expérimentation et de test de l'IA pour le *manufacturing*¹⁹. Elle permet d'accompagner les entreprises sur le sujet, mais aussi plus largement sur l'ensemble des enjeux du numérique : du *cloud* à l'embarqué, de la collecte des données, de leurs échanges, et de leur usage en simulation (*via* des jumeaux numériques) et avec l'intelligence artificielle.



Figure 6 : Vers une continuité numérique : penser large (Source : CEA List).

Références

- ADEDJOUMA M. *et al.* (2024), "Defining operational design domain for autonomous systems: A domain-agnostic and risk-based approach", SoSE.
- ARMINGAUD R. *et al.* (2024), « CEA-List@EvalLLM2024 : prompter un très grand modèle de langue ou affiner un plus petit ? », EvalLLM2024.
- CHIHANI Z. (2021), "Formal methods for AI: Lessons from the past, promises of the future", CAID.
- D'AGOSTINO S. *et al.* (2024), "DenRAM: Neuromorphic dendritic architecture with RRAM for efficient temporal processing with delays", *Nature Commun.*
- GABREAU C. *et al.* (2024), "A study of an ACAS-Xu exact implementation using ED-324/ARP6983", ERTS.
- MIRO-PANADES I. *et al.* (2024), "A 772µJ/frame ImageNet feature extractor accelerator on HD images at 30FPS", IEEE APCCAS.
- NG A. (2017), Artificial intelligence is the new electricity, Presentation at the Stanford MSx future forum.
- OWEN M. P. *et al.* (2019), "ACAS Xu: Integrated collision avoidance and detect and avoid capability for UAS", IEEE/AIAA DASC.

¹⁵ "Field-programmable gate array" : réseau de portes programmables *in situ*. Les systèmes combinant processeur et FPGA offrent une bande passante de données extrêmement élevée. Les FPGA sont très utilisés pour les applications intensives en calcul : puissants et avec des délais d'exécution rapides.

¹⁶ <https://pixano.cea.fr/>

¹⁷ <https://www.caisar-platform.com/>

¹⁸ <https://expressif.cea.fr/>

¹⁹ <https://ai-matters.eu/>

Les enjeux d'une IA responsable

Par **Brigitte D'ANDREA NOVEL**

Professeure Mines Paris - PSL, directrice de l'Institut des Transformations Numériques

David RYCKELYNCK

Professeur Mines Paris - PSL, Ingénierie Digitale des Systèmes Complexes

et **Georges-André SILBER**

Enseignant-chercheur Mines Paris - PSL, centre de recherche en informatique, porteur de la chaire « logiciel responsable »

En 2022, avant l'explosion de l'IA générative, le numérique représentait déjà 4,4 % de l'empreinte carbone annuelle de la France et 11 % de sa consommation électrique, soit 51,5 TWh, ou 65 TWh si on prend en compte la consommation électrique des centres de données situés à l'étranger. Parallèlement, le développement des ENR, des réseaux intelligents, des villes intelligentes, des jumeaux numériques, implique une utilisation toujours plus importante de l'IA, et plus particulièrement des méthodes d'apprentissage. Comment résoudre cette dualité ?

Les chercheurs au sein de l'Institut des Transformations Numériques de Mines Paris explorent et développent des pratiques innovantes pour un numérique plus respectueux de l'environnement et des ressources, au moyen de différents leviers, à savoir : une utilisation raisonnée de la donnée, la pertinence des réseaux basés sur la connaissance a priori, la prise en compte des contraintes matérielles / logicielles liées à l'implémentation et la reproductibilité des résultats.

Vers une utilisation raisonnée de la donnée

Nous nous intéressons dans cet article aux domaines où l'on dispose à la fois de données et de connaissances sur ce qui a produit ces données, et plus particulièrement au domaine de l'ingénierie scientifique. De façon concise, l'ingénierie scientifique se caractérise par l'utilisation de modèles mathématiques intégrant plus ou moins de connaissances *a priori*. Une approche frugale du traitement des données revient à mettre en œuvre le principe que toutes les données ne sont pas informatives. En effet, la taille des données, en nombre d'instances ou en volume par instance, a un impact au premier ordre sur les ressources informatiques exploitées pour leur stockage et leur traitement. Cette frugalité a été bien comprise dans le développement de l'IA pour l'industrie manufacturière, par exemple dans Martinez-Arellano G. et Ratchev S. (2024).

Un ensemble de données cache en lui l'information de ce qui est important pour sa représentation simplifiée. On parle d'espace latent. Par exemple, si un ensemble de vecteurs 3D a systématiquement des instances de vecteurs avec une composante nulle et que c'est toujours la même composante qui est nulle, alors une représentation 2D des vecteurs est plus appropriée. Les algorithmes les plus avancés

pour trouver une représentation latente des données ne nécessitent aucune annotation, ni intervention humaine. Ils peuvent donc traiter de très gros volumes si besoin. Ces algorithmes sont dits auto-supervisés. On les retrouve dans des tâches de compression de données, mais aussi dans les Large Language Models (LLM) et les modèles génératifs les plus récents. L'algorithme le plus simple, bien antérieur à l'IA, est celui de la décomposition en valeurs singulières qui extrait d'une matrice de données (série d'instances de vecteurs) le rang de celle-ci et sa représentation simplifiée. Mais cette décomposition n'est adaptée qu'à certains espaces latents (sous-espaces vectoriels).

Ces dernières décennies les images et les vidéos d'objets techniques se sont généralisées pour diverses tâches de contrôle, d'analyse ou de prévision, dans le cadre de la conception, la production ou l'exploitation de systèmes ou de services. On peut citer par exemple le contrôle qualité par imagerie d'aubes fan de turbo-réacteurs au moment de leur fabrication par un célèbre motoriste français du secteur aéronautique. Mais bien au-delà du format commun des images en deux dimensions d'objets réels, l'ingénieur scientifique exploite des formats de données variés pour numériser des objets réels ou virtuels, mais aussi pour réaliser des prévisions. Par exemple, l'image d'une prévision météo telle que l'on en voit dans les journaux d'information est bien plus

qu'une simple image en couleur puisqu'elle contient des informations calculées par des modèles très sophistiqués. D'ailleurs, la comparaison des prévisions à ce qui a été réellement observé permet de calibrer les modèles physiques sous-jacents aux prévisions.

Confondre la taille du verre et le volume d'eau est vraiment sans conséquence. Mais ignorer que les données cachent une information latente de petite dimension, c'est un véritable frein au déploiement du numérique dans l'ingénierie scientifique. L'argument de dire que nous n'avons pas les moyens financiers ni informatiques pour l'analyse des données massives par des algorithmes d'apprentissage est l'argument de ceux qui regardent le volume du verre et non le volume d'eau dans l'illustration ci-dessus. Beaucoup de ceux qui ont des grandes capacités de calcul et de stockage des données nous poussent à cette erreur d'analyse.

Un autre frein au déploiement d'algorithmes d'apprentissage en ingénierie scientifique est la confidentialité des données et leur manque de variété. On peut faire une analogie très utile avec le domaine de la médecine, où la rigueur scientifique du diagnostic doit s'appliquer à différents types de données et d'images, avec des bases de données plutôt éparses. Des avancées technologiques pour un apprentissage fédéré ont été réalisées dans ce domaine afin que différents établissements hospitaliers puissent partager un algorithme commun, de façon collaborative, sans jamais partager de données (Owkin, 2021). Ces algorithmes sont encore trop méconnus en ingénierie scientifique au sens large.

Architecture des réseaux

Les réseaux neuronaux physiquement informés (PINN : Physics-Informed Neural Networks) sont un type d'approximateurs de fonctions universelles qui peuvent intégrer dans le processus d'apprentissage la connaissance des lois physiques sous-jacentes permettant de modéliser le comportement des données, notamment par des équations différentielles ordinaires EDO ou des équations aux dérivées partielles EDP (Raissi *et al.*, 2019, 2024). La faible disponibilité des données pour la résolution de certains problèmes, en biologie par exemple, limite la robustesse des modèles conventionnels d'apprentissage automatique utilisés pour ces applications. Une connaissance *a priori* des lois physiques prend donc tout son intérêt lors de l'apprentissage des réseaux neuronaux en limitant l'espace des solutions admissibles, qui doivent respecter les contraintes imposées par la physique, comme en mécanique des fluides les lois de conservation de masse, de volume, de quantité de mouvement (Bastin *et al.*, 2008 ; Michel *et al.*, 2024), permettant ainsi à l'algorithme d'apprentissage de trouver la bonne solution même avec une faible quantité d'exemples d'apprentissage. Les PINN permettent d'aborder un large éventail de problèmes, et représentent un moyen de développer de nouvelles classes de solveurs numériques pour les EDP, alternatives aux approches traditionnelles par éléments finis, analyse modale, itératives, multi-grilles, etc. (cf. par exemple Bouche et

Weens, 2024), et sont à même de résoudre les problèmes d'inversion de modèles et d'identification de systèmes à partir des données (Raissi *et al.*, 2019).

L'avantage de ce type d'approche par réseau neuronal est que le respect de la dynamique de l'EDP, des conditions aux bords et initiales est garanti par la minimisation de la fonction de perte qui encode à la fois les résidus de l'EDP et ces conditions. Ces réseaux s'adaptent à des EDP non linéaires ou couplées, sur des domaines complexes ou irréguliers ne nécessitant pas de maillage explicite. Ils sont également bien adaptés pour résoudre des problèmes en dimensions élevées, où les méthodes traditionnelles deviennent coûteuses (comme en mécanique quantique ou en dynamique des fluides). Ils peuvent également être renseignés par des données expérimentales lorsque certaines conditions initiales sont inconnues. Les PINN bénéficient de l'utilisation d'outils modernes de différentiation automatique pour calculer efficacement les gradients lors de la phase d'optimisation des paramètres du réseau, et s'implémentent efficacement sur des GPU (Graphics Processing Unit pour processeurs graphiques) parallèles.

Les réseaux neuronaux informés par la physique (PINN) constituent une approche prometteuse qui combine la puissance des réseaux neuronaux et l'interprétabilité de la modélisation physique. Les PINN ont montré de bonnes performances pratiques dans la résolution des équations différentielles partielles et dans les scénarios de modélisation hybride, où les modèles physiques améliorent les approches basées sur les données. Cependant, même si les PINN constituent une approche prometteuse, il est essentiel d'établir leurs propriétés théoriques afin de comprendre pleinement leurs capacités et leurs limites en termes de précision et robustesse, notamment dans le cas de problèmes raides¹ et mal conditionnés (gradients évanescents²), on pourra consulter par exemple Doumeche *et al.* (2023) où les auteurs font appel à des outils de l'analyse fonctionnelle et du calcul des variations, permettant dans le cas d'EDP linéaires une régularisation de type Sobolev implémentable, permettant de reconstruire une solution qui non seulement atteint la précision statistique mais maintient également la cohérence avec la physique sous-jacente, validant ainsi les résultats obtenus par l'IA.

Logiciels d'IA responsables

Les avancées scientifiques impressionnantes des dix dernières années dans le domaine de l'apprentissage automatique, notamment l'essor des grands modèles de langues (LLM) dans leurs variantes génératives, offrent des perspectives de progrès dans de nombreux domaines. Malheureusement, ces avancées scientifiques

¹ Une équation différentielle ordinaire est dite raide si les méthodes numériques usuelles pour ses solutions nécessitent un pas anormalement petit.

² L'évanescence du gradient est le problème posé, dans un réseau de neurones profond, par la diminution très rapide des valeurs des gradients pendant la rétropropagation, entraînant l'annulation du gradient et par conséquent l'arrêt de l'apprentissage.

et technologiques aggravent deux aspects néfastes de l'utilisation croissante du numérique : son empreinte environnementale et la mainmise exclusive d'entités privées sur ces nouvelles technologies.

L'entraînement des LLM nécessite une puissance de calcul gigantesque, fournie par des grappes de plusieurs milliers de machines équipées de plusieurs processeurs et GPU, ainsi que d'une très grande quantité de mémoire et de stockage. Les réseaux de neurones contenant ces modèles sont entraînés grâce à des opérations d'algèbre linéaire massives dans des espaces vectoriels aux dimensions très grandes. La règle actuelle est que la performance des modèles croît avec la quantité de données et la taille du réseau. Elle engendre une course effrénée à la puissance et à la création de centres de données titanesques.

Une fois un modèle entraîné, son utilisation dans un réseau de neurones (inférence) consomme encore plus d'énergie, car en plus de solliciter des centres de données, elle devient ubiquitaire sur de nombreux appareils. Sur tous nos smartphones et ordinateurs, avec des modèles plus petits dérivés de ceux de grande taille, l'IA est utilisée pour améliorer les applications, notamment de bureautique : classement des e-mails, rédaction de documents ou encore retouche d'images.

La construction des centres de données dédiés à l'IA nécessite de nombreux matériaux, pour l'infrastructure, l'approvisionnement électrique, le refroidissement, mais aussi pour la construction des ordinateurs et équipements réseaux. L'utilisation de l'IA sur nos machines personnelles nécessite également un renouvellement rapide de nos ordinateurs et smartphones. Ainsi, la création de modèles d'IA et leur utilisation ont un impact environnemental très important, que ce soit en termes de consommation d'énergie, d'émission de GES, de consommation de ressources minérales et d'eau.

En plus de la puissance de calcul, un LLM a besoin d'une très grande masse de données pour son entraînement : les entreprises créant de tels modèles utilisent quasiment toutes les données textuelles disponibles sur Internet. En dehors des projets académiques produisant des modèles moins ambitieux, aucune entreprise ne fournit son jeu de données d'entraînement. Certaines entreprises comme Meta ou Mistral AI publient les poids de leurs réseaux en *open data*, contrairement à d'autres comme OpenAI. Enfin, l'architecture des réseaux, les hyper-paramètres utilisés ou la description du processus d'apprentissage et de raffinement ne sont pas toujours publiés. Cette fermeture, aggravée par le fait que la recherche académique publique n'a pas accès aux mêmes moyens de calcul que la recherche privée américaine, pose un problème de reproductibilité des résultats. Par ailleurs, d'un point de vue sociétal, le fait qu'une technologie majeure soit uniquement aux mains d'entités privées pose des questions démocratiques importantes.

Au sein de l'Institut des Transformations Numériques et plus particulièrement dans sa chaire « logiciel responsable », nous menons des recherches afin d'améliorer l'utilisation du matériel par le logiciel, pour qu'il

consomme moins d'énergie, de mémoire, qu'il soit plus rapide, et qu'en conséquence, il nécessite moins de renouvellement du matériel. Les logiciels que nous étudions et améliorons sont des logiciels libres fondamentaux : compilateurs, systèmes d'exploitation, bibliothèques de base. Ils sont au cœur de nos infrastructures et sont fondamentaux pour la transition numérique.

Pour conclure, il est urgent de renforcer la recherche scientifique académique publique en IA, afin d'explorer de nouvelles méthodes d'apprentissage utilisant mieux les données, les connaissances *a priori*, et consommant moins d'énergie (la puissance d'un cerveau humain varie entre 20 et 40 W). Cette recherche doit se faire dans un mode ouvert – poids, données, logiciels –, nécessaire à la reproductibilité et à la compréhension des phénomènes mathématiques à l'œuvre, encore mal compris.

Bibliographie

- BASTIN G., CORON J. M. & d'ANDRÉA-NOVEL B. (2008), "Boundary feedback control and Lyapunov stability analysis for physical networks of 2×2 hyperbolic balance laws", *Proc. 47th Conference on Decision and Control*, Cancun, Dec. 2008.
- BOUCHE D. & WEENS W. (2024), « Analyse quantitative des schémas numériques pour les équations aux dérivées partielles », EDP Sciences, collection PROFIL, p. 248.
- DOUMECHE N., BIAU G. & BOYER C. (2023), "Convergence and error analysis of PINNs". arXiv:2305.01240 [math.ST].
- MARTINEZ-ARELLANO G. & RATCHEV S. (2024), "Towards frugal industrial AI: A framework for the development of scalable and robust machine learning models in the shop floor", *The International Journal of Advanced Manufacturing Technology*, pp. 1-23.
- MICHEL T., ANSALDI A., VIQUERAT J., MELIGA P. & HACHEM E. (2024), "Combining machine learning and computational fluid dynamics for solar panel tilt angle optimization in extreme winds", *Physics of Fluids*, 36(12).
- RAISSI M., PERDIKARIS P., KARNIADAKIS G.E. (2019), "Physics informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations", *Journal of Computational Physics*, n°378, pp. 686-707.
- OWKIN (2021), Federated learning in healthcare: Building trust through traceability.
- BRILLAND T., FANGEAT E., MEYER J. & WELLHOF M. (2025), « Évaluation de l'impact environnemental du numérique en France », mise à jour de l'étude, ADEME-Arcep, janvier 2025.

La stratégie chinoise en matière d'IA

Par Paul JOLIE

Conseil général de l'Économie

Les États-Unis sont depuis longtemps à la pointe de l'IA, grâce à une recherche de pointe et à un secteur privé fort, mais l'approche implacable de la Chine, impulsée par l'État, réduit rapidement l'écart.

En 2023, il y avait 4 482 entreprises d'intelligence artificielle en Chine, selon le recensement de l'Académie chinoise des technologies de l'information et de la communication.

En mars 2024, la Chine comptait 52 entreprises licornes dans le secteur de l'IA, dont 13 nouvelles. Le rapport 2023 sur l'indice mondial d'innovation en IA montre que la Chine se classe au premier rang mondial, représentant 36,7 % des principaux articles sur l'IA et 34,7 % des brevets d'IA accordés parmi les principaux pays, démontrant ainsi une forte dynamique d'innovation en IA.

L'un des aspects les plus frappants de la stratégie chinoise en matière d'IA est son écosystème LLM *open source*, qui progresse à un rythme qui a attiré l'attention des communautés mondiales de l'IA.

Introduction

Les États-Unis sont depuis longtemps à la pointe de l'IA, grâce à une recherche de pointe et à un secteur privé fort, mais l'approche implacable de la Chine, impulsée par l'État, réduit rapidement l'écart.

En 2023, il y avait 4 482 entreprises d'intelligence artificielle en Chine, selon le recensement de l'Académie chinoise des technologies de l'information et de la communication.

En mars 2024, la Chine comptait 52 entreprises licornes dans le secteur de l'IA, dont 13 nouvelles.

Le rapport 2023 sur l'indice mondial d'innovation en IA montre que la Chine se classe au premier rang mondial, représentant 36,7 % des principaux articles sur l'IA et 34,7 % des brevets d'IA accordés parmi les principaux pays, démontrant ainsi une forte dynamique d'innovation en IA.

Les débuts de l'IA en Chine

Le développement de l'IA en Chine a commencé dans les années 1970, lorsque des projets liés à l'IA ont été intégrés dans les plans de recherche nationaux. Ces recherches fondamentales ont jeté des bases solides pour les progrès futurs de l'IA en Chine.

En 2016, il y eut un premier effet « Spoutnik » pour la Chine lorsque AlphaGo, un programme de l'entreprise DeepMind (ensuite rachetée par Google), a battu au jeu de Go successivement le champion sud-coréen Lee Sedol et Ke Jie, le meilleur joueur chinois à l'époque. Ce fut une prise de conscience de l'importance de l'intelligence artificielle par le monde politique du pays du Milieu, qui a alors déclenché un ensemble de mesures pour renforcer sa position dans l'IA.

Rappelons que depuis l'instauration en 1949 de la République populaire de Chine, le pays cherche à retrouver sa splendeur passée. Un de ses objectifs politiques est de dépasser les États-Unis d'ici 2049 (centenaire de la proclamation de la RPC), en tant que première puissance mondiale.

Si on voit le « combat » pour le *leadership* mondial comme une partie de Go, l'intelligence artificielle est une des zones de « combat » sur laquelle il est important de prendre position, afin de gagner d'autres terrains en s'appuyant ensuite sur la position acquise. La Chine « joue » différentes positions stratégiques du *goban*, avec par exemple sa stratégie sur les énergies renouvelables.

Ainsi, depuis 2016, la Chine cherche à créer au fil des ans un environnement de développement favorable de l'IA en multipliant les actions de soutien.

Par exemple, en juillet 2017, le Conseil d'État a publié le Plan de développement de l'intelligence artificielle de nouvelle génération, qui définit les orientations du développement de l'IA au niveau de la stratégie nationale, notamment la recherche théorique fondamentale, les percées dans les technologies clés et les applications industrielles, ce qui a suscité un enthousiasme considérable pour les investissements dans l'IA dans divers secteurs.

En décembre 2017, le ministère de l'Industrie et des Technologies de l'information a publié le Plan d'action triennal pour la promotion du développement de l'industrie de l'intelligence artificielle de nouvelle génération (2018-2020).

En juillet 2022, le ministère des Sciences et Technologies, en collaboration avec cinq autres départements, a publié les Avis d'orientation pour favoriser une application généralisée de l'IA, encourageant les entreprises et autres parties prenantes à ouvrir des scénarios d'application, accélérant ainsi l'intégration de l'IA dans divers secteurs.

En mars 2024, le rapport d'activité du gouvernement chinois a présenté pour la première fois l'initiative « AI+ ». L'objectif de l'initiative étant de promouvoir la transformation des industries traditionnelles et l'amélioration de la qualité de vie des citoyens grâce à l'IA.

En parallèle, la Chine a cherché à renforcer ses infrastructures d'IA.

En septembre 2024, la capacité de calcul totale en Chine était de 246 EFLOPS, ce qui la plaçait parmi les *leaders* mondiaux. Poussée par les applications prometteuses de la technologie de l'IA, la Chine a connu une augmentation importante de la construction de centres de calcul intelligents.

Cela est complété par la poursuite agressive par la Chine d'une stratégie nationale de données conçue pour améliorer la disponibilité et la qualité des données, ce qui est crucial pour la formation de modèles d'IA avancés.

Les avancées récentes de l'IA en Chine

Mais c'est surtout le deuxième effet « Spoutnik » avec l'émergence de ChatGPT fin 2022 qui a dopé le développement coordonné de grands modèles à usage général et verticaux, en créant un écosystème dynamique pour l'innovation industrielle.

En 2023, les entreprises chinoises d'IA ont accéléré l'application commerciale de grands modèles, qui comprennent à la fois des grands modèles fondamentaux dotés de capacités générales et des grands modèles spécifiques à l'industrie.

Parmi les premiers, on peut citer ERNIE Bot de Baidu, Tongyi Qianwen (Qwen) d'Alibaba, Zhipu Qingyan de Zhipu AI et Doubao de ByteDance, qui représentent tous l'exploration de la Chine à la pointe de la technologie.

Parmi les seconds, on peut citer Pangu de Huawei, qui sert pour le modèle météorologique, le modèle de vagues océaniques et le modèle pour l'industrie financière, entre autres...

En octobre 2024, près de 200 grands modèles d'IA génératifs étaient accessibles au public en Chine. Les entreprises qui publient ces modèles comprennent non seulement des entreprises d'IA de premier plan, mais aussi de nouvelles licornes émergeant de cette vague d'innovation de grands modèles, telles que Zhipu AI, Moonshot AI et Minimax. Ces entreprises, souvent dirigées par des anciens de Tsinghua, sont à l'avant-garde du développement de grands modèles linguistiques (LLM), dont certains surpassent désormais leurs homologues américains dans les tests de référence bilingues et commencent également à établir des références dans diverses applications.

Et pourtant, le financement des start-ups d'IA en Chine ne représente qu'une fraction de ce que lèvent les start-ups d'IA américaines : la Chine a vu son financement de l'IA atteindre 5,2 milliards de dollars en 2024, soit 7 % des 76,3 milliards de dollars des États-Unis, c'est-à-dire beaucoup moins...

Private foundation model developers valued at \$1B+ based in China (as of 1/27/2025)

Company	Valuation	Notable investors
Moonshot AI	\$3.3B	Alibaba, Tencent, HONGSHAN 红杉中国, Meituan
ZHIPU AI	\$3.0B	Alibaba, Tencent, HONGSHAN 红杉中国, Meituan, Prosperity**
百川智能 BAICHUAN AI	\$2.8B	Alibaba, Tencent, xiaomi
MINIMAX	\$2.5B	Alibaba, Tencent, HONGSHAN 红杉中国
零一万物 01.AI	\$1.0B	Alibaba, 创新工场

Source: CB Insights valuation data. **Previously the China investment arm of Sequoia Capital. **Venture fund of Saudi Aramco. CBINSIGHTS

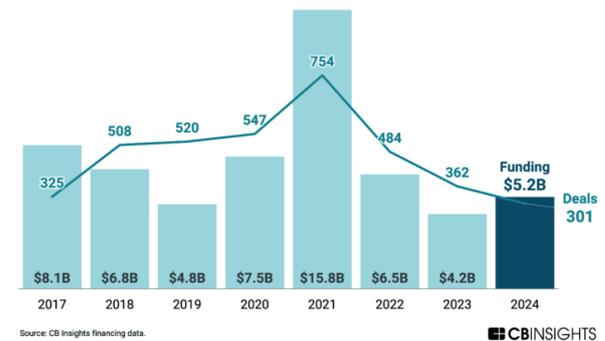
Figure 1 : Les développeurs de modèles de fondations privées évalués à 1 Md \$ et plus, basés en Chine (Source : CBInsight).

	Open	Closed
Biggest public companies*		
Top-funded private companies		
Notable investors**		
Total equity funding since 2020	\$14.9B	\$37.5B

Source: CB Insights data (as of 12/31/2024). *By market cap. **Based on investment activity and public statements. Note: xia (01.1B in equity funding) is classified as open-source based on Elon Musk's stated commitment (October 2024) to open-sourcing its models. CBINSIGHTS

Figure 2 : Les fonds propres investis depuis 2020 pour les entreprises les plus importantes aux modèles ouverts et aux modèles fermés (Source : CBInsight).

Disclosed equity funding and deal count to China-based AI startups



Source: CB Insights financing data. CBINSIGHTS

Figure 3 : Les fonds propres et le nombre de transactions pour les start-ups d'IA basées en Chine (Source : CBInsight).

En janvier 2025, ironie de l'histoire, il y a eu un troisième moment « Spoutnik », mais inversé ! Les États-Unis ont pris conscience que la Chine pouvait les concurrencer, voire les dépasser. NVIDIA a chuté en une seule séance le 27 janvier de 17 %, perdant 590 milliards de dollars de capitalisation ! L'entreprise chinoise DeepSeek a en effet bouleversé les hypothèses sur ce qu'il faut pour développer des modèles d'IA puissants. La société a publié un modèle de raisonnement *open source* (nommé R1) qui rivalise avec les performances du modèle de raisonnement o1 d'OpenAI. DeepSeek affirme avoir entraîné son modèle de base avec des puces limitées et environ 5,6 millions de dollars de puissance de calcul – une fraction des 100 millions de

dollars et plus que ses rivaux américains ont dépensés pour entraîner des modèles similaires – grâce à des techniques astucieuses¹.

Une vision orientée applications – en se concentrant sur certaines industries et activités d'avenir

En Chine, la technologie de l'IA est désormais largement appliquée dans l'agriculture, la fabrication et les services. Cela est très important pour obtenir des modèles d'affaires qui permettent au bout du compte le retour sur investissement des travaux de R&D réalisés.

Par exemple, le grand modèle « Xiaotian » du groupe Yimutian a appris des données sur la distribution des produits agricoles dans plus de 2 800 comtés. Il couvre des modules tels que les nouvelles variétés, les nouvelles technologies, les tendances de l'offre et de la demande et la mise en correspondance intelligente de la production et des ventes.

Dans le secteur manufacturier, les entreprises chinoises intègrent la technologie de l'IA à différentes étapes (R&D, conception, production, opérations et gestion, services de produits). Par exemple, les usines intelligentes de Haier utilisent la technologie de vision artificielle pour les inspections de qualité des produits, réduisant ainsi les erreurs.

Dans le secteur des services, la richesse des données d'interaction homme-machine accumulées en Chine à l'ère de l'Internet mobile permet au secteur de mieux servir le public, en fournissant des services plus pratiques et plus sûrs aux entreprises comme aux particuliers.

Dans les services aux consommateurs, par exemple, le secteur de la santé utilise des technologies d'IA comme Tencent AIMIS pour pallier les pénuries de ressources médicales, permettant le diagnostic à distance, la téléconsultation et l'assistance au diagnostic, améliorant de fait la qualité et l'accessibilité des services de santé.

Dans les services de production, le secteur financier illustre l'intégration de l'IA. Ant Group s'appuie sur son grand modèle fondateur pour développer un modèle financier personnalisé.

Sans parler de ce qui nous attend dans divers secteurs de l'industrie, par exemple l'automobile...

Les enjeux autour de l'éthique

La Chine met l'accent sur l'éthique comme priorité dans la gouvernance de l'IA. Dès septembre 2021, le Comité professionnel national chinois sur la gouvernance de l'IA de nouvelle génération a publié les Normes éthiques pour l'intelligence artificielle de nouvelle génération, visant à intégrer l'éthique dans l'ensemble du cycle de vie de l'IA ; et en mars 2022, le Bureau général

du Conseil d'État a publié l'Avis sur le renforcement de l'éthique et de la gouvernance dans les sciences et les technologies, classant l'IA aux côtés des sciences de la vie et de la médecine comme domaines prioritaires.

La Chine accorde une grande importance au développement d'outils de gouvernance pour garantir la sécurité et la fiabilité des produits d'IA.

L'industrie et le monde universitaire explorent également activement le développement d'outils de gouvernance. Ainsi Alibaba, en collaboration avec l'université Tsinghua, l'Université du Zhejiang et d'autres institutions, a achevé le projet *open source* EasyRobust, qui améliore efficacement la robustesse des modèles. Ce projet a reçu le prix IEEE Open Source Science Award 2023.

En décembre 2021, pour résoudre les problèmes découlant des algorithmes de recommandation, la Chine a exigé le dépôt des algorithmes pour améliorer la transparence. Ce dépôt favorise la transparence des algorithmes pour le public afin d'améliorer la surveillance générale... Chacun se fera une opinion sur le sens de cette orientation...

La Chine a également publié en juillet 2023 la première législation au monde spécifiquement sur l'IA générative (pour l'instant en « mesures provisoires »), envoyant un signal positif pour la croissance réglementée de l'IA générative.

En outre, des lois telles que la loi sur le commerce électronique, la loi sur la protection des informations personnelles et la loi anti-monopole ont été mises à jour pour inclure des dispositions relatives aux algorithmes d'IA dans leurs champs d'application réglementaires respectifs, formant dès lors un cadre juridique complet pour l'IA en Chine.

On l'aura compris, la Chine cherche à accélérer l'élaboration de normes et de lignes directrices pour promouvoir la mise en œuvre de règles et de réglementations. Elle cherche à être la plus rapide pour proposer (imposer ?) sa vision². Dès juillet 2020, l'Administration nationale de normalisation et quatre autres départements ont publié les Lignes directrices pour la construction d'un système national de normes d'intelligence artificielle de nouvelle génération. Et en juin 2024, le ministère de l'Industrie et des Technologies de l'information a publié les Lignes directrices pour la construction d'un système de normalisation complet pour l'industrie nationale de l'intelligence artificielle (2024). Le document couvre tous les segments de la chaîne industrielle de l'IA et fournit un soutien de normalisation complet pour un développement industriel de haute qualité.

Pour exporter sa vision, l'industrie chinoise participe activement à l'élaboration de normes internationales. Citons à titre d'exemple iFLYTEK.

De toute évidence, la Chine veut participer activement au développement du système mondial de gouvernance de l'IA. Dès octobre 2023, le président chinois

¹ Ce qui tend à relativiser au moins partiellement les comparaisons en flux d'investissements : le rapport AI Index de 2024, <https://aiindex.stanford.edu/report/> mentionne 67,2 G\$ investis par les États-Unis en IA en 2023 contre 7,76 G\$ en Chine. Ce rapport note aussi qu'en 2023, 61 modèles d'IA notables émanaient des États-Unis contre 21 dans l'Union européenne et 15 en Chine.

² À titre de comparaison en 2023, les États-Unis ont édicté 25 réglementations sur l'IA contre une seule en 2016. Ce chiffre en 2023 était en croissance de 56,3 % par rapport à 2022. Toutefois, un contre-mouvement s'est fait jour avec le changement gouvernemental de janvier 2025. L'Europe, quant à elle, a édicté le règlement sur l'IA en 2024.

Xi Jinping a proposé l'Initiative de gouvernance mondiale de l'IA. Elle dit adhérer à une approche centrée sur l'humain et l'IA pour le bien, où le développement de l'IA doit viser à améliorer le bien-être humain, avec la sécurité sociale et le respect des droits de l'homme comme conditions préalables. La gouvernance de l'IA doit être conforme aux lois internationales applicables, s'aligner sur les valeurs humaines partagées et travailler collectivement pour prévenir et combattre l'utilisation abusive de la technologie de l'IA.

Toujours dans une logique de rattrapage vis-à-vis des avancées américaines qui développent plutôt des systèmes fermés, la Chine appelle à un environnement ouvert, inclusif et non discriminatoire pour le développement de l'IA, permettant à tous les pays de partager les avantages de l'IA.

Elle utilise cela comme arme géopolitique en défendant le multilatéralisme, tout en visant à renforcer la représentation et la voix des pays en développement sur les questions d'IA. Elle est même allée jusqu'à plaider la création d'un organe international de gouvernance de l'IA dans le cadre de l'ONU.

En septembre 2024, la Chine a organisé l'atelier sur le renforcement des capacités en matière d'IA, auquel ont participé des représentants de plus de 30 pays, dont la Thaïlande, Singapour, la Malaisie, le Vietnam et l'Indonésie.

Une tension grandissante entre États-Unis et Chine

Les États-Unis ont pris récemment une conscience plus aiguë de la stratégie chinoise sur l'intelligence artificielle. En se rendant compte que les Chinois entraînaient les nouveaux modèles en utilisant les puces américaines NVIDIA, ils ont interdit l'exportation des dernières technologies de puces dès le mois de septembre 2022.

Cela a eu un effet court terme de ralentissement des progrès en Chine, mais le pays a su s'organiser pour mutualiser les composants de dernière génération déjà achetés et aussi mettre en place des circuits d'approvisionnement parallèles.

Malgré les tensions, les deux pays cherchent à coopérer sur les sujets concernant les risques technologiques liés à l'IA, la gouvernance mondiale et d'autres questions d'intérêt commun. De même, le premier dialogue intergouvernemental Chine-États-Unis sur l'intelligence artificielle s'est tenu à Genève, en Suisse, le 14 mai 2024.

La Chine utilise son avance technologique en IA pour fédérer d'autres pays autour d'elle

Le Premier ministre chinois Li Qiang a annoncé que la Chine lancerait une initiative visant à renforcer la capacité technologique de l'ASEAN en matière de développement grâce à l'autonomisation de l'IA.

La Chine a ainsi proposé de construire des plateformes de coopération multilatérale ouvertes, inclusives et également participatives pour l'IA.

Le Centre de développement et de coopération de l'intelligence artificielle Chine-BRICS (Brésil, Russie, Inde, Chine et Afrique du Sud) a été officiellement lancé en juillet 2024, respectant l'esprit d'ouverture, d'inclusion et de coopération gagnant-gagnant des BRICS.

Conclusion

La Chine adopte une approche équilibrée du développement et de la sécurité, en mettant en œuvre des réglementations prudentes, en créant un environnement propice à la croissance, en prônant une coopération mondiale dans la gouvernance de l'IA et en soulignant l'importance de combler le fossé entre les deux domaines.

Les efforts des États-Unis pour ralentir la Chine ne porteront probablement pas leurs fruits à long terme.

Aujourd'hui, la Chine et les États-Unis sont au coude-à-coude en matière de recherche sur l'IA générative, chaque pays contribuant à des milliers d'articles qui repoussent les limites de ce que l'IA peut accomplir.

Espérons que cette compétition permettra de construire un meilleur futur, et que cela ne soit pas le début d'un nouveau type de course à l'armement. L'avenir nous le dira.

S'agissant de l'Europe, le volet régulation est entré en application le 2 février 2025. Le sommet mondial de l'IA organisé à Paris les 10 et 11 février 2025 vise aussi à ce que la France et l'Europe n'apparaissent pas seulement préoccupées de réguler et d'avoir un dialogue international sur le sujet, mais aussi de développer des offres et des usages de façon autonome.

Réussir l'après-Paris 2025 : quelles ambitions pour la France à la suite du Sommet pour l'action sur l'IA ?

Par Arno AMABILE

Conseiller de l'envoyée spéciale pour le Sommet pour l'action sur l'IA, Présidence de la République,
Rapporteur général de la Commission IA

Entre espoir et fatalisme, la France cherche sa place dans la révolution de l'intelligence artificielle. Pour devenir une « puissance de l'IA », elle doit tant développer sa puissance sur l'IA (maîtrise technologique et adoption par la société) que sa puissance par l'IA (poids dans la chaîne de valeur et la géopolitique de l'IA).

Si la France est bien placée pour suivre rapidement le développement de l'IA générale, elle doit encore trouver les secteurs et marchés dans lesquels son écosystème peut tirer parti de son savoir-faire ou de sa localisation, sans se trouver enfermé dans de petits marchés isolés.

Le rapport de la Commission IA en mars 2024, la troisième phase de la stratégie IA en février 2025 et le Sommet pour l'action sur l'IA au même mois concourent à cette ambition, sans illusions sur nos dépendances et notre retard, mais sans renoncement face au potentiel de l'IA pour la France.

Depuis le lancement de ChatGPT en novembre 2022 et la création de Mistral AI début 2023, nous oscillons entre espoir et fatalisme. Espoir que l'intelligence artificielle, révolution technologique ancrée dans les mathématiques, soit l'occasion d'un renouveau économique et géopolitique de la France. Fatalisme, lié à la faiblesse de la France dans l'économie numérique, qui se répèterait dans l'IA.

En mars 2024, la Commission IA, co-présidée par Philippe AGHION et Anne BOUVEROT, a présenté un rapport intitulé « IA : notre ambition pour la France ». Ce document propose 25 mesures et un plan ambitieux visant à consolider les atouts français et à transformer le dynamisme de notre écosystème en succès durable. Les propositions ont conduit à des annonces présidentielles en mai 2024, suivies par l'organisation d'un Sommet pour l'action sur l'IA, à Paris les 10 et 11 février 2025. Lors de ce Sommet, le gouvernement a précisé l'ambition de la France : devenir « une puissance de l'IA ».

On peut comprendre cette ambition de deux manières : la puissance sur l'IA et la puissance par l'IA. D'abord, la France cherche à maîtriser la technologie pour l'adopter selon ses propres termes. Dans un monde qui se diviserait entre producteurs et consommateurs d'IA ("AI givers" et "AI takers"), elle vise clairement le premier groupe. Ensuite, elle souhaite peser suffisamment dans la chaîne de valeur de l'IA pour conserver ou accroître son influence internationale.

La tentation de concentrer nos efforts sur l'IA générale, au détriment des applications, est compréhensible. Après tout, une IA « plus performante que les humains

à la plupart des tâches économiquement valorisées »¹ pourrait notamment planifier des façons de s'améliorer, écrire et tester ce code, puis le déployer. Cette perspective soulève deux craintes majeures : la perte de contrôle sur la technologie et la domination écrasante du pays qui la développerait en premier.

Cette dynamique diffère des précédentes vagues technologiques. Historiquement, inventer une technologie n'assure pas la domination de sa chaîne de valeur. L'exemple du moteur à explosion, inventé en France et en Allemagne, le prouve : ces pays n'ont pas maintenu une domination sans partage sur l'industrie automobile. La maîtrise des technologies complémentaires, des produits et des techniques de production, développées en partie aux États-Unis puis renouvelées au Japon, ont également joué un rôle crucial.

Le caractère exponentiel de l'IA générale semble donc nécessiter toutes nos attentions, si ce n'était les désaccords entre chercheurs et la commoditisation des modèles. Les experts divergent sur la définition même de l'IA générale et son horizon temporel. Aucun modèle n'a su conserver son avance au-delà de quelques mois, tandis que les modèles ouverts ou non américains rattrapent très rapidement les modèles fermés. Les modèles de DeepSeek en sont un exemple éclatant qui n'a rien d'isolé.

Face à ces incertitudes, la course à l'IA générale est à la fois trop incertaine pour former l'unique objectif

¹ Selon la définition d'OpenAI, définition importante puisque atteindre ce niveau bloquerait l'accès de Microsoft aux modèles d'OpenAI.

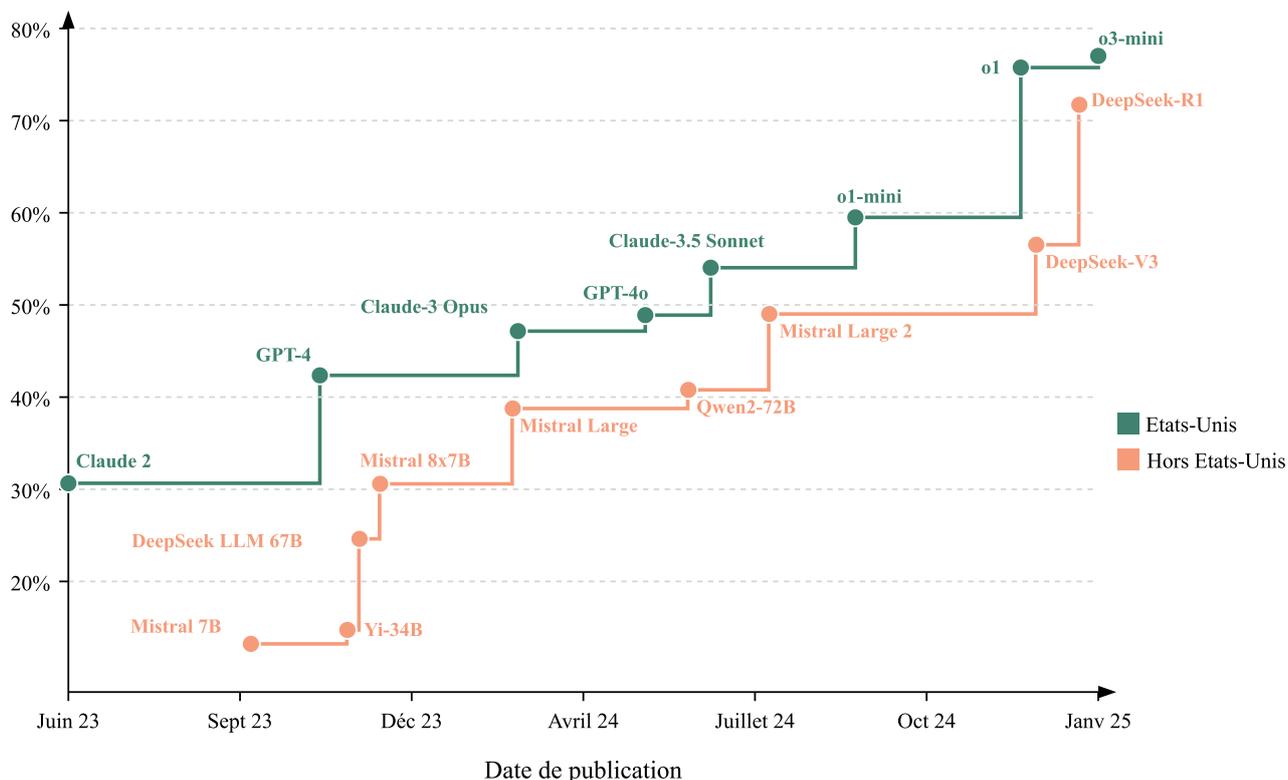


Figure 1 : Les meilleurs modèles américains et fermés ont quelques mois d'avance sur des modèles non américains et ouverts : les modèles sont évalués d'après le test GPQA Diamond, qui teste les connaissances scientifiques (Source : Epoch AI, AI Benchmarking Hub, 2025).

stratégique de la France, et trop importante pour être ignorée. Ici, l'ambition de la France pourrait être de rester proche de la frontière technologique pour maîtriser rapidement ces avancées dès leur émergence.

Cette stratégie de *"fast follower"* est réaliste. La diffusion rapide des innovations dans le domaine le prouve, tout comme le développement de modèles de pointe par des équipes françaises (Bloom, Llama 2 et 3, Mistral Large). La France récolte ici les fruits de son école en mathématiques et informatique ainsi que ceux de la première phase de la stratégie IA faisant suite au rapport Villani, qui a su retenir en France de nombreux talents exceptionnels.

Au-delà du sujet épineux des semiconducteurs, la disponibilité de la puissance de calcul reste le principal obstacle au développement et au fonctionnement de modèles d'IA européen autonomes. Malgré son vaste réseau de supercalculateurs publics, l'Europe ne peut les dédier entièrement à l'entraînement d'un modèle. Or la puissance de calcul privée en Europe ne représente que 4 % de la puissance de calcul mondiale, et son fonctionnement est grevé par des coûts de l'énergie 1,5 à 3 fois plus importants qu'aux États-Unis².

Pour maintenir sa position de *"fast follower"*, la France doit donc agir sur deux leviers : augmenter sa puissance de calcul et consolider son attractivité pour la recherche. La troisième phase de la stratégie IA, annoncée en février 2025, cible précisément ces enjeux.

²DOAN R., LEVY A. & STORCHAN V., (2025), Financer les infrastructures pour une IA européenne compétitive, Groupe d'études géopolitiques

Plusieurs initiatives viennent renforcer la capacité de calcul en France : simplification des procédures d'implantation des *datacenters*, maintien des tarifs préférentiels d'électricité sous conditions environnementales, accès à une énergie décarbonée et stable, construction d'un nouveau supercalculateur public, et investissements privés de 109 milliards d'euros dans les *datacenters* d'ici 2030³. Au vu de la croissance rapide des besoins, les réflexions de l'Union européenne autour de "AI Megafactories" ou les propositions d'un "CERN de l'IA"⁴ sont les bienvenues.

Pour consolider l'attractivité pour les talents en IA, la France déploie des moyens significatifs : 900 millions d'euros pour les IA-clusters via France 2030, maintien du Passeport talents et du French Tech Visa, et concentration des financements sur des projets stratégiques. Attirer plus de talents, alors que la concurrence internationale pour attirer les talents s'exacerbe⁵ et que le système de recherche américain est remis en cause, nécessitera de continuer les efforts sur la rémunération des chercheurs publics et surtout un effort massif sur la réduction de leur charge administrative, recommandation clé de la Commission IA.

³À noter que comme le projet Stargate, ces investissements ne visent pas uniquement, ou même majoritairement, l'entraînement des IA, mais également l'inférence, i.e. le fonctionnement des modèles.

⁴Voir par exemple Artificial Intelligence Is Stuck. Here's How to Move It Forward, New York Times, Juillet 2017 ou l'initiative du CAIRNE, Confederation of Laboratories for Artificial Intelligence Research in Europe.

⁵Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence, Octobre 2024

Malgré ces atouts, une réalité demeure : les États-Unis et la Chine dominent encore largement le développement des modèles d'IA de frontière. 81 % des grands modèles développés entre 2022 et 2024 l'ont été dans l'un de ces deux pays. Même le succès des modèles ouverts dépend largement de Llama, créé par Meta et de jeux de données d'entraînement en partie tirés de dialogues de ChatGPT. La Silicon Valley, avec sa concentration unique de talents, données, capital et puissance de calcul, reste la plus susceptible de développer la première IA générale, sauf à ce qu'elle ne se replie sur elle-même.

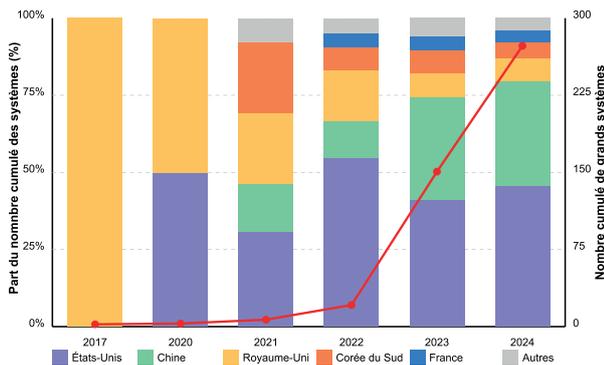


Figure 2 : Nombre cumulé de grands modèles développés par pays : Epoch définit un grand modèle comme requérant plus de 10^{23} opérations en virgule flottante (FLOPS) (Source : Epoch (2024) retraité par Our World in Data).

Une approche alternative nécessitant moins de ressources pourrait certes apporter une percée, rebattant totalement les cartes. Mais l'histoire des avancées en IA généraliste a jusqu'ici plutôt favorisé des paradigmes sachant tirer parti de l'augmentation continue de la quantité de données et de la puissance de calcul⁶, même si pour une tâche donnée, des modèles plus petits peuvent aussi être plus performants et moins consommateurs d'énergie⁷.

Dans une course à l'IA générale dominée par la puissance brute des données et du calcul, par la quantité de financement et l'absence de régulation, la France et l'Europe ont des désavantages structurels. Que l'Europe puisse espérer développer une technologie aussi fondamentale qu'une IA générale quelques mois ou une grosse année seulement après son invention devrait finalement être une source d'espoir plutôt que d'abattement.

L'ambition française ne peut toutefois pas se limiter à cette seule course. Trois raisons l'imposent, déjà citées par la Commission IA : la technologie seule ne garantit pas les bénéfices économiques et sociaux⁸ ; l'IA offre

déjà de nombreuses opportunités avant d'atteindre le stade général ; et notre souveraineté exige que nous déployions l'IA selon nos valeurs, plutôt que de dépendre uniquement de solutions étrangères⁹.

Pour devenir une puissance de l'IA, la France ne peut se limiter à la recherche dans des modèles de frontière, elle doit occuper des positions dans le développement et la distribution de produits intégrant de l'IA. Cette approche est d'autant plus cruciale au vu de la commoditisation actuelle des modèles, qui conduit la valeur à se déplacer vers les deux extrémités de la chaîne : les infrastructures, d'un côté, les applications et leur distribution, de l'autre.

Le retard français et européen tant en amont qu'en aval des modèles est préoccupant. Comme le souligne la Commission IA, malgré l'excellence de nos talents et le dynamisme de notre écosystème, l'Europe souffre d'un déclassement technologique et économique qui menace sa prospérité et son indépendance¹⁰. Alors que les modèles d'IA créent un renouveau de la robotique, la faiblesse de l'Europe dans ce domaine devraient renforcer notre préoccupation.

Les États-Unis accueillent le *cluster* mondial concentrant talents et expertise dans le numérique et l'IA. Que cela soit dû à l'investissement militaire des années 1960, à une culture entrepreneuriale spécifique à la côte Ouest ou à une combinaison comportant une part de chance et d'effet de réseaux¹¹, cette concentration est particulièrement criante dans le développement de produits : nous manquons en France moins de développeurs de modèles que de talents expérimentés dans le développement de produits, l'intégration d'IA dans le hardware ou la distribution à grande échelle.

Toute ambition française nécessite donc à la fois de réduire ce retard structurel tout en capitalisant sur nos forces spécifiques. Ces forces incluent notre expertise en IA frugale et embarquée, notre maîtrise des systèmes de confiance, et notre proximité avec l'industrie européenne, riche en données et en potentiels clients.

Les entreprises françaises, des start-ups comme Mistral AI, H, Helsing, Pleias, Probabl et Potoroom aux groupes établis comme Thalès et Dassault Systèmes, excellent chacune à leur manière à développer des IA de confiance, frugales ou embarquées. Leur succès repose non pas sur la surenchère de performance face aux plus grands modèles, mais sur des solutions ciblées à forte valeur ajoutée – comme l'illustre le partenariat entre Mistral AI et Stellantis pour l'IA embarquée automobile.

Cette différenciation, bien qu'avantageuse, présente un risque : confiner l'écosystème français à des niches de marché où l'adoption est lente et le passage à l'échelle difficile. C'est pourquoi ces entreprises ne se présentent pas comme les leaders d'un sous-segment du marché de l'IA « de confiance, frugale et embarquée »

⁶ SUTTON R. (2019), The Bitter Lesson et COLLINS J. & YOUSEFI M. (2024), Learning the Bitter Lesson: Empirical Evidence from 20 Years of CVPR Proceedings, Oct.

⁷ VAROQUAUX G., LUCCIONNI A. S. & WHITTAKER M. (2024), Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI. et article 4 « Les enjeux d'une IA responsable » dans ce numéro.

⁸ IA : notre ambition pour la France, Commission de l'intelligence artificielle, 2024, p. 8.

⁹ « Nous n'aurons pas la maîtrise de notre avenir par le seul déploiement de l'IA des autres », *Ibid*, p. 10.

¹⁰ *Ibid*, p. 4.

¹¹ KRUGMAN P. (2024), Why does U.S. technology rule? Maybe it's just in the right place, *New York Times*

mais plutôt comme des acteurs aux ambitions internationales produisant des IA performantes dont la proposition de valeur repose sur l'une de ces qualités.

Sans parler uniquement des plus grandes entreprises du numérique dont la valorisation dépasse les milliers de milliards d'euros, la France et l'Europe devraient en effet viser à développer des champions de l'IA, dont la valorisation dépasserait la centaine de milliards d'euros comme SAP, Shopify, Spotify ou Palantir.

Si elle veut devenir une puissance de l'IA, la France doit aider ses entreprises de l'IA à passer à l'échelle, que ce soit en accélérant l'adoption sur le marché domestique, ou en facilitant l'expansion internationale, au-delà même des frontières européennes.

Les initiatives de février 2025, de la troisième phase de la stratégie IA au Sommet pour l'action sur l'IA, mettent ainsi logiquement l'accent sur le déploiement et la diffusion de cette technologie.

Au niveau national, la stratégie complète le renforcement de la puissance de calcul et de l'attractivité pour les talents par trois axes majeurs : formation massive (passer de 40 000 à 100 000 étudiants et professionnels par an) ; sensibilisation et appropriation pour le grand public (deux millions de Français *via* les « Cafés IA » d'ici 2027) ; et transformation de l'État (équipement de tous les ministères en solutions d'IA, y compris commerciales *via* des partenariats technologiques).

Au niveau européen, le Sommet pour l'action sur l'IA a catalysé trois initiatives majeures : l'annonce d'une simplification réglementaire dans le champ de l'IA, à commencer par une mise en œuvre équilibrée du règlement sur l'IA, le lancement du fonds InvestAI de 50 milliards d'euros pour les start-ups, et le lancement du réseau des « Champions européens de l'IA ». Cette dernière initiative, peut-être la plus significative au vu des ambitions françaises, fédère 120 grandes entreprises européennes autour de leur volonté de déployer l'IA pour renforcer leur propre compétitivité, avec des solutions européennes autant que possible. Elle a conduit des investisseurs importants comme KKR, DST Global, EQT ou Blackstone à annoncer leur volonté d'investir 150 milliards d'euros dans l'écosystème européen d'IA sur cinq ans.

Au niveau international, le Sommet poursuivait trois objectifs interconnectés : politique (rassurer sur la capacité de nos sociétés à comprendre et gérer des risques liés à l'IA) ; économique (faciliter le déploiement de cas d'usage concrets de l'IA, notamment pour l'intérêt général) ; et diplomatique (poser les bases d'une coopération internationale). Ces objectifs s'alignent parfaitement avec l'ambition française d'être un pays *leader* d'une IA de confiance, frugale et donc rentable, embarquée.

Politiquement, le Sommet a particulièrement ciblé les préoccupations environnementales et sociales, freins importants à l'adoption de l'IA en Europe. La création de la Coalition pour l'IA durable, avec 96 partenaires dont 39 grandes entreprises, 16 pays et 6 organisations internationales, et l'engagement de plus de 60 entreprises multinationales pour une IA de confiance dans

le monde du travail démontrent une volonté d'action concrète vers une IA plus responsable.

Économiquement, l'accès aux données de qualité reste le défi majeur des entreprises d'IA. Le partenariat CurrentAI, doté d'une première levée de fonds de 400 millions d'euros et du soutien de 10 pays, vise à abaisser les barrières pour les innovateurs de tous les pays et à leur fournir la seule chose qu'ils ne peuvent compenser par leur inventivité : l'accès à des données de qualité.

Diplomatiquement, la demande est claire : les pays cherchent à diversifier leurs fournisseurs d'IA et souhaitent un écosystème plus ouvert et compétitif. Cette attente représente une opportunité d'expansion pour les acteurs français, explicitement encouragés par des pays asiatiques comme latino-américains à être plus ambitieux dans leur développement international.

Naturellement, ces succès et ces annonces ne règlent pas tous les problèmes. On pense en particulier à notre dépendance à une infrastructure ou des composants étrangers ou aux difficultés de déploiement que rencontreront les administrations à cause de l'hébergement imparfait du numérique public. On peut parier toutefois que des grands champions de l'IA seraient les mieux à même de tirer avec eux des fournisseurs de services d'hébergement ou de composants. La dynamique observée à l'occasion du Sommet, tout comme la coopération avec l'Inde qui a co-présidé le Sommet, ont montré que la France pourrait devenir une puissance de l'IA, à condition de renforcer sa maîtrise de cette technologie.

L'histoire des révolutions technologiques le prouve : aucune domination n'est permanente, que ce soit par un pays ou par une entreprise. La France possède les atouts nécessaires pour réussir dans l'IA. Notre seul véritable risque serait de renoncer à cette ambition.

L'IA générative dans l'administration française : le cas d'« Albert »

Par Ulrich TAN

Ingénieur en chef des Mines et chef du DataLab à la direction interministérielle du Numérique (DINUM)

Dans cet article, nous explorons le projet « Albert », l'intelligence artificielle générative développée par la direction interministérielle du Numérique (DINUM). La stratégie sous-jacente émane d'une vision technologique ambitieuse pour l'administration : s'approprier une technologie nouvelle et prometteuse, pour ne pas la subir, et assurer la souveraineté numérique de l'État. Lancé en réponse à l'essor fulgurant de l'IA, Albert vise à fournir des solutions adaptées aux besoins de l'administration française tout en garantissant la confidentialité des données et la puissance de calcul nécessaire. Nous discuterons des trois piliers du projet, ses applications concrètes, et les défis technologiques et éthiques à relever pour une utilisation optimale de l'IA dans le secteur public.

Dans cet article, nous nous intéressons à « Albert », présenté comme l'intelligence artificielle générative souveraine de l'État français. La stratégie sous-jacente émane d'une vision technologique ambitieuse pour l'administration : s'approprier une technologie nouvelle et prometteuse, pour ne pas la subir, et assurer la souveraineté numérique de l'État.

Genèse et contours du projet : maîtriser ses données en plein boom de l'IA

Le projet Albert débute concrètement en juin 2023¹, soit six mois après la sortie de ChatGPT (30 novembre 2022). Le succès sans précédent de l'application d'OpenAI² a accéléré l'engouement général autour de l'IA, et l'administration n'a pas échappé à ce mouvement. Les initiatives se multiplient alors dans la sphère publique pour explorer les promesses de l'IA et en particulier l'IA générative.

Dans ce contexte, Albert a été lancé par le DataLab de la DINUM pour répondre à deux limitations récurrentes : la confidentialité des données et la puissance de calcul. Lorsque vous utilisez un service comme ChatGPT, les données que vous lui envoyez sont en effet captées par les serveurs de l'entreprise américaine, dont vous utilisez également la puissance de calcul. À l'inverse, les solutions déployées par le DataLab dans le cadre d'Albert le sont sur des serveurs maîtrisés par l'administration, ce qui lui impose par là-même de disposer d'une puissance de calcul suffisante.

¹Le projet n'a été nommé « Albert » que lorsqu'il a été dévoilé publiquement le 5 octobre 2023.

²L'application connaît la plus forte croissance jamais enregistrée, atteignant 100 millions d'utilisateurs actifs par mois en moins de deux mois. Source : <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>

La technologie repose sur les grands modèles de langage (ou LLM pour *Large Language Models*) et leur intégration dans des systèmes d'IA (SIA) permettant de générer du texte, ou du code³, à partir d'une instruction formulée en langage naturel (fournie dans une invite textuelle, un « prompt », ou à partir d'un média comme du son ou une image). Elle s'inscrit ainsi dans la catégorie des IA génératives, une famille d'IA qui faisait déjà parler d'elle avec les hypertrucages (*deepfakes*) quelques années avant l'émergence des LLM⁴. Ces SIA peuvent couvrir un large spectre d'usages : résumer du texte, générer des réponses, assister un agent dans sa recherche d'information, transcrire une visio-conférence ou une déposition, faire une traduction à la volée...

Il faut noter que la DINUM ne développe pas directement des LLM. Cela nécessite en effet des ressources dont elle ne dispose pas. Elle utilise donc des modèles préexistants (modèles pré-entraînés, ou modèles de fondation), qu'elle adapte (*fine tuning*, *prompt engineering*) et intègre dans des SIA (ajoutant des capacités comme la génération augmentée par récupération documentaire – RAG, *Retrieval Augmented Generation*). La DINUM a d'ailleurs été la première administration française civile à faire du *fine tuning* et à intégrer du RAG dès l'été 2023. La diversité des modèles disponibles, en particulier les modèles dont l'usage est libre comme

³La génération de son, d'image ou encore de vidéo est également possible, mais ne fait actuellement pas partie des cas d'usage étudiés.

⁴Un historique complet et rigoureux des IA génératives dépasse largement le cadre de cet article. On mentionnera simplement l'apparition des réseaux antagonistes génératifs en 2014 [1], et celle des modèles auto-attentionnels, ou *transformers*, en 2017 [2]. Ces derniers sont utilisés dans la majorité des LLM aujourd'hui (mais il existe d'autres architectures de LLM).

ceux de Mistral AI⁵, permet à Albert de ne pas dépendre d'un seul fournisseur de modèle. Le DataLab a d'ailleurs déployé plus d'une dizaine de modèles différents et change régulièrement de modèles de fondation⁶. Tous ces déploiements sont faits dans le respect de ses normes de confidentialité et de sécurité⁷.

Pas juste un modèle : Trois piliers pour s'approprier l'IA

Face à la diversité d'usages anticipée, Albert a été pensé comme un dispositif technologique. Albert n'est donc pas à proprement parler une IA (il n'y a pas un seul modèle ou un seul SIA), il recouvre en fait une gamme de services et de produits d'IA, bâtie autour de trois piliers :

- le sur-mesure, où le DataLab développe une solution reposant sur les LLM, en co-construction directe avec les utilisateurs finaux ;
- l'IA à la demande (*IA as a service*), où il s'agit de mettre à disposition des services prêts à l'emploi, consommables directement *via* des applications en ligne (API en ligne ou logiciel en tant que service – SaaS, *Software as a Service*) ;
- le commun numérique où les codes, les modèles, les jeux de données, mais aussi les résultats de recherches sont publiquement partagés de manière ouverte et sous licences libres, et peuvent faire l'objet de contributions externes, notamment de la communauté du logiciel libre (*open source*).

Premier pilier

Le premier pilier a été assez naturellement le premier à être actif. C'est celui qui permet non seulement de monter en compétence, mais aussi de favoriser l'adoption des outils en assurant qu'ils soient bien adaptés aux besoins des utilisateurs.

On peut par exemple citer le cas d'Albert France Services, développé avec l'Agence nationale de la cohésion des territoires (ANCT). Il s'agit de fournir aux agents des France Services un assistant virtuel avec lequel ils peuvent échanger pour chercher une solution à un problème concret d'un usager. Le projet a été lancé en conditions opérationnelles en janvier 2024, et couvre aujourd'hui 44 France Services dans 6 départements. L'enjeu est double : vérifier que la technologie est suffisamment pertinente, *i.e.* que l'IA répond bien aux questions posées, et s'assurer que la solution est utile aux agents. On notera que la diversité des situations rencontrées par France Services

⁵ Mistral AI propose à la fois des modèles libres et ouverts (*i.e.* dont les poids sont accessibles), mais aussi des modèles propriétaires payants.

⁶ Il utilise aussi bien des modèles de Mistral que ceux de Meta, Lighton, Google ou encore PleIAs.

⁷ Le DataLab utilise en particulier des serveurs répondant à la certification de sécurité SecNumCloud, délivrée par l'Agence nationale de la Sécurité des systèmes d'information (ANSSI). Il échange régulièrement avec l'ANSSI sur les questions de sécurité, ainsi qu'avec la Commission nationale de l'informatique et des libertés (Cnil) sur le sujet de la protection des données personnelles.

est un véritable défi, et il a fallu itérer à plusieurs reprises en composant avec les limites des modèles actuels⁸.

Ces expériences sont précieuses pour l'administration, qui se forge ainsi un véritable savoir-faire dans la conduite de projet d'IA générative. Par ailleurs, les développements sur Albert France Services, en particulier le RAG, sont majoritairement repris dans tous les autres projets de la famille Albert (mutualisation et externalités positives).

Deuxième pilier

Le deuxième pilier (*IA as a service*) a ainsi largement bénéficié des travaux précédents. La première brique du pilier 2 a été le déploiement d'Albert API, qui met à disposition des administrations des fonctionnalités d'IA (notamment des LLM) *via* une API⁹ en ligne. Albert API, en service depuis juin 2024, est le premier service en ligne de ce type déployé sur des serveurs SecNumCloud. Il permet aux administrations de faire l'économie d'un déploiement d'IA (mobilisant puissance de calcul et compétences) pour leurs projets de SIA utilisant des LLM¹⁰. Les premiers utilisateurs ont rapporté un gain de temps pouvant aller jusqu'à trois mois dans le développement de leurs SIA.

Aujourd'hui, 25 projets dans l'administration publique utilisent Albert API. Il a par exemple permis d'automatiser des tâches de résumé, qui prenaient deux à trois jours habituellement, et qui prennent désormais moins d'une journée à être traitées¹¹. Albert API est utilisé par LaSuite (suite bureautique collaborative de la DINUM), qui a mis à disposition des capacités d'IA générative dans son éditeur de texte (Docs) et son tableur (Grist). Tous les agents utilisant ProConnect ont de fait accès à Albert¹².

Le pilier 2 est dès lors un vecteur structurant de la diffusion de l'IA générative dans l'administration, rendant l'IA accessible aux agents de l'État, directement dans leurs outils du quotidien (comme avec Docs). Il est encore en construction, et doit être complété par une offre de logiciels en tant que services, tels qu'un agent conversationnel généraliste¹³.

⁸ Au moins trois modèles spécialement ré-entraînés ont été testés, et l'interface utilisateur a régulièrement évolué en fonction des retours.

⁹ Interface de programmation applicative (API, *Application Programming Interface*).

¹⁰ Par ailleurs, Albert API respecte les normes imposées par les usages (celles d'OpenAI), de sorte qu'un projet développé sur une API LLM différente peut facilement migrer vers Albert API (*via* redirection d'URL).

¹¹ Dans le cas d'usage rapporté : la machine peut traiter en 20 minutes l'ensemble des textes à résumer, et les agents n'ont plus qu'à vérifier les résultats.

¹² LaSuite est accessible aux agents en se connectant simplement *via* ProConnect (<https://www.proconnect.gouv.fr/>). Aujourd'hui, 1 000 agents bêta-testent l'IA dans Docs.

¹³ Un premier agent conversationnel a été développé et testé à titre purement expérimental au DataLab dès le mois de novembre 2023. Un autre, accessible directement *via* la messagerie Tchap (Albert Tchap), a également été développé et présenté en juin 2024, puis ouvert à tous les agents de la DINUM en septembre 2024. Plusieurs versions ont été testées (dont une version multimodale). Ces expérimentations permettent de tester en conditions réelles les questions de protection des données personnelles, et de sécurité informatique.

Troisième pilier

Enfin, le troisième pilier est celui qui doit sceller l'appropriation collective de la technologie. Il vise à la mobilisation d'une communauté numérique où les administrations, avec la communauté *open source*, mais aussi les acteurs privés (éditeurs et / ou intégrateurs de logiciels, opérateurs, start-ups, hébergeurs *cloud*...) se seront réapproprié entièrement le projet¹⁴.

Les codes, les modèles, et les jeux de données utilisés par Albert sont librement disponibles et ouverts à contributions sur les plateformes publiques de référence Github (<https://github.com/etalab-ia/>), et Hugging Face (<https://huggingface.co/AgentPublic>). Le DataLab y a notamment mis à disposition le plus grand jeu de données administratives ouvertes au monde¹⁵. L'intégralité des données ouvertes de la direction de l'Information légale et administrative (DILA), incluant le site Légifrance, y est en particulier disponible sous une forme directement interrogeable par les SIA (format « vectorisé »)^{16, 17}.

Ce troisième pilier se construit sur un temps long. Les interventions en conférences, formations, les publications ou encore l'animation de communautés participent à ce travail.

Les perspectives : faire émerger les applications à forte valeur ajoutée

Albert est donc avant tout un programme de mise à disposition technologique (expertise, services, infrastructure, commun numérique), répondant à une forte demande. Il permet en particulier de réduire le risque de fuite de données par une utilisation inappropriée de services numériques non encadrés (*shadow IT*). En complément d'Albert, la DINUM a mis en place un incubateur de produits IA, Alliance¹⁸. Ce dispositif a pour objectif premier d'orienter les projets IA de l'administration vers la réussite au sens de l'impact réel des solutions proposées. C'est un sujet que ne traite que partiellement Albert, *via* son pilier 1 (les projets du pilier 1 sont d'ailleurs menés dans le cadre d'Alliance). Dans une start-up, Albert serait le CTO¹⁹, là où Alliance jouerait plutôt le rôle du CPO²⁰.

Par ailleurs, dans la perspective d'un déploiement durable de la technologie, les défis restent nombreux. La frugalité des SIA est un enjeu majeur. Dans ce domaine, le DataLab travaille sur la taille optimale des

modèles dans un esprit *lean* (chercher les modèles les plus petits possibles à niveau de performance donné). Cela nécessite d'évaluer correctement les performances des IA, ce qui reste un domaine de R&D très actif²¹.

Des limites inhérentes aux LLM posent également des questions de gouvernance. Ces IA font des erreurs, voire « hallucinent » (ce sont des modèles statistiques). C'est d'autant plus trompeur que, non seulement les réponses générées ont une forme assertive, mais aussi, plus elles seront performantes, plus elles seront crédibles. En cas d'erreur, cela pose des problèmes du fait de l'opposabilité des réponses d'une administration, soumise au Code des relations entre le public et l'administration (CRPA). Dans ces conditions, ces IA sont aujourd'hui destinées à être une aide aux agents, qui restent responsables de la relation avec les usagers. Mais les agents doivent être formés pour être conscients des limites des IA. De plus, l'explicabilité n'est pas assurée avec ces algorithmes, au sens des obligations de transparence algorithmique. Ces IA ne doivent donc pas être utilisées dans un processus de décision administrative.

Conclusion

Cet aperçu des défis à relever est loin d'être exhaustif. Plutôt que chercher à être holistique, l'approche du DataLab, et plus largement de la DINUM, se veut pragmatique avec des méthodes itératives (agile), et visant la frugalité (*lean*). La transformation numérique de l'État met à l'épreuve les capacités d'adaptation de l'administration. Dans le domaine de l'IA générative, l'État s'est montré pro-actif, et la France peut se féliciter de faire partie des pays les plus en avance sur le sujet.

Références

- [1] GOODFELLOW I., POUGET-ABADIE J., MIRZA M., BING X., WARDE-FARLEY D., OZAIR S., COURVILLE A. & BENGIO Y. (2014), "Generative adversarial nets", *Proceedings of the 27th International Conference on Neural Information Processing Systems*, MIT Press, 10 juin (arXiv 1406.2661).
- [2] VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L. & POLOSUKHIN I. (2017), "Attention is all you need", *Advances in Neural Information Processing Systems*, 30, 12 juin, p. 15 (arXiv 1706.03762).

¹⁴ Le code étant *open source*, tout le monde peut le réutiliser librement.

¹⁵ 380 milliards de tokens : https://huggingface.co/datasets/AgentPublic/open_government

¹⁶ Pour faire des recherches documentaires, par exemple pour du RAG, les SIA en question ont besoin de vectoriser les données, c'est-à-dire les transformer en représentations numériques (on parle aussi de plongement sémantique ou *embedding*), <https://huggingface.co/datasets/AgentPublic/DILA-Vectors>

¹⁷ Ce travail s'inscrit en fait dans le cadre d'un projet plus vaste, encore en développement (Albert Data).

¹⁸ <https://alliance.numerique.gouv.fr/>

¹⁹ *Chief Technology Officer*.

²⁰ *Chief Product Officer*.

²¹ Le DataLab mettra prochainement à disposition un outil d'évaluation en ligne (projet EG1 en interne).

Comment le système éducatif français se saisit-il de l'intelligence artificielle générative ?

Par **Émilie-Pauline GALLIÉ**

Inspectrice générale de l'éducation, du sport et de la recherche

et **Erwan PAITEL**

Inspecteur général de l'éducation, du sport et de la recherche

Depuis 2022, l'IA générative (IAG) est accessible à une grande partie de la population mondiale grâce à Internet, mais son utilisation soulève des questions sur la maîtrise des outils, les biais, et les limites éthiques et environnementales. En France, le système éducatif a encouragé les expérimentations de cette technologie. Cette première étape passée, il devient désormais nécessaire de développer des stratégies robustes pour intégrer les usages de ces technologies dans le parcours des élèves.

Pour passer d'initiatives individuelles à une éducation à l'IA pour tous, il faut une approche systématique et inclusive, garantissant l'accès aux outils et aux formations nécessaires pour tous les élèves. En particulier, il est crucial de former les enseignants à l'utilisation de l'IAG et de développer des compétences spécialisées. Enfin, les défis techniques, pédagogiques et environnementaux doivent être relevés pour assurer une intégration réussie de l'IA dans l'éducation. Cela inclut la mise en place d'infrastructures adéquates et la promotion de pratiques durables pour minimiser l'impact environnemental de l'IA.

À travers la stratégie nationale de l'intelligence artificielle (IA), les programmes d'investissement d'avenir (PIA3, France 2030), les initiatives « instituts interdisciplinaires en IA » (3IA) et *IA-cluster*, la France a, depuis 2018, massivement soutenu la recherche et l'innovation en IA. Ces efforts se matérialisent dans un récent classement de l'Université de Stanford¹, qui la positionne au sixième rang mondial en matière de dépôts de brevets liés à l'IA. Au-delà de ce succès scientifique, l'appropriation et l'utilisation de l'IA par les citoyens et les entreprises deviennent tout aussi cruciales pour tirer les bénéfices espérés de cette technologie générique.

L'IA générative (IAG), à la grande différence des technologies de rupture précédentes, est accessible quasiment à l'ensemble des humains, à partir du moment où ils ont une connexion Internet, ce qui correspond à 67,5 % de la population mondiale² et 93 % de la population européenne³. Cette situation est encourageante

en termes de réduction des inégalités numériques et plus largement des inégalités d'accès aux savoirs. Mais avoir accès à l'IAG ne signifie pas savoir s'en servir et encore moins en connaître les potentialités, les biais et les limites éthiques et environnementales. En outre, l'utilisation de l'IAG est encore interrogée que ce soit dans les écoles⁴ ou dans les entreprises⁵, notamment pour des questions de sécurité et de souveraineté.

Composé de 15 millions d'apprenants et près d'un million d'enseignants, le monde de l'éducation en France (au sens éducation nationale et enseignement supérieur) joue un rôle crucial dans l'appropriation éclairée des technologies. Après quelques hésitations, les tutelles ministérielles des établissements ont opté pour un encouragement de l'usage de l'IAG, sans attendre, d'une part, d'en maîtriser l'ensemble des enjeux et des difficultés, et, d'autre part, de disposer de solutions souveraines. Elles entendent permettre dès à présent aux enseignants de se saisir des possibilités offertes et préparer les jeunes générations à un avenir où cette technologie sera omniprésente. Elles

¹ <https://hai.stanford.edu/news/global-ai-power-rankings-stanford-hai-tool-ranks-36-countries-ai>

² <https://www.statista.com/statistics/617136/digital-population-worldwide/>

³ <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/ddn-20241217-1>

⁴ <https://acteurspublics.fr/articles/lia-un-outil-damelioration-du-systeme-educatif-a-double-tranchant>

⁵ <https://www.forbes.fr/technologie/faut-il-interdire-lia-generative-au-bureau/>

proposent déjà un catalogue fourni de formations aux enseignants^{6, 7} pour les accompagner.

Cependant, les enjeux de l'IA dans l'éducation sont encore multiples, et les réponses actuelles restent partielles, en particulier pour s'assurer qu'indépendamment des origines socio-économiques ou territoriales, tout jeune ait accès aux outils et aux connaissances nécessaires pour utiliser l'IA de manière consciente, efficace et éthique.

Cet article est organisé en trois parties. La première présente la place de l'IA dans l'éducation. La seconde insiste sur la nécessité de passer à l'échelle les nombreuses expérimentations déjà présentes, en raison notamment des bouleversements sur le métier d'enseignant et le contenu des formations lui-même que l'IAG va interroger. Enfin, la troisième partie revient sur les défis restants pour une appropriation complète de l'IA dans le système éducatif.

L'IAG a une place croissante mais hétérogène dans le système éducatif

Dans l'ensemble de la société française, l'usage de l'IAG est très variable. Comme souvent quand il s'agit de technologies, les jeunes, en particulier les jeunes adultes (18-29 ans), sont plus enclins à utiliser l'IAG tandis que l'adoption diminue avec l'âge. Une enquête Ifop⁸ relève que 25 % des Français utilisent des IA génératives, avec une adoption plus marquée chez les jeunes. Une autre enquête de l'agence Heaven révèle que 97 % des étudiants connaissent au moins une IAG et plus de la moitié d'entre eux s'en servent régulièrement⁹.

La confiance dans les IA génératives varie également selon l'âge. Les jeunes utilisateurs tendent à faire davantage confiance aux informations générées par l'IA, tandis que les utilisateurs plus âgés sont plus critiques et prudents¹⁰.

Si l'utilisation de l'IAG parmi les enseignants est inférieure à celle de leurs étudiants (34 % vs 55 %¹¹, août 2023), ils en perçoivent le potentiel pour la création de contenus pédagogiques, leur adaptation aux besoins particuliers des élèves, favorisant ainsi une approche personnalisée, l'aide à la rédaction de documents, à la production d'exercices, au suivi des apprentissages, à l'évaluation des compétences des élèves, à

la correction des examens, etc. Ils appréhendent également les conséquences importantes sur la manière de transmettre, de former et d'évaluer, et donc sur la manière d'exercer leur métier. Ce dernier point peut générer, notamment pour les enseignants les moins familiers de l'IA et plus généralement des outils numériques, de fortes inquiétudes.

Ce contexte contribue à expliquer que l'intégration de l'IAG dans le système éducatif français est, selon les établissements ou leur tutelle, absente, en phase de réflexion¹², de consultation¹³, d'expérimentation¹⁴ ou de cadrage¹⁵.

Il est nécessaire de passer d'une phase expérimentale à un accès à l'éducation à l'IA pour tous et par l'IA accessible à tous

Initialement portée par des expérimentations appuyées par le ministère¹⁶ et des initiatives individuelles ou territoriales, pour certaines aussi labellisées¹⁷ et financées¹⁸, l'intégration de l'IA dans l'éducation doit désormais pouvoir être généralisée. Si l'approche expérimentale est louable et illustre une capacité des services et des enseignants à fonctionner en mode agile, l'effet, au sein d'un système éducatif qui se veut égalitaire, est partiel et ne permet pas d'assurer une appropriation et une diffusion à grande échelle. L'expérimentation et les initiatives doivent donc laisser place progressivement à une stratégie et une coordination nationale à long terme s'appuyant sur les initiatives déjà enclenchées. Ainsi, pour maximiser les bénéfices de ces initiatives, il est essentiel de les poursuivre, de les recenser de manière systématique et de les évaluer afin de créer une base de connaissances partagée, de favoriser les échanges de bonnes pratiques lors du passage à l'échelle, et de veiller à l'efficacité budgétaire des étapes initiales et de s'assurer d'un vecteur budgétaire de déploiement national pérenne.

⁶ https://magistere.education.fr/local/magistere_offers/index.php?v=formation

⁷ https://www.reseau-canope.fr/ia-en-classe?mtm_campaign=ia_en_classe_2024&mtm_source=ia_en_classe_2024_home_rc&mtm_medium=home_rc

⁸ <https://www.ifop.com/publication/2eme-vague-du-barometre-sur-la-perception-et-lusage-des-ia-generatives-par-les-francais/>

⁹ <https://www.lesphinx-developpement.fr/wp-media/uploads/2023/10/Infographie-IA-dans-lenseignement-1.pdf>

¹⁰ <https://www.ifop.com/wp-content/uploads/2024/03/120667-presentation.pdf>

¹¹ <https://www.lesphinx-developpement.fr/wp-media/uploads/2023/10/Infographie-IA-dans-lenseignement-1.pdf>

¹² <https://www.univ-gustave-eiffel.fr/luniversite/pages-speciales/vue-detaillee/chatgpt-dans-lenseignement-superieur-la-vraie-question-nest-pas-son-interdiction-mais-son-usage>

¹³ <https://www.aefinfo.fr/depeche/721922-ia-anne-genetet-lance-une-consultation-aupres-des-enseignants-et-eleves-pour-etablir-des-lignes-directrices-claires>

¹⁴ <https://www.aefinfo.fr/depeche/719522-ia-nous-avons-cree-une-infrastructure-securisee-qui-fonctionne-avec-du-rag-o-wong-hee-kam-universite-de-rennes>

¹⁵ <https://www.pantheonsorbonne.fr/formation/lia-et-moi-luniversite>

¹⁶ <https://eduscol.education.fr/1603/le-dispositif-edu>

¹⁷ <https://tne.reseau-canope.fr>

¹⁸ <https://www.banquedesterritoires.fr/sites/default/files/2022-06/TNE.pdf>

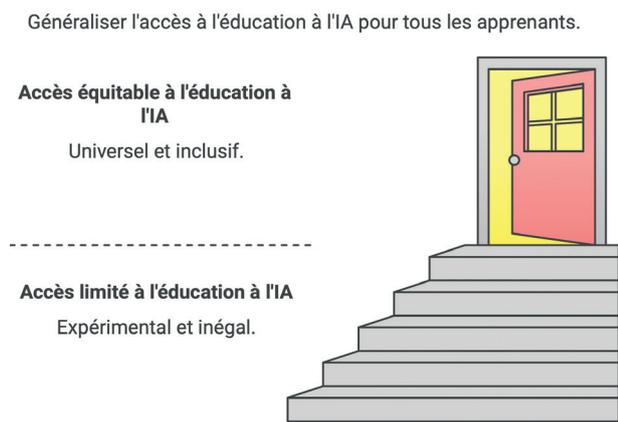


Figure 1 : Généraliser l'accès à l'éducation à l'IA pour tous les apprenants (Source : Illustration générée par les auteurs avec l'IA générative Napkin.ai).

Côté élèves, pour passer à une éducation à – et avec – l'IA pour tous, il est crucial de développer une approche systématique et inclusive. Cela implique de s'assurer que tous les élèves, indépendamment de leur origine socio-économique ou territoriale, aient accès aux outils et aux formations nécessaires. Enseigner aux élèves à utiliser l'IA ne se limite pas à la formation à son usage : l'utiliser et la maîtriser de manière critique et réfléchie est fondamental. Ils doivent comprendre les limites et les biais potentiels de ces technologies pour pouvoir les utiliser de manière éthique et responsable. Il existe également des opportunités avec l'IA de réduction des écarts pour les élèves concernés par des troubles spécifiques des apprentissages¹⁹. L'IA peut modifier des ressources (application Cantoo pour les publics dyslexiques par exemple) et les adapter afin de garantir une éducation plus personnalisée pour des profils particuliers²⁰. Si la tendance au tout-IA n'est pas d'actualité, il est nécessaire de rappeler que l'école est bien plus qu'un lieu d'apprentissage de savoirs. De fait, l'apprentissage personnalisé permis par l'IA est une opportunité à saisir mais il ne doit pas se faire au détriment de la construction d'un collectif qui se joue dans les salles de classe et via l'interaction sociale.

L'enseignement et l'utilisation de l'IAG ne doivent pas être vus uniquement sous l'aspect « maîtrise de la technologie », mais également comme des compétences nécessaires à chaque discipline. Les enseignants devront aussi être formés pour développer des compétences spécialisées, combinant leur domaine d'expertise avec l'IA (souvent désigné par la formule « X + IA »). Cette double compétence est essentielle pour préparer les élèves à un avenir où l'IA jouera un rôle central, probablement incontournable, dans de nombreux secteurs professionnels.

Enfin, très vite va se poser la question des compétences et connaissances à acquérir, dans un environnement où l'IA devient très présente et semble tout savoir. Certains étudiants se demandent déjà à quoi bon apprendre alors que l'IAG apporte toutes les réponses à leurs

questions. Ces réflexions sont notamment menées par le GIP PIX en lien avec les branches professionnelles, afin d'aboutir à une certification de type « PIX IA ».

Côté enseignants, ceux-ci pourront se servir de l'IAG pour la création de contenus ou de scénarii pédagogiques ainsi que dans leurs pratiques pédagogiques quotidiennes ou pour confier du travail aux élèves à la maison (l'application Nolej répond notamment à ce cahier des charges), mais aussi potentiellement pour répondre à des tâches administratives. L'ambition est de trouver avec l'IA un assistant à l'enseignant, un outil d'aide aux apprentissages et non un remplaçant, l'interaction humaine dans la classe jouant un rôle vital dans l'éducation et les apprentissages des élèves.

Étant donné les enjeux, l'intégration de l'IAG dans les pratiques pédagogiques nécessite une réflexion collective, multidimensionnelle et critique sur ses usages au sein de la communauté éducative. Sa plus-value doit être démontrée par la recherche scientifique.

Le défi de l'éducation : rendre accessible l'IAG à 15 millions d'apprenants et à un million d'enseignants

Le déploiement d'outils à IA doit s'accompagner d'actions pour s'assurer de l'appropriation de ces outils et de leur bonne utilisation.

Le développement de chartes d'utilisation constitue une première étape nécessaire pour garantir et sécuriser une appropriation collective de l'IAG. On observe déjà l'émergence de telles chartes au sein d'établissements d'enseignement supérieur ou d'académies^{21,22,23}. Le ministère de l'Éducation nationale publie au printemps 2025 un cadre d'usages avec des principes généraux et des lignes directrices à visées opérationnelles pour définir les usages encouragés et les lignes rouges, qui pourront s'adapter avec les évolutions des technologies, de la réglementation et des usages. Ces chartes permettent aux institutions, aux enseignants et aux apprenants, de disposer d'un cadre qui les protège, de définir des niveaux d'acculturation nécessaires, des limites claires des usages et des pratiques éthiques et respectueuses d'un développement durable. Il y a un enjeu à converger vers un modèle de charte commun pour l'ensemble des acteurs qui reprennent les mêmes grands principes, plutôt que chacun crée sa propre charte.

¹⁹ <https://www.inserm.fr/dossier/troubles-specifiques-apprentissages/>

²⁰ <https://hal.science/hal-04557108/document>

²¹ <https://pedagogie.ac-toulouse.fr/emi/charte-ethique-des-usages-des-intelligences-artificielles#:~:text=Les%20intelligences%20artificielles%20g%C3%A9n%C3%A9ratives%20mises,le%20domaine%20de%20l'%C3%A9ducation.>

²² https://www.univ-orleans.fr/upload/public/2024-06/dircom_2024_charte_IA_UO_R.pdf

²³ <https://www.sciencespo-grenoble.fr/sites/default/files/2024-07/Charte%20Intelligence%20artificielle%202024-1.pdf>

Chartes d'utilisation de l'IAG

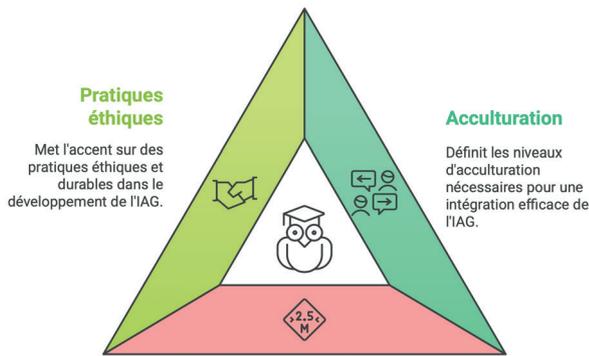


Figure 2 : Chartes d'utilisation de l'IA
(Source : Illustration générée par les auteurs avec l'IA générative Napkin.ai).

Il faut ensuite s'assurer qu'une appropriation collective s'opère auprès de tous ceux qui éduquent : les futurs enseignants, tout comme ceux en exercice, doivent donc tous être formés à l'IA afin d'être en capacité de l'intégrer dans leurs pratiques et de former les apprenants, et cela quelle que soit leur discipline d'exercice puisque l'IAG va se diffuser dans tous les domaines.

L'intégration généralisée des outils à IA dans l'éducation ne doit pas masquer les questions concernant leur coût environnemental. Les technologies d'IA consomment beaucoup de ressources²⁴ et d'énergie²⁵, ce qui a un impact significatif et croissant sur l'environnement. Dans une stratégie d'usage généralisé en éducation, il est crucial de considérer les conséquences sur l'empreinte environnementale et de promouvoir des pratiques durables. Cela pourrait inclure l'utilisation de centres de données écologiques, le développement de LLM (*Large Language Models*) frugaux²⁶ pour réduire la consommation d'énergie et la sensibilisation des utilisateurs aux impacts environnementaux de l'IA, via par exemple une autorisation d'un nombre limité de prompts pour chaque séquence d'utilisation de l'IAG. Cette pratique aurait un impact environnemental mais aussi pédagogique pour les élèves, qui devraient réfléchir à la formulation de leur requête.

Au-delà du cadre d'usage, il est crucial de garantir la disponibilité des capacités techniques nécessaires dans les établissements scolaires et d'enseignement supérieur. Cela inclut l'accès à des infrastructures adéquates, telles que des connexions Internet haut débit, des équipements informatiques modernes et des plateformes sécurisées pour l'utilisation de l'IA, et en capacité de répondre à l'ensemble des requêtes, qui vont fortement augmenter avec la généralisation de l'utilisation de l'IAG.

²⁴ <https://www.gstatic.com/gumdrop/sustainability/google-2023-environmental-report.pdf>

²⁵ <https://www.polytechnique-insights.com/tribunes/energie/ia-generative-la-consommation-energetique-explose/>

²⁶ <https://www.campusmatin.com/numerique/pedagogie/nous-ne-voulions-pas-un-simple-chatbot-l-universite-de-rennes-teste-l-ia-en-circuit-ferme.html>

Plus largement, l'intégration de l'IAG dans l'éducation doit être pensée globalement en s'intéressant aux usages afin de ne pas être enfermé dans une technologie qui pourrait être vite dépassée. En effet, si les modèles larges de langage (LLM) sont aujourd'hui le support technique de l'IAG, le dynamisme de la recherche laisse à penser que d'autres modèles d'IA, plus performants et/ ou moins énergivores, pourraient supplanter les LLM dans les années à venir. Le monde de l'éducation doit pouvoir s'adapter rapidement à ces potentielles évolutions techniques.

Intégration de l'IA dans l'éducation

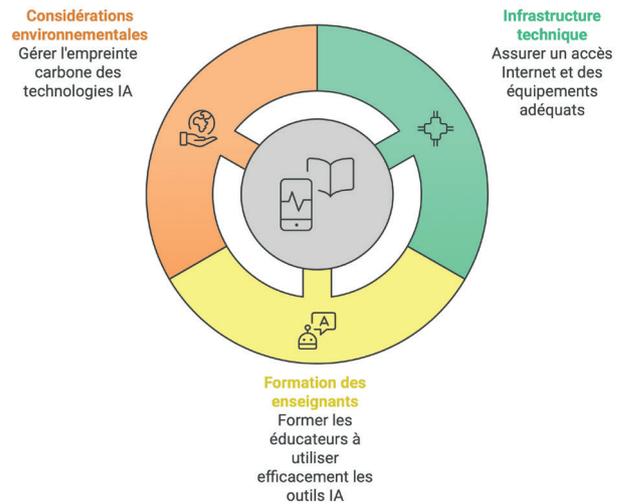


Figure 3 : Intégration de l'IA dans l'éducation
(Source : Illustration générée par les auteurs avec l'IA générative Napkin.ai).

Conclusion

La France a pris des mesures significatives pour permettre l'intégration de l'IA dans son système éducatif, mais il reste encore beaucoup à faire pour garantir une adoption équitable et durable. En développant des stratégies nationales, co-construites avec les enseignants, en formant les enseignants et en sensibilisant les élèves aux enjeux éthiques et environnementaux, la France peut s'assurer que l'IA devienne un outil puissant pour l'éducation de demain, et répondre aux besoins de la société et des entreprises. La généralisation de l'IA doit aussi être une occasion de repenser le rôle des enseignants mais aussi des élèves dans les processus d'apprentissage.

Droit et intelligence artificielle : interactions et transformations

Par Yannick MENECEUR

Magistrat, inspecteur de la justice et maître de conférences associé à l'université de Strasbourg

Les applications de l'intelligence artificielle (IA) dans les domaines du droit et la justice ont quitté depuis bien longtemps les laboratoires et se banalisent. Dans ce contexte, elles peuvent être analysées sous deux angles : celui de leur encadrement juridique, en tant qu'objets de droit, et celui de leur influence sur le droit, en véritable nouvel acteur.

Cet article a vocation à présenter les contours des interactions et des transformations en cours, en décrivant notamment les nouveaux cadres juridiques européens et les enjeux sociétaux du déploiement d'IA au sein des systèmes judiciaires, en allant au-delà de la classique balance entre bénéfices et risques.

Les applications de l'intelligence artificielle (« IA¹ ») dans les domaines du droit et de la justice sont sorties des laboratoires et deviennent de plus en plus banalisées : occultation automatique des données sensibles dans les décisions de justice en *open data*, outils de synthèse, de recherche juridique ou de jurimétrie sont autant d'exemples déjà exploités quotidiennement par les professions juridiques et judiciaires. Mais l'intégration de ces technologies exige le respect d'un cadre rigoureux de déploiement, afin de minimiser les risques de renforcement de discriminations illégitimes, d'atteinte à la vie privée et, d'une manière plus générale, de violation des droits fondamentaux. Autant de préoccupations qui ont justifié l'adoption en 2024 de nouvelles dispositions juridiquement contraignantes au niveau européen pour encadrer de manière transversale les usages des systèmes d'IA (SIA²).

Il est vrai que les réelles potentialités de « l'IA » n'ont pas toujours été simples à objectiver du fait d'un marketing extrêmement agressif de l'industrie numérique dans son ensemble. Les recherches en études des sciences et techniques (*Science and Technic Studies*) nous offrent pourtant un cadre, trop peu souvent mobilisé dans le cadre de la prise de décision publique, pour dépasser les simples questions instrumentales liées à « l'IA » et ne pas se satisfaire d'une balance classique entre les bénéfices et les risques.

¹Afin de se garder de tout anthropomorphisme et par commodité éditoriale, l'expression d'intelligence artificielle sera présentée sous la forme de son acronyme et entre guillemets.

²L'acronyme de systèmes d'intelligence artificielle, « SIA », sera employé pour désigner précisément les systèmes visés dans l'article 3 du règlement sur l'intelligence artificielle de l'Union européenne et de l'article 2 de la convention-cadre du Conseil de l'Europe.

Les rapports entre le droit et « l'IA » peuvent donc être analysés à la fois sous l'angle de l'alignement de cette technologie sur un cadre de confiance par le droit, et sous celui de la concurrence que cette même technologie exerce sur le droit.

L'intelligence artificielle, objet de droit

Il sera d'abord examiné les nouveaux cadres juridiques applicables à « l'IA », avant d'approfondir la question des mécanismes de responsabilité.

La régulation et l'encadrement juridique de l'intelligence artificielle

L'émergence d'une régulation spécifique de l'intelligence artificielle

D'intenses controverses ont animé l'ensemble des parties prenantes sur l'intérêt et la manière d'encadrer l'usage non seulement des réseaux de neurones profonds (*deep neural networks*), mais également, parfois dans une totale confusion, de la vaste gamme des différentes méthodes d'apprentissage automatique (*machine learning*) et d'autres approches de traitement statistique des données. Passée une assez intense vague cherchant à instrumentaliser l'éthique de la fin des années 2010³, c'est le principe d'un encadrement juridique contraignant qui s'est imposé au début des années 2020 dans les politiques publiques européennes.

L'impulsion est venue des institutions européennes, dès 2021 avec la proposition de règlement sur l'intelligence

³Sur l'instrumentalisation des discours éthiques et la nécessité d'adopter une réglementation juridiquement contraignante, voir Y. Meneceur, *L'intelligence artificielle en procès : plaidoyer pour une réglementation internationale et européenne*, Bruylant, 2020, pp. 219-225.

artificielle (RIA ou AI Act⁴) par la Commission européenne et en 2022 avec l'ouverture des négociations au Conseil de l'Europe d'un traité international au sein d'un comité sur l'intelligence artificielle (CAI), après plus de deux années d'instruction par le comité *ad hoc* sur l'intelligence artificielle (CAHAI). À l'échelle internationale, des organisations telles que l'OCDE⁵ et l'UNESCO⁶ ont également adopté des instruments juridiques non contraignants sous la forme de recommandations, contribuant à établir un cadre de gouvernance internationale de cette technologie⁷.

En 2024, l'Union européenne a adopté le RIA⁸, législation de sécurité des produits fondée notamment sur le nouveau cadre législatif européen (NLF – New Legislative Framework, ensemble adopté en 2008⁹), et le Conseil de l'Europe a ouvert à la signature une convention-cadre sur l'intelligence artificielle, traité international de haut niveau, posant les principes utiles à des instruments ultérieurs sectoriels plus précis, sans déterminer de prohibitions précises, et laissant une assez large marge de manœuvre aux États signataires pour en achever les objectifs.

Les spécificités des cadres juridiques européens relatifs à l'intelligence artificielle

Le RIA se donne pour objectif de promouvoir l'adoption de cette technologie au sein des 27 États membres et d'harmoniser les règles de mise sur le marché, afin de protéger la santé, la sécurité et les droits fondamentaux des citoyens européens. Ce texte s'approprie dans ses fondements les lignes directrices en matière d'éthique pour une intelligence artificielle digne de confiance du groupe d'experts de haut niveau sur l'intelligence artificielle (GEHN IA), adoptées en avril 2019¹⁰, lignes directrices présentées elles-mêmes comme opérationnalisant

la Charte des droits fondamentaux de l'Union européenne dans le contexte particulier de cette technologie.

La convention-cadre sur l'intelligence artificielle du Conseil de l'Europe a été formellement adoptée par le Comité des ministres le 17 mai 2024 et ouverte à la signature à Vilnius le 5 septembre 2024¹¹. Les dispositions de ce texte ne visent pas directement les personnes, physiques ou morales, développant ou utilisant des SIA. Elles s'adressent aux États parties à la convention-cadre, qui doivent mettre en œuvre des mesures pour atteindre les objectifs fixés par le traité.

La responsabilité juridique de l'IA Vers une personnalité juridique pour l'IA ?

L'idée de conférer un statut juridique à des systèmes automatisés comme des robots ou des « IA » repose sur leur capacité croissante à prendre des décisions de manière autonome et à parfois produire des résultats n'étant pas prévus par leurs concepteurs. Certains avancent que, pour des systèmes dotés d'un très haut degré d'autonomie, cette reconnaissance pourrait simplifier l'attribution de responsabilités dans des situations où les dommages résultent directement des actions de l'IA, en faisant jouer des mécanismes de collectivisation des risques¹².

Cependant, cette proposition soulève des interrogations fondamentales quant à ses conséquences sur la responsabilité civile et pénale. Accorder une personnalité juridique à « l'IA » pourrait diluer la responsabilité des concepteurs, développeurs ou utilisateurs, en créant une entité intermédiaire difficile à sanctionner. En matière pénale, cela poserait également des problèmes juridiques et pratiques, où le régime de responsabilité des personnes morales ne pourrait être simplement dupliqué : comment, notamment, évaluer l'intentionnalité d'une machine ? Ces limites ont conduit à privilégier une approche où « l'IA » est perçue comme un outil sophistiqué, replaçant l'humain au cœur des mécanismes de responsabilité.

L'adaptation des régimes de responsabilité existants

Dans le sillage du RIA, la Commission européenne a entendu adapter en matière civile deux régimes de responsabilité applicables aux SIA : celui, sans faute, de responsabilité du fait des produits défectueux, et un autre, pour faute, en matière extracontractuelle¹³.

⁴Il sera préféré ici l'emploi de l'acronyme français « RIA ». La formulation de « IA Act », contraction de l'acronyme français de la technologie et du terme anglais « Act », parfois employé par la doctrine et certaines publications, est impropre et ne sera pas employée ici.

⁵Recommandation (OCDE/LEGAL/0449) du Conseil sur l'intelligence artificielle du 22 mai 2019 amendée le 3 mai 2024.

⁶Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle, adoptée en séance plénière le 23 novembre 2021.

⁷Voir le site Internet, coordonné par huit organisations internationales, <https://globalpolicy.ai> pour une vue globale de ces initiatives.

⁸Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n°300/2008, (UE) n°167/2013, (UE) n°168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle)

⁹Cet ensemble législatif regroupe le règlement (CE) n°765/2008 fixant les prescriptions relatives à l'accréditation et à la surveillance du marché des produits, la décision 768/2008 relative à un cadre commun pour la commercialisation des produits, qui comprend des dispositions de référence à incorporer dans les révisions de la législation sur les produits et le règlement (UE) 2019/1020 sur la surveillance du marché et la conformité des produits.

¹⁰Lignes directrices en matière d'éthique pour une IA digne de confiance rédigées par le groupe d'experts de haut niveau sur l'intelligence artificielle, 8 avril 2019.

¹¹Convention-cadre (STCE n°225) du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, accessible sur <https://rm.coe.int/1680afae3d> (consulté le 5 janvier 2025).

¹²Voir les débats sur la résolution du Parlement européen sur les règles de droit civil sur la robotique : A. Tabuteau, « Intelligence artificielle : des experts se mobilisent contre la création d'une personnalité juridique pour les robots », Lefebvre Dalloz, 18 avril 2018, accessible sur https://www.efl.fr/actualite/intelligence-artificielle-experts-se-mobilisent-contre-creation-personnalite-juridique-robots_R-1216c84f-c958-480d-b969-90f16e194df3 (consulté le 4 janvier 2025)

¹³Pour une présentation des deux propositions de directives, voir Y. Meneceur, « Les nuances des nouveaux régimes de responsabilité des IA en Europe », Actu-Juridique.fr, 15 mars 2023, accessible sur <https://www.actu-juridique.fr/ntic-medias-presse/les-nuances-des-nouveaux-regimes-de-responsabilite-des-ia-en-europe/> (consulté le 4 janvier 2025)

L'Union européenne a adopté le 23 octobre 2024 une directive abrogeant les dispositions de 1985 sur la responsabilité du fait des produits défectueux, l'adaptant non seulement aux SIA mais également, de manière plus large, aux enjeux présentés de manière générale par les logiciels¹⁴. L'objectif est de faciliter l'accès à la preuve pour les demandeurs et d'établir une présomption de causalité en cas de complexité technique ou scientifique. Le projet de directive en matière de responsabilité civile extracontractuelle (régime de responsabilité pour faute), proposée par la Commission le 28 septembre 2022¹⁵, a été abandonné en début 2025 au regard des difficultés d'articulation avec le RIA.

L'intelligence artificielle, nouvel acteur du droit

Si « l'IA » est bien devenue une alliée du droit et de ses professions, elle concurrence paradoxalement le droit en instrument de gouvernance de la société.

L'intelligence artificielle, alliée du droit et de ses professions

L'automatisation sophistiquée des tâches juridiques

L'une des premières applications de « l'IA » attendues par les juristes porte sur l'optimisation de la veille et de la recherche juridique, l'analyse de masses importantes de documents et, plus spécifiquement pour les avocats, l'aide à l'élaboration de stratégies avec la prévision du montant d'indemnité en cas de saisine d'une juridiction¹⁶. Du côté de l'administration de la justice, il est attendu des SIA de pouvoir améliorer l'accès à la justice des citoyens, au moyen par exemple d'agents conversationnels (*chatbots*) capables de simplifier la rédaction de plaintes ou d'expliquer des décisions dans un langage accessible, ou encore de libérer des ressources pour des audiences approfondies dans les affaires complexes, tout en accélérant le traitement des cas simples. Des outils d'aide à la qualification pénale sont également expérimentés dans les tribunaux¹⁷. Par ailleurs, ces applications ne se limitent pas à améliorer l'efficacité de la justice, elles ambitionnent également d'améliorer la sécurité juridique. Les SIA sont ainsi attendus pour identifier les incohérences ou omissions dans des documents juridiques, réduisant dès lors les risques d'erreurs humaines.

¹⁴ Directive (UE) 2024/2853 du Parlement européen et du Conseil du 23 octobre 2024 relative à la responsabilité du fait des produits défectueux et abrogeant la directive 85/374/CEE du Conseil.

¹⁵ Proposition de directive COM(2022)496 final du Parlement européen et du Conseil du 28 septembre 2022 relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA).

¹⁶ Pour une analyse des cas d'usage « d'IA » génératives dans le domaine des professions du droit, voir Y. Meneceur, « IA générative et professionnels du droit – Comprendre et s'approprier la langue des probables », *LexisNexis*, 2024, pp. 60-78.

¹⁷ Outil en phase de test au moment de l'écriture de cet article, voir « L'intelligence artificielle au sein de la justice française », Info.gouv.fr, 10 décembre 2024, accessible sur <https://www.info.gouv.fr/actualite/lintelligence-artificielle-au-sein-de-la-justice-francaise> (consulté le 4 janvier 2025).

Toutefois, l'administration de la justice craint, dans le même temps, une massification des demandes avec l'automatisation de la génération des saisines des tribunaux : l'adaptation des moyens logistiques de traitement serait donc à anticiper¹⁸. Au-delà de la quantité, c'est aussi la qualité de cette massification qui sera à interroger : loin des ambitions de fiabilisation, les « IA » génératives grand public, utilisées dans des domaines techniques aussi spécialisés que le droit, ne produisent pas un contenu suffisamment fiable pour être employé sans la vérification d'un professionnel. Le défi logistique redouté par la justice sera donc accompagné d'un défi qualitatif, où des moyens bien humains resteront vraisemblablement encore nécessaires.

La transformation des professions juridiques

L'émergence de ces technologies a déjà des répercussions sur les savoir-faire dans le secteur juridique. Les avocats, magistrats et autres acteurs du droit, sans devenir eux-mêmes des ingénieurs ou techniciens informatiques, ont vocation à dépasser la simple maîtrise bureautique. L'article 4 du RIA, qui est déjà entré en vigueur et applicable aux professions du droit, exige des fournisseurs et des déployeurs de prendre « des mesures pour garantir, dans toute la mesure du possible, un niveau suffisant de maîtrise de l'IA pour leur personnel et les autres personnes s'occupant du fonctionnement et de l'utilisation des systèmes d'IA pour leur compte, en prenant en considération leurs connaissances techniques, leur expérience, leur éducation et leur formation, ainsi que le contexte dans lequel les systèmes d'IA sont destinés à être utilisés, et en tenant compte des personnes ou des groupes de personnes à l'égard desquels les systèmes d'IA sont destinés à être utilisés ».

Cette évolution implique donc une transformation des cursus universitaires et de la formation continue des professionnels : des universités adaptent déjà depuis quelques années leurs programmes pour inclure des modules sur le droit numérique, voire créer des filières dédiées. L'École nationale de la magistrature a créé un cycle approfondi sur le numérique, en association avec l'École des Mines de Paris, l'École nationale des greffes et le Campus du numérique public de la Direction interministérielle du numérique. L'ambition de la formation est de permettre aux professionnels de la justice d'appréhender les enjeux du numérique dans l'institution judiciaire, d'acquérir une connaissance pratique des dispositifs numériques existants, de piloter un projet de transformation numérique au niveau local ou national, et de s'approprier les perspectives d'usages permises par le numérique.

L'intelligence artificielle, concurrente du droit ?

L'intelligence artificielle : une technologie n'étant ni neutre, ni un simple outil

« L'IA » n'est en réalité ni une technologie « neutre », ni un simple « outil »¹⁹. Ses capacités sont mieux

¹⁸ *Ibid.*

¹⁹ Sur ce thème, voir les développements dans Y. Meneceur, « IA générative et professionnels du droit », *op. cit.*, pp. 41-49.

appréhendées au travers de son affordance²⁰, c'est-à-dire des potentialités concrètes qu'elle autorise avec le traitement mathématique, statistique et probabiliste des données d'un environnement, plutôt qu'au travers de grilles de lectures classant de « bons » ou de « mauvais » usages. Une des questions centrales est en effet de savoir, avant toute considération d'un possible contrôle, si ces méthodes algorithmiques, très performantes dans des environnements fermés avec des données quantitatives, peuvent véritablement restituer la « texture ouverte » du droit, caractérisée par d'innombrables ambiguïtés d'interprétation et des nuances contextuelles. Contrairement à un arbre décisionnel rigide, le raisonnement juridique autorise, dans un système juridique donné, des solutions contradictoires pour une même situation, un défi que les différents types contemporains « d'IA » peinent à relever²¹.

L'argument de la complémentarité humain-machine dans le processus décisionnel, souvent présentée comme permettant de combiner les capacités des professionnels et des « IA », ne permet pas de rejeter cette critique. La collaboration harmonieuse entre « l'IA » et les professionnels, trop souvent fantasmée, sous-estime les risques d'une délégation excessive résultant, notamment, de biais cognitifs comme les biais d'automatisation et les biais d'ancrage²².

L'introduction de SIA d'aide à la décision auprès de juges peut donc influencer en profondeur sur leur office avec, comme opportunité, une meilleure harmonisation des décisions en introduisant une nouvelle forme de contrôle interne qualifiée par un ancien premier président de la Cour de cassation de « collégialité élargie²³ ». Mais si la standardisation est perçue par certains professionnels comme une garantie d'équité, elle peut également entraîner une perte de définition nécessaire pour traiter les spécificités de chaque affaire. Cette tension entre généralisation et individualisation reste un point critique dans l'usage de « l'IA » dans le domaine de la justice.

Un lent glissement d'un État de droit vers un « État des algorithmes » ?

L'autre question soulevée par l'introduction de SIA dans le domaine de la justice est celle du transfert de

gouvernementalité à l'œuvre, conférant aux concepteurs de ces systèmes un tout nouveau pouvoir d'influence sur les métiers qu'ils assistent. Ce phénomène, étudié depuis des années sous des appellations telles que « gouvernementalité algorithmique²⁴ », « algocratie²⁵ » ou « État des algorithmes²⁶ », traduit la transformation des mécanismes de gouvernance résultant de l'usage croissant d'outils algorithmiques dans la prise de décision et substituant le calcul à la délibération politique et à l'État de droit. Même si les divers algorithmes se présentent aujourd'hui comme « exécutant » une norme, ils deviendraient (ou plus exactement leurs concepteurs deviendraient) producteurs de normes : un programme informatique, quel que soit sa nature, est toujours une somme de décisions et d'arbitrages qui sont le reflet de l'interprétation par le concepteur du problème qu'il a à résoudre.

Ce glissement est exacerbé par un contexte sociétal fortement empreint d'hyper-individualisme et d'une méfiance accrue envers les institutions démocratiques. Certains voient dans la technologie un salut, espérant des mécanismes décisionnels objectifs et dépourvus d'affect, renforçant ainsi une technocratie automatisée²⁷.

Le rôle des décideurs publics et des professionnels du droit dans l'accompagnement de cette transformation est essentiel. Ils doivent veiller à ne pas céder à un « technosolutionnisme » naïf, qui confierait des responsabilités cruciales à des systèmes technologiques sans compréhension de leurs capacités, ni la mise en place d'un clair dispositif de gouvernance organisant leur exploitation. Leur expertise juridique, combinée à une compréhension technique, est nécessaire pour préserver les principes fondamentaux de l'État de droit, et garantir une gouvernance équilibrée et équitable.

²⁰ L'affordance est un concept qui a été proposé par le psychologue américain James Gibson dans les années 1960. Il qualifie la faculté des êtres vivants à guider leurs actions selon ce que l'environnement offre en termes de potentialités d'actions. L'affordance est, comme l'exprime James Gibson : « une structure relationnelle multifacée entre un objet / une technologie et l'usage et les comportements que celui-ci permet ou contraint dans un contexte donné ».

²¹ S'agissant des larges modèles de langage, voir par exemple M. Dahl, V. Magesh, M. Suzgun, D.E.Ho, "Large legal fictions: Profiling legal hallucinations in Large Language Models", *Journal of Legal Analysis*, 26 juin 2024, accessible sur <https://academic.oup.com/jla/article/16/1/64/7699227> (consulté le 5 janvier 2025)

²² Le biais « d'automatisation » est une trop grande confiance accordée à un mécanisme automatisé. Le biais « d'ancrage » est relatif à la difficulté à se départir d'une information fournie antérieurement à une décision émanant des algorithmes.

²³ B. Louvel, « La justice prédictive », site Internet de la Cour de cassation, 12 février 2018, accessible sur <https://www.courdecassation.fr/toutes-les-actualites/2018/02/12/bertrand-louvel-la-justice-predictive> (consulté le 4 janvier 2025).

²⁴ A. Rouvroy et T. Berns, « Gouvernamentalité algorithmique et perspectives d'émancipation », *Réseaux* 2013/1, n°177, 2013, pp. 163-196

²⁵ P. Gueydier, *Pouvoir régalién et algorithmes, vers l'algocratie ?*, Optic, 2018.

²⁶ Sur la notion « d'État des algorithmes », voir Y. Meneceur, *L'intelligence artificielle en procès – Plaidoyer pour une réglementation internationale et européenne de l'intelligence artificielle*, Bruylant, 2020, pp. 227-241.

²⁷ Sur le développement de « l'IA » dans le contexte de cet hyperindividualisme, voir Y. Meneceur, « Intelligence artificielle, État de droit et œuvre de justice : regard prospectif sur l'évolution de la justice à l'ère de l'hyperindividualisme et de la transition numérique », *Lexbase*, 5 décembre 2024, accessible sur <https://www.lexbase.fr/article-juridique/112861575-dossier-special-decryptage-de-lia-appliquee-au-droit-intelligence-artificielle-etat-de-droit-et-uvre> (consulté le 5 janvier 2024).

Bibliographie sélective

Textes juridiques

CEPEJ, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, CEPEJ (2018)14, décembre 2018.

Convention-cadre (STCE n°225) du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit.

Directive (UE) 2024/2853 du Parlement européen et du Conseil du 23 octobre 2024 relative à la responsabilité du fait des produits défectueux et abrogeant la directive 85/374/CEE du Conseil.

Lignes directrices en matière d'éthique pour une IA digne de confiance rédigées par le groupe d'experts de haut niveau sur l'intelligence artificielle, 8 avril 2019.

Recommandation (OCDE/LEGAL/0449) du Conseil sur l'intelligence artificielle du 22 mai 2019 amendée le 3 mai 2024.

Recommandation de l'Unesco sur l'éthique de l'intelligence artificielle, adoptée en séance plénière le 23 novembre 2021.

Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n°300/2008, (UE) n°167/2013, (UE) n°168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle).

UNESCO, Draft Guidelines for the Use of AI Systems in Courts and Tribunals, CI/DIT/2024/GL/01, août 2024.

Ouvrages

BENSAMOUN A. & LOISEAU G. (2022) (dir.), *Droit de l'intelligence artificielle*, LGDJ, 2^e édition.

CASTETS-RENARD C. & EYNARD J. (2023) (dir.), *Un droit de l'intelligence artificielle – Entre règles sectorielles et régime général, perspectives comparées*, Bruylant.

G'SELL F. (2024), "Regulating under uncertainty: Governance options for generative AI", Stanford Cyber Policy Center, July.

LATILA. (2024), *Le droit du numérique – Une approche par les risques*, Dalloz, 2^e édition.

MENECEUR Y. (2020), *L'intelligence artificielle en procès – Plaidoyer pour une réglementation internationale et européenne*, Bruylant.

MENECEUR Y. (2024), *IA générative et professionnels du droit – Comprendre et s'approprier la langue des probables*, LexisNexis.

Articles et autres publications

« Quel impact de l'IA dans les directions juridiques ? », *Le Monde du Droit*, 23 juillet 2024.

ALBERT Ch. (2024), « Les interviews des acteurs de l'IA juridique : la position de l'éditeur Wolters Kluwer », *Village de la justice*, 18 juillet.

BARTHE E. (2024), « Intelligence artificielle en droit : derrière la "hype", la réalité », blog *Precisement.org*, mis à jour le 23 janvier.

DAHL M., MAGESH V., SUZGUN M. & HO D. E. (2024), "Large legal fictions: Profiling legal hallucinations in Large Language Models", *Journal of Legal Analysis*, 26 juin.

MENECEUR Y. (2024), « [Étude] de la justice prédictive à l'intelligence artificielle générative... mirages révolus et nouvelles ambitions d'une justice mathématisée », *Village de la Justice*, 28 février.

MENECEUR Y. (2024), « Intelligence artificielle, État de droit et œuvre de justice : regard prospectif sur l'évolution de la justice à l'ère de l'hyperindividualisme et de la transition numérique », *Lexbase*, 5 décembre.

ROUVROY A. & BERNS T. (2013), « Gouvernamentalité algorithmique et perspectives d'émancipation », *Réseaux* 2013/1, n°177, pp. 163-196.

L'office de l'avocat à l'heure du numérique et de l'IA : entretien avec M^e Christiane FÉRAL-SCHUHL

Propos recueillis par Arnaud MAZIER
Direction générale des Entreprises (DGE)

Bien que les outils numériques améliorent le quotidien matériel des avocats et facilitent leur relation avec les clients, un avocat ne peut se contenter d'être un utilisateur ordinaire de l'intelligence artificielle. Il doit respecter scrupuleusement les principes essentiels régissant la profession d'avocat ainsi que les principes fondamentaux de la justice. Certes, ces principes sont profondément enracinés dans la conscience professionnelle des avocats, mais ils peuvent être érodés par l'environnement numérique, parfois sans que l'avocat s'en rende compte.

L'intelligence artificielle, en particulier celle générative, offre indéniablement de nouvelles possibilités dans l'exercice de la profession d'avocat. Cependant, son intégration doit se faire dans le respect du droit continental, tout en maintenant la responsabilité pleine et entière de l'avocat. Les enjeux de formation sont cruciaux pour soutenir l'évolution de la profession.

Les questions ont été « raccourcies » afin de tenir compte des contraintes éditoriales

A. M. : Quel rôle joue le numérique dans les activités quotidiennes d'un avocat ?

M^e Féral-Schuhl : Pendant longtemps, l'avocat a investi dans des éléments de *standing* visant à rassurer le client : des locaux, une belle adresse, du personnel, etc. Aujourd'hui, l'atout principal pour un avocat est celui d'être joignable facilement et directement par son client. Ce dernier lui réclame son numéro de portable, son courriel, et il n'hésite pas à échanger par SMS ou encore par WhatsApp. Si le bouche-à-oreille reste important, le client collecte, *via* les moteurs de recherche ou les réseaux sociaux, des informations sur son avocat. Il le suit sur les réseaux sociaux et se sent libre de commenter et de réagir en ligne.

Dans ce contexte, les avocats ont rapidement investi les réseaux sociaux, les blogs, les sites d'information, etc. Ces nouveaux moyens numériques leur permettent de se positionner sur le marché du droit, où chacun doit être en mesure de communiquer sur ses missions, ses compétences, son savoir-faire, son offre en matière de prestations juridiques, mais aussi ses tarifs.

Dans leur quotidien, les avocats sont désormais connectés, et ils utilisent des bases de données, la visioconférence, les messageries et, de plus en plus souvent, les systèmes d'intelligence artificielle.

Pour autant, l'avocat n'est pas un utilisateur comme les autres. En effet, il convient de rappeler qu'il a l'obligation

Présentation de Maître Christiane FÉRAL-SCHUHL

Christiane Féral-Schuhl est avocate au Barreau de Paris et co-fondatrice du cabinet Féral, spécialisée dans le droit du numérique, des données personnelles et de la propriété intellectuelle. Elle est également médiatrice. En parallèle de son activité professionnelle, elle s'implique dans le domaine institutionnel en étant seconde vice-présidente du Conseil national de la Médiation, après avoir présidé le Conseil national des Barreaux et le Barreau de Paris.

Elle a aussi co-présidé la Commission parlementaire sur le droit et les libertés à l'âge du numérique avec le député Christian Paul, et siégé au Haut Conseil à l'égalité entre les femmes et les hommes ainsi qu'au Conseil Supérieur des tribunaux administratifs et des cours d'appel administratives. Investie dans la formation, elle a été marraine de promotion d'élèves avocats.

Maître Féral-Schuhl est l'auteure de plusieurs ouvrages et articles, notamment *Cyberdroit : le droit à l'épreuve de l'Internet* (Dalloz, 2020) et *La protection des données personnelles* (Dalloz, 2019). Elle a également co-écrit *La médiation, à l'épreuve de la pratique avec Michèle Jaudel* (Dalloz, édition 2025/2026), ainsi que *Cybersécurité, mode d'emploi avec Xavier Leonetti* (PUF, 2022). Plus récemment, elle a collaboré avec l'illustratrice Tiphaine Mary sur une bande dessinée intitulée *Adélaïde. Lorsque l'Intelligence Artificielle casse les codes* (Dalloz, 2024).

de respecter les principes essentiels qui gouvernent la profession d'avocat, et les principes fondamentaux de la justice. Or, si ces principes sont bien ancrés dans l'esprit des avocats, ils peuvent être battus en brèche du fait de l'environnement numérique, parfois à l'insu de l'avocat lui-même.

Prenons comme exemple les réseaux sociaux. L'avocat est tenu d'être loyal à l'égard de son client, mais aussi à l'égard de la partie adverse¹. Cette obligation de loyauté contraint l'avocat à une vigilance permanente, car ses comptes peuvent être passés au crible par ses clients, ses prospects et, plus largement, son réseau de relations. Il ne faut pas sous-estimer le sentiment de trahison que pourrait éprouver un client qui découvrirait sur un réseau social les liens de son avocat avec l'avocat adverse ou encore avec le directeur juridique de la société adverse, son principal concurrent.

De même, les rapports entretenus avec les magistrats peuvent interpeller alors qu'une affaire est en cours. L'avocat est libre, dans le cadre de sa vie privée, de nouer des relations avec les personnes de son choix, et il n'est évidemment pas répréhensible en soi de côtoyer des magistrats. Pour autant, le fait d'afficher ces relations publiquement peut poser certaines difficultés, à tout le moins un ressenti négatif chez les justiciables concernés.

De la même manière, l'avocat est tenu de respecter des principes de prudence, modération, délicatesse et courtoisie. À ce titre, il doit s'exprimer en évitant les propos polémiques ou indéliçables² qui peuvent non seulement nuire à sa réputation et son image, mais également rejaillir sur la profession tout entière. Il doit donc prêter attention aux contenus qu'il poste et à ses prises de position non mesurées. L'avocat doit en particulier « faire preuve de prudence et de modération dans ses propos, et en aucun cas d'animosité personnelle, ce qui exclut toute attaque *ad hominem* »³. Il doit aussi, en toutes circonstances, « observer les devoirs que lui imposent les règles, traditions et usages professionnels, notamment envers les magistrats, les membres du barreau et les clients [...] » et « en toutes circonstances, faire preuve de délicatesse, particulièrement lorsque sa qualité d'avocat est connue [...] »⁴.

Or, non seulement, nous observons une inflation des *posts* et des commentaires virulents publiés par des avocats à propos de leurs clients, de leurs adversaires ou encore des magistrats, mais nous pouvons de plus craindre une violation du secret professionnel. Ce dernier a un caractère « général, absolu et illimité »⁵ et ce « quels qu'en soient les supports, matériels ou immatériels ». La ligne rouge peut être vite franchie, car, même lorsque l'avocat ne dévoile pas le nom de son client ou ne communique pas des données nominatives sur un dossier ou sur son agenda, certaines fonctionnalités peuvent livrer des informations

confidentielles. La géolocalisation, par exemple, souvent paramétrée par défaut, permet de suivre les déplacements de l'avocat, de situer l'endroit où il est, d'opérer des recoupements qui peuvent ainsi conduire à identifier une adresse, le nom d'une société, l'adresse du client⁶, etc., et, ainsi, à dévoiler tout ou partie d'un secret. De même, l'utilisation de l'intelligence artificielle ouvre de nouvelles zones de risques, notamment avec les systèmes ouverts qui peuvent favoriser la fuite de données confidentielles, en mutualisant les réponses apportées.

Les cabinets d'avocats, à l'image des entreprises, peuvent ainsi être la cible de cyberattaques pour des causes diverses, qu'il s'agisse d'espionnage, de vengeance ou encore de négligence.

Un dispositif de protection trop léger, une utilisation sans encadrement d'objets connectés, un recours à des WiFi publics non sécurisés ou encore un manque de formation des collaborateurs aux règles essentielles de l'« hygiène informatique » sont autant de causes à l'origine des vulnérabilités constatées. À titre d'exemple, un cabinet new-yorkais a été victime d'une attaque de rançongiciel⁷. Après avoir refusé de payer la rançon de 42 millions de dollars exigée par les hackers, les données confidentielles des clients ont été rendues publiques, à l'instar de celles concernant la chanteuse et actrice Lady Gaga, occasionnant un très lourd préjudice économique et « réputationnel » au cabinet. Par ailleurs, dans l'affaire des Panama Papers, le cabinet d'avocat, victime de la fuite de 11,5 millions de documents provoquée par un lanceur d'alerte anonyme, n'a pas survécu et a dû cesser ses activités deux ans plus tard !

Conscient de ces risques, le Conseil national des barreaux a publié plusieurs guides pratiques, notamment l'« Utilisation des systèmes d'intelligence artificielle générative⁸ » et « Les avocats et le Règlement général sur la protection⁹ ». Ils ont vocation à accompagner et à sensibiliser les avocats aux risques précités et aux mesures préventives à prendre.

Parallèlement, il est intéressant de constater que les clients n'hésitent plus à exiger de leur avocat des mesures de cybersécurité pour garantir la confidentialité de leurs dossiers et le secret professionnel. Il n'est pas rare de découvrir, dans les appels d'offres, des questions ciblées sur les outils technologiques de l'avocat.

⁶ Par exemple : Cons. const., 3 déc. 2021, n°2021-952 QPC, M. Omar Y.

⁷ Le rançongiciel (*ransomware*) est un logiciel malveillant qui empêche l'accès aux données stockées sur un ordinateur et propose leur récupération contre le paiement d'une rançon. Ce logiciel prend en otage les données présentes sur un ordinateur en les chiffrant. Une fois les données rendues inaccessibles pour l'utilisateur, le pirate exige le paiement d'une rançon en contrepartie de la livraison de la clé de déchiffrement.

⁸ https://encyclopedie.avocat.fr/GEIDFile/CNB_GT_IntelligenceArtificielle_2024.pdf?Archive=132021395020&File=Telecharger%5Fle%5Fguide%5Ffici&verif=480312480317473152480453469030450538480306470024488826479274

⁹ https://encyclopedie.avocat.fr/GEIDFile/CNB_20230524_guide-RGPD-2023%5bA-K%5d.pdf?Archive=131511395979&File=Telechargez%5Fle%5Fguide%5Fpratique%5Ffici&verif=480312480317473152480453469030450538480306470024488824480274

¹ *Règles de la profession d'avocat*, Dalloz Action, 16^e édition, 2018-2019, n°332.31

² Avis n°183/20.1658, 25 mai 2010.

³ *Règles de la profession d'avocat*, Dalloz Action, 16^e édition, 2018-2019, n°336.23.

⁴ Règlement intérieur du Barreau de Paris, art. P 10.01.

⁵ *Règles de la profession d'avocat*, Dalloz Action, 16^e édition, 2018-2019, n°332.31.

Les algorithmes et l'IA prédictive peuvent-ils aider à analyser des données pour détecter des tendances, enrichir une défense, identifier des failles ou évaluer la pertinence d'un contentieux ?

Certainement.

D'ailleurs, l'intelligence artificielle n'est plus une option pour les avocats.

Pour autant, l'avocat reste maître du jeu et responsable de l'utilisation des réponses apportées par l'intelligence artificielle, d'autant que celle-ci peut se tromper. Nous en avons eu plusieurs illustrations, notamment avec la mésaventure d'un avocat américain qui a été berné par ChatGPT il y a quelques mois à peine. Il est devenu célèbre pour avoir cité, à l'appui de son argumentaire, six cas de jurisprudences imaginaires proposés par ChatGPT. Il a été sanctionné d'une peine d'amende de 5 000 dollars¹⁰. Il y a également l'histoire de cette avocate qui, en Colombie-Britannique, a engagé une action judiciaire sur le fondement de deux jurisprudences inventées par ChatGPT, qu'elle n'avait pas vérifiées. Comme la partie adverse lui a réclamé ces jurisprudences introuvables, l'avocate de la requérante a dû admettre son erreur devant le juge. Elle a été condamnée à payer les frais de la partie adverse¹¹.

Il appartient donc à l'avocat, utilisateur de l'intelligence artificielle, de vérifier les sources, et il ne peut certainement pas considérer avoir été induit en erreur. En ce sens, la Cour d'appel du Québec, dans un avis en date du 8 août 2024¹², invite les plaideurs à faire preuve de prudence lorsqu'ils incluent, dans leurs observations écrites ou orales des sources juridiques trouvées ou des analyses réalisées à l'aide de l'intelligence artificielle. Elle leur enjoint d'« appuyer leurs observations sur des textes législatifs ou réglementaires, de la jurisprudence et de la doctrine provenant de sources fiables tels les sites *web* des tribunaux et les éditeurs ou banques de données reconnus ». L'avocat est donc responsable de l'exactitude de ses observations écrites et orales. S'il utilise l'intelligence artificielle pour les préparer, il doit vérifier et contrôler l'information fournie. Il lui revient d'effectuer les recoupements nécessaires avec des banques de données fiables, afin de s'assurer que les références et leur contenu résistent à un examen minutieux et contribuent à maintenir l'intégrité et la fiabilité de notre système de justice.

Nous pouvons aussi nous interroger sur la responsabilité de l'avocat qui n'utiliserait pas l'intelligence artificielle. En ce sens, une décision de la Cour supérieure

de justice de l'Ontario en date du 22 novembre 2018¹³ retient l'attention. Dans cette affaire, le juge canadien a refusé d'allouer à la demanderesse qui obtenait gain de cause la somme de 900 dollars au titre des recherches juridiques effectuées pour préparer le dossier. La Cour a considéré que l'avocat aurait pu réduire le temps passé aux recherches juridiques s'il avait utilisé un moteur de recherche équipé d'intelligence artificielle. Si une telle décision paraît peu probable en France, elle soulève certaines interrogations, notamment l'appréciation du coût de l'avocat dès lors qu'il existe de plus en plus d'outils pour réduire le temps de recherche qu'il investit.

Sur le terrain de la responsabilité, nous pouvons en outre nous questionner sur la responsabilité de l'avocat qui ne suivrait pas les pistes proposées par l'intelligence artificielle. Est-ce qu'un client pourrait lui reprocher de s'être fourvoyé dans un argumentaire en rupture avec les « précédents » ? La même question se pose pour le juge qui opérerait un revirement de jurisprudence.

L'IA générative s'installe progressivement dans nombre de tâches intellectuelles. Est-elle une disruption pour l'office de l'avocat ?

L'intelligence artificielle générative peut certainement offrir aux avocats une assistance de qualité pour la résolution de problèmes, la reconnaissance de motifs, la prise de décisions ou encore l'analyse de données et l'automatisation de rapports, telles que pour les « dues diligences ». Elle permet de générer des textes académiques, des résumés d'articles et de synthèses de données. Elle facilite le travail d'étude des cas et de création des scénarios et hypothèses, des visuels, des vidéos ou des documents audios. Elle peut même répondre très rapidement à des questions simples et donc à la satisfaction des clients. Elle ouvre à l'évidence de belles perspectives à la profession. Les avocats peuvent optimiser leur temps, élargir leurs compétences et ainsi se recentrer sur des tâches à valeur ajoutée.

Cependant, il faut être vigilant sur les risques liés aux biais. À titre d'illustration, avec le décret DataJust du 27 mars 2020 (projet désormais abandonné), la justice avait prévu le développement d'un dispositif algorithmique permettant de recenser, par type de préjudice, les montants alloués aux victimes en indemnisation de leur préjudice corporel dans les décisions de justice rendues en appel par les juridictions administratives et judiciaires sur la période s'étalant de 2017 à 2019. Certains n'y ont vu qu'un référentiel indicatif d'indemnisation des préjudices corporels. Selon eux, il s'agissait seulement d'un support à la décision des juges, en ayant recours à l'intelligence artificielle pour évaluer les préjudices corporels. Mais la profession

¹⁰F. Bayard, « Cet avocat a utilisé ChatGPT... et s'en est mordu les doigts », 01net, 29 mai 2023 : <https://lext.so/3sLEJp>.

¹¹Legalnews, « Sanction pour une avocate ayant cité des cas fictifs générés par ChatGPT », *Le monde du droit*, 11 mars, 2024 : <https://lext.so/fJ7ZoE>.

¹²https://courdappelduquebec.ca/fileadmin/Fichiers_client/Procdures_et_avis/Liste_des_avis/Avis_utilisation_intelligence_artificielle_-_FR_-_13_aout_2024.pdf

¹³Aurore-Angélique Hyde, « Avocat et intelligence artificielle : quelles obligations, quelles responsabilités ? Réflexions à partir d'une décision de la Cour supérieure de justice de l'Ontario du 22 novembre 2018 », *Recueil Dalloz*, 2019, p. 2107.

d'avocat, par la voie du Conseil national des barreaux¹⁴ (CNB), s'est mobilisée sur le sujet. Le CNB soulevait tout particulièrement une suspicion de biais discriminant, car les femmes victimes de préjudices corporels pourraient être moins indemnisées que les hommes victimes. En effet, leurs salaires sont statistiquement moins élevés. Comment, dès lors, le juge qui s'appuie sur cette base de données intelligente pourrait-il corriger ce biais si l'information n'est pas portée à sa connaissance, ou s'il n'en a pas conscience ?

Ce n'est pas une hypothèse d'école puisqu'il suffit de rappeler que les populations afro-américaines ont identifié un biais racial dans le logiciel COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), qui est utilisé aux États-Unis par une soixantaine de juridictions. Cet outil, au moyen de 137 questions très différentes, évalue la probabilité de récidive d'un accusé à partir d'un système de notation allant de 1 à 10.

Mais nous pouvons aussi considérer que l'intelligence artificielle est un danger pour les professions juridiques, car elle peut faire perdre la dimension humaine de l'exercice professionnel. Nous pouvons rappeler que, au-delà de la compétence et de la connaissance juridique, il faut de l'expérience et du savoir-être pour maîtriser la procédure judiciaire et pour appréhender la singularité du dossier que l'on défend.

On peut aussi dénoncer le risque de voir voler en éclats les garanties protectrices du justiciable : le secret professionnel, la confidentialité, le respect du contradictoire, la formation professionnelle obligatoire, etc.

Les bases de connaissances juridiques sont des outils essentiels pour chaque avocat. Comment l'IA révolutionne-t-elle cette industrie ?

Le principal risque, c'est de perdre l'approche de droit continental qui caractérise le système juridique français. Le recours à l'intelligence artificielle pourrait conduire à un glissement de cette construction continentale du droit vers un système de *common law*, d'autant que la justice prédictive est portée par des éditeurs juridiques à dimension internationale, souvent de culture anglo-saxonne.

Nous devons donc nous assurer que la logique décisionnelle contenue dans les algorithmes sera conçue selon notre modèle juridique : le droit continental diffère en effet de la *common law* en ce que la règle de droit l'emporte sur la jurisprudence. Or, avec la démocratisation des outils de justice prédictive, nous pouvons légitimement craindre un glissement vers un « principe du précédent ».

¹⁴ Le Conseil national des barreaux (CNB) a pris position le 3 avril 2020 pour dénoncer le décret n°2020-356 du 27 mars 2020 portant création d'un traitement automatisé de données à caractère personnel (DataJust) et pour demander d'être associé aux travaux de conception de l'algorithme afin d'en garantir la transparence et l'absence de biais.

Comment voyez-vous l'évolution de la formation des avocats avec l'essor des outils numériques et de l'IA ?

Il est impératif de faire évoluer la formation et de l'adapter aux nouvelles générations d'étudiants et d'élèves avocats. Il faut s'assurer que, dans dix ou vingt ans, lorsque des générations d'avocats auront été formées exclusivement avec des outils d'intelligence artificielle, un avocat sera à même de faire une analyse critique et de « contrer » l'algorithme.

Il nous faut donc former les jeunes futurs avocats – mais aussi les magistrats – afin qu'ils puissent challenger l'intelligence artificielle, qui fournit des réponses en bloc, sans favoriser l'ouverture de la réflexion. Il faut ici s'assurer que ces avocats en devenir et magistrats soient en mesure de provoquer des revirements de jurisprudence, à rebours du robot dont les décisions risquent de se conformer à une justice passéiste et figée.

Je perçois néanmoins un risque, celui de voir les cabinets qui s'équipent d'intelligence artificielle tentés de contourner le recrutement des jeunes générations. Or, la transmission du savoir et des usages doit se faire dans les premières années d'exercice.

L'utilisation de l'IA pour améliorer la qualité et la réactivité dans les travaux de la DGFIP : avancer et innover par vents budgétaires contraires

Par Esther MAC NAMARA

Conseillère référendaire à la Cour des comptes

et Thomas BINDER

Responsable du département IA de la Délégation à la Transformation numérique de la DGFIP

La direction générale des Finances publiques (DGFIP) intègre l'intelligence artificielle (IA) dans de nombreux processus, par exemple pour optimiser la gestion fiscale, lutter contre la fraude et améliorer la gestion comptable. Depuis 2016, elle utilise le *data mining* pour cibler les fraudes, et déploie des solutions comme le projet « Foncier innovant » pour détecter les constructions non déclarées. L'IA générative (IAG), bien que prometteuse, rencontre encore des limites techniques et organisationnelles.

En 2024, la DGFIP a élaboré une première feuille de route stratégique visant à intégrer l'IAG dans ses processus de manière sécurisée. Pour 2025, elle prévoit de mettre en place une gouvernance stratégique et des formations pour accompagner son déploiement, tout en optimisant l'emploi des ressources.

Introduction

La direction générale des Finances publiques (DGFIP) joue un rôle central dans la gestion des finances publiques en France, allant de la collecte des impôts à la gestion des dépenses publiques. C'est une administration tournée vers la donnée, qui a rapidement et massivement investi sur le potentiel de l'informatique pour améliorer l'exercice de ses missions. Ses 5 000 informaticiens maintiennent et développent près de 800 applications¹.

Les technologies d'intelligence artificielle fascinent par leur caractère anthropomorphique. Elles correspondent à une forme additionnelle d'automatisation, comme l'ont été d'autres technologies avant elles, et qui restent pertinentes (algorithmie simple, RPA², etc.). L'intelligence artificielle générative (IAG) bouleverse déjà les métiers

de la rédaction et du droit, mais également ceux de la programmation informatique ou de la data science.

Le recours à l'intelligence artificielle répond aux trois enjeux principaux auxquels la DGFIP fait face : premièrement, une exigence de qualité, de réactivité, de simplicité des services publics requise par les citoyens, les entreprises et les administrations partenaires ; deuxièmement, la participation indispensable de la DGFIP à l'effort de redressement des finances publiques, avec une gestion plus efficace des dépenses et des recettes fiscales ; et enfin, un enjeu d'attractivité des talents dans un contexte marqué par un marché du travail dynamique, des départs à la retraite nombreux et un recours massif par les jeunes générations aux technologies d'IAG.

La DGFIP a investi rapidement dans ces nouvelles possibilités. Mais ses investissements s'inscrivent dans un cadre budgétaire contraint qui l'oblige à l'optimisation des moyens engagés. L'intégration de l'IA s'inscrit également dans le contexte de la doctrine informatique de la DGFIP, marquée par une forte internalisation de

¹ Elle met à disposition des autres administrations des infrastructures critiques, comme Nubo, son *cloud* interne, ou ses données.

² *Robotic process automation* : il s'agit d'une technique d'automatisation des processus sur des logiciels grâce à des logiciels robots.

l'ensemble de la chaîne de valeur, qui s'explique par une attention marquée aux enjeux de cybersécurité, de protection du secret fiscal et une surface d'action qui lui permet d'amortir ses investissements et parfois de proposer ses capacités aux autres administrations.

Le déploiement du *data mining* et du *machine learning* par la DGFIP

Le déploiement du *data mining* et du *machine learning* ont démontré la capacité de la DGFIP à s'emparer de ces technologies depuis dix ans.

L'IA a transformé de nombreux processus, notamment grâce à sa capacité à analyser de grandes quantités de données, à automatiser des tâches répétitives et fournir des prédictions précises. Dans le domaine des finances publiques, les cas d'usage se sont multipliés.

Lutte contre la fraude fiscale

L'IA permet d'identifier les contribuables à risque et de cibler les actions de recouvrement de manière plus précise.

Depuis 2016, la DGFIP a recours au *data mining*, dans le cadre du programme « Ciblage de la fraude et valorisation des requêtes » (CFVR) créé en 2014 à titre expérimental, puis pérennisé en 2019. Ce programme, qui repose sur une infrastructure et une équipe dédiées, est mis en œuvre de façon centralisée. Il alimente le programme des équipes du contrôle fiscal à hauteur de 50 % de leur programmation de travail environ. Les équipes restent libres de suivre ou non les recommandations des algorithmes. En 2024, ce programme rapporte plus de 2,5 Mds d'euros.

Optimisation de la collecte fiscale

Initié en 2022, le projet « Foncier innovant » s'appuie sur l'IA pour automatiser l'exploitation des images aériennes publiques de l'IGN, afin de détecter des incohérences déclaratives dans le patrimoine foncier et automatiser la mise à jour graphique dans le plan cadastral, avec un premier cas d'usage sur les piscines et les bâtis.

Après un pilote réussi, le déploiement à l'échelle nationale de ce nouvel outil permet de collecter plus de 43 M€ par an en identifiant les piscines qui n'étaient pas déclarées. Les investissements engagés (24 M€), financés grâce au FTAP³, ont ainsi été amortis dès la première année de déploiement.

Détection des anomalies en matière comptable

À partir de 2018, l'introduction de systèmes d'IA dans les outils comptables a également permis d'optimiser le contrôle hiérarchisé de la dépense dans les processus comptables. Comme en matière fiscale, l'IA appuie le comptable en l'aidant dans la détection des

mandats à vérifier prioritairement, mais l'agent reste, *in fine*, responsable de ses actions.

Comme le souligne la Cour des comptes⁴, ces outils ont amélioré la productivité des actions menées sans réduire les effectifs associés à ces missions. Qualitativement, la structure des emplois s'est modifiée avec l'arrivée de *data scientists* et experts en IA, regroupés sous forme d'équipes *ad hoc*, mais introduits également de manière isolée dans les équipes métiers pour jouer le rôle d'interface avec les équipes spécialisées.

L'émergence de l'IA générative à partir de 2023

Plusieurs projets internes ont apporté la preuve de la valeur ajoutée de cette technologie, mais se sont heurtés à des limites matérielles et organisationnelles.

La DGFIP a installé en 2021 une délégation à la transformation numérique, directement rattachée à la direction générale⁵. Au sein de cette structure, des spécialistes de l'IA ont été chargés d'explorer le potentiel lié à cette technologie et de conduire des actions de sensibilisation et d'information.

Le modèle opérationnel mis en place en 2023 a été celui du programme d'immersion. Chaque membre de l'équipe disposait d'un portefeuille de structures métiers dans lesquelles il passait un jour par semaine pour créer des liens informels, mais aussi observer au quotidien les problèmes concrets et apporter de la valeur à chaque fois que possible.

À titre d'exemple, le projet LLaMandement, qui porte le résumé et l'affectation automatique aux services des amendements parlementaires, a mis à jour l'ampleur du potentiel de transformation des activités de la direction de la Législation fiscale, pour la DGFIP, mais également pour l'ensemble des administrations centrales.

Ces expérimentations ont également révélé les limites de ces modèles : les « hallucinations » n'autorisant pas simplement par exemple d'en faire un chat à la disposition immédiate des usagers, en toute confiance, sans la relecture experte des agents publics.

Le dispositif se heurtait néanmoins à plusieurs limites qui empêchaient la consolidation des projets exploratoires ou le démarrage de nouveaux travaux avec les métiers. Sur le plan technologique, la DGFIP ne disposait d'aucun moyen de mise à disposition sécurisée et pérenne de ses capacités d'intelligence artificielle. Elle ne disposait ni de l'infrastructure matérielle (GPU) ni des plateformes nécessaires au déploiement de services, quelles que soient l'échelle et la volumétrie utilisateurs considérées, du pilote jusqu'à une industrialisation.

⁴ <https://www.ccomptes.fr/sites/default/files/2024-10/20241022-S2024-1165-L-intelligence-artificielle-dans-les-politiques-publiques-exemple-du-MEF.pdf>

⁵ En 2024, la DTNum était dotée d'un effectif de 130 ETP et d'un budget de 30 M€.

³ Fonds pour la transformation de l'action publique.

Sur le plan humain, elle avait besoin de recruter des compétences opérationnelles en intelligence artificielle (MLOps et LLMOps) pour déployer et maintenir à l'échelle les technologies d'IA précédemment développées de manière exploratoire. Enfin, les nouveaux processus, normes et la gouvernance nécessaires à l'adoption de ces technologies nécessitaient d'être coconstruits avec toutes les parties prenantes pour une adoption maîtrisée, sécurisée et flexible de ces technologies.

Ne disposant ainsi pas des fondamentaux infrastructurels, humains ou la gouvernance pour la consolidation des projets IA, les équipes ne pouvaient pas industrialiser leurs réalisations, les empêchant par la même occasion de terminer leurs travaux pour démarrer de nouvelles expérimentations.

En 2024, la DGFIP intègre mieux l'IA avec son système d'information

En 2024, la DGFIP ancre sa première feuille de route stratégique, préparant un déploiement plus large de l'IA par une meilleure intégration avec son système d'information.

La DGFIP a défini en 2024 ses orientations stratégiques d'intégration de l'intelligence artificielle à large échelle dans ses processus informatiques et métiers. Cette feuille de route prévoit :

- la consolidation d'un processus d'innovation technologique avec et pour les services informatiques ;
- la mise à disposition aux équipes métiers d'outils, logiciels et plateformes en cohérence avec les expérimentations réussies ;
- l'accélération des expérimentations au plus proche des enjeux opérationnels pour identifier les vecteurs de valeur ajoutée les plus impactants ;
- la définition d'un parcours de sensibilisation et de formation des agents pour utiliser efficacement ces technologies.

Cette orientation a tout d'abord nécessité le recrutement et la mise en place d'une équipe pluridisciplinaire et intégrée de spécialistes de l'intelligence artificielle, capable d'accompagner les métiers sur toute la chaîne de valeur : immersion, formations, sensibilisation, ateliers d'idéation, développement de pilotes (preuves de concept), construction de plateforme et outils, mise à l'échelle. Elle inclut à ce jour des *data scientists*, développeurs, DevOps, MLOps, architectes techniques et responsables produits.

Sur le volet technique, la DGFIP s'est dotée d'un socle technologique à l'état de l'art, sécurisé et résilient incluant la puissance de calcul nécessaire au déploiement à l'échelle des technologies d'IA (GPU). Ce socle a permis de décupler les capacités d'expérimentation comme d'industrialisation de la DGFIP. Elle permet aux bureaux d'expérimenter et d'industrialiser en autonomie ou accompagnés par des experts, de nouveaux cas d'usage : assistance à la programmation informatique, fiabilisation des données, outillage bureautique

augmenté, croisement d'information intelligent, maintenance prédictive, recherche documentaire augmentée ou encore aide à la réponse aux demandes des usagers.

Construit sur cette plateforme, et en combinant l'IA générative avec de puissantes capacités de recherche documentaire (Retrieval Augmented Generation), un prototype d'assistant IA nommé Caradoc a été conçu par la DGFIP pour que ses agents accèdent plus rapidement aux informations dont ils ont besoin⁶.

Perspectives pour 2025

Au-delà des enjeux technologiques, le déploiement maîtrisé à plus large échelle nécessitera de mettre en place la structure de gouvernance qui régulera l'emploi des capacités, et veillera à la montée en compétence des agents.

Pour être acceptées, les technologies d'IA doivent être comprises de tous. Des actions de sensibilisation ont été initiées depuis 2023, mais restent parcellaires. Elles prennent la forme d'actions de formation de plusieurs heures, de conférences ou de « cafés IA » en lien avec les « champions du numérique », réseau interne d'ambassadeurs.

Ces premières actions devront être complétées par un plan de montée en compétences articulé autour de formations certifiantes pour permettre aux agents de la DGFIP qui le souhaiteraient d'acquérir les compétences techniques en lien avec les technologies d'IA et d'IA générative plus spécifiquement (en data science, LLMOps, juridique, etc.).

Ce dispositif devra adopter un rythme aussi rapide que celui des évolutions technologiques auxquelles nous assistons dans le domaine de l'IA générative, et embarquer la direction générale à travers un dispositif de gouvernance *ad hoc*.

Cette structure de gouvernance devra veiller à la bonne mutualisation des moyens au niveau de la direction générale, en évitant la construction de capacités spécifiques à telle ou telle direction. Elle devra simultanément prioriser les cas d'usage éligibles au déploiement de l'IA, en lien avec ses priorités stratégiques, dans un contexte de tension élevée sur les moyens et sur les investissements.

Un effort constant devra en parallèle être mené par les équipes techniques afin de limiter la taille des modèles appelés pour ajuster le plus finement possible la consommation énergétique aux besoins réels en évitant la sur-qualité.

Conclusion

L'utilisation de l'IA et l'introduction de l'IA générative représentent une opportunité nouvelle pour améliorer la qualité et la réactivité des travaux de la DGFIP. Les cas d'usage sont nombreux, les attentes politiques élevées, l'adoption par les agents rapide.

⁶ <https://gitlab.adullact.net/dgfip/projets-ia/caradoc>

Simultanément, la capacité d'investissement dans les machines (CPU et GPU) qui sont indispensables à cette technologie n'a jamais été si faible dans un contexte financier particulièrement tendu, et qui le restera encore de nombreuses années.

Plus que jamais cette situation impose une forme de sobriété dans l'emploi de ces ressources. Cela suppose en premier lieu de ne recourir à l'IAG que lorsqu'il a été démontré qu'aucune autre solution ne permettait de résoudre plus simplement les problèmes rencontrés. Le design, l'amélioration de la qualité des données, de leur circulation et de leur exploitation, l'automatisation, parfois simplement l'amélioration du fonctionnement des organisations, sont autant de leviers au potentiel immense et encore largement sous-exploités.

L'intelligence artificielle générative reste spectaculaire par sa vitesse de déploiement. Néanmoins, cette technologie reste encore à un stade peu mature dans son modèle économique comme dans ses capacités. Elle nous invite à un effort de pensée critique permanent pour préserver le lien de confiance entre l'État et les citoyens et réfléchir à l'optimisation des moyens. Les cas d'usage prometteurs hier, comme le *chat bot* à usage universel par les usagers, ne le sont plus aujourd'hui. En revanche, les possibilités d'appui aux équipes en lien avec les usagers semblent élevées (détection de l'intention et orientation des questions vers les bonnes équipes, préparation des réponses, résumé des demandes, multimodalité d'interactions, identification des signaux faibles de mécontentement, etc.).

La DGFIP dispose d'une taille qui lui permet de rentabiliser rapidement ses investissements en IA et de structurer une offre de services interne mutualisée. À plus long terme, la question de l'ouverture des services d'IAG aux autres administrations par l'intermédiaire du *cloud* Nubo pourrait être posée dans une logique de mutualisation de ses capacités industrielles et de son expertise.

L'IA pour la médecine de précision en oncologie

Par Thomas WALTER

Professeur de Mines Paris PSL, directeur du Centre de Bioinformatique (CBIO) Mines Paris PSL, Chef de l'équipe « Apprentissage statistique et modélisation des systèmes biologiques », Directeur adjoint de l'Unité « Oncologie Computationnelle (U1331) » de l'Institut Curie / INSERM / Mines Paris - Chaire à Paris Artificial Intelligence Research Institute (PRAIRIE)

Comment choisir le meilleur traitement possible pour un patient atteint d'un cancer ? C'est l'enjeu de la médecine de précision, qui vise à proposer un traitement personnalisé à chaque patient. Ce choix doit se baser sur des données qui caractérisent la tumeur et son micro-environnement.

L'intelligence artificielle nous fournit une technique capable de créer des modèles prédictifs à partir de données massives et hétérogènes, et ainsi prédire le traitement optimal pour chaque patient.

Ici, nous allons présenter les fulgurants progrès de l'intelligence artificielle au travers d'un exemple d'application pour la pathologie digitale, une technique essentielle pour le diagnostic en oncologie. L'émergence de concepts avancés tels que l'apprentissage par instances multiples, l'auto-supervision et des modèles de fondation (*foundation models*), a contribué à un nombre impressionnant de nouvelles méthodes et applications, encore impensables il y a quelques années. Par exemple, dans les cancers du sein et de l'ovaire, l'IA a permis de prédire le défaut de recombinaison homologue, un défaut de la réparation d'ADN, à partir de simples lames histologiques. Ces avancées ouvrent la voie à une meilleure personnalisation des traitements, rendant l'IA incontournable pour la médecine de demain.

Contexte : cancer et médecine de précision

En France, plus de 430 000 nouveaux cas de cancer ont été diagnostiqués en 2023, soit un doublement par rapport à 1990. Aujourd'hui, un homme a environ une chance sur deux de développer un cancer au cours de sa vie, contre une chance sur trois pour une femme. Contrairement à l'incidence, qui ne cesse d'augmenter, le taux de mortalité est en constante diminution. Cette évolution est en grande partie attribuable aux progrès réalisés dans le diagnostic, la prise en charge et les traitements innovants. Cependant, le cancer demeure la première cause de mortalité en France.

Ainsi, l'amélioration du diagnostic apparaît comme un levier essentiel pour accroître les chances de survie des patients. En effet, la détection précoce des tumeurs, associée à une spécification de leurs caractéristiques biologiques, permet non seulement d'optimiser les stratégies thérapeutiques, mais aussi de réduire les traitements inutiles ou excessivement agressifs. On parle alors de désescalade thérapeutique.

Dans ce contexte, la médecine de précision joue un rôle fondamental. La médecine de précision vise à adapter les traitements aux spécificités de chaque patient. Celles-ci incluent les caractéristiques génétiques et moléculaires de la tumeur, la composition de son micro-environnement, la présence ou l'absence de certains types cellulaires et

la caractérisation morphologique de la tumeur, et bien d'autres aspects encore. De fait, le nombre de paramètres mesurables ne cesse d'augmenter, couvrant un spectre de plus en plus large d'aspects biologiques et cliniques. Identifier les caractéristiques qui peuvent nous informer sur l'évolution de la maladie ou les chances de succès d'un traitement s'apparente alors à chercher une aiguille dans une botte de foin.

C'est ici qu'intervient l'intelligence artificielle (IA), capable d'identifier des liens statistiques subtils entre les données d'entrée (des données recueillies sur la maladie du patient) et les données de sortie (survie, réponse au traitement). L'objectif ne se limite donc pas à proposer un modèle prédictif, mais vise également à comprendre quels éléments – quelle mutation, quelle surexpression, quel phénotype cellulaire – ont permis d'aboutir à la bonne prédiction. Ces éléments sont alors appelés biomarqueurs.

La pathologie digitale

L'anatomopathologie est une discipline médicale qui étudie les altérations des tissus et des cellules afin de diagnostiquer et caractériser des maladies. En oncologie, elle repose essentiellement sur l'analyse microscopique de coupes de tissus colorées, issues de biopsies ou de pièces opératoires, afin d'examiner l'architecture tissulaire ainsi que les altérations induites par la maladie.

La pathologie digitale repose sur la numérisation des lames histopathologiques, transformant profondément la manière dont elles sont stockées, organisées, partagées, visualisées et annotées. Cette transition ouvre également la voie au développement d'algorithmes d'intelligence artificielle (IA) permettant l'analyse automatique de ces lames, avec trois objectifs.

Assistance au diagnostic

Certaines tâches en histopathologie sont particulièrement fastidieuses, comme le comptage de certains types de cellules. L'automatisation de ces tâches par l'IA permet un gain de temps considérable et une analyse plus exhaustive. Dans ce scénario, l'IA n'analyse pas les lames de manière totalement autonome : une quantification automatique est proposée en complément de la procédure de diagnostic manuelle, facilitant ainsi le travail du pathologiste. La détection des cellules en division (mitoses) illustre bien ce cas d'usage. Ce problème a fait l'objet de nombreux projets de recherche en vision par ordinateur (*computer vision*), avec plusieurs grandes compétitions internationales [1] et un large éventail de méthodes développées. Aujourd'hui, certaines entreprises proposent des solutions de comptage automatique des mitoses, rendant cette tâche plus efficace, reproductible et standardisée.

Analyse autonome des lames histopathologiques

Un scénario plus complexe se présente lorsque la lame est analysée de manière autonome par une IA : plutôt que de fournir une information en soutien à l'anatomo-pathologiste, l'IA délivrerait un résultat sans qu'aucun humain n'ait regardé la lame. Ce cas d'usage reste rare à l'heure actuelle, car les performances des algorithmes sont souvent loin de répondre aux exigences de robustesse requise. De plus, dans la majorité des cas, l'analyse d'une lame par un anatomo-pathologiste nécessite une compréhension approfondie de la physiologie et des mécanismes sous-jacents des maladies, une capacité que l'IA actuelle ne possède pas. Toutefois, certaines applications se prêtent mieux à une analyse autonome, notamment l'identification des métastases ganglionnaires. Cet examen, réalisé systématiquement, revêt une importance capitale pour le choix du traitement. Il s'agit d'une tâche particulièrement adaptée à l'IA, car elle ne requiert pas une compréhension approfondie des processus biologiques sous-jacents, mais consiste essentiellement à détecter la présence de régions tumorales dans une lame. Cette application a marqué l'un des premiers succès de l'IA en pathologie, avec l'organisation de la première grande compétition dédiée à la détection des métastases dans les ganglions lymphatiques [2]. Aujourd'hui, des solutions basées sur l'IA permettent l'analyse autonome des échantillons ganglionnaires, et sont proposées par des entreprises spécialisées.

Prédictions au-delà de l'automatisation du diagnostic

Dans ce dernier volet, l'IA ne se limite plus à reproduire les analyses humaines, mais vise à révéler des informations biologiques pertinentes pour la variable de sortie.

Par exemple, elle vise à prédire le sous-type moléculaire d'un cancer, ses mutations génétiques, le risque de rechute du patient, sa survie ou encore sa réponse au traitement. Souvent, les éléments morphologiques présents dans l'image qui permettent de telles prédictions ne sont pas encore connus. L'IA a donc un double objectif : proposer un modèle prédictif, tout en identifiant un ensemble de biomarqueurs morphologiques susceptibles d'expliquer pourquoi une prédiction a été faite. Un exemple notable est la prédiction de la HRD (*homo-logous recombination deficiency*) [3], qui sera détaillée plus loin.

Données et annotations

Tous les algorithmes d'IA nécessitent des données associées à une vérité terrain (*ground truth*) afin d'être entraînés. On remarque que, selon les différents scénarios, le type d'annotations permettant d'entraîner ces modèles varie considérablement.

Si l'objectif est de reproduire l'œil du pathologiste, par exemple pour détecter des cellules mitotiques ou des régions métastatiques, il faut s'appuyer sur une annotation manuelle des cellules ou des régions d'intérêt. Cela représente un effort considérable et donc un coût important en raison du temps et de l'expertise nécessaires.

En revanche, si l'on cherche à prédire une donnée de nature moléculaire (mutation, signature génétique, etc.), il est nécessaire de disposer de données expérimentales multimodales, associant à la fois des images histopathologiques et une caractérisation moléculaire issue d'analyses complémentaires.

Enfin, lorsqu'il s'agit de prédire une variable clinique, comme la survie du patient ou sa réponse au traitement, on cherche à anticiper un événement futur. Il est donc indispensable de suivre les patients sur une longue période, ce qui est complexe et coûteux. Heureusement, ces données sont déjà collectées dans le cadre d'essais cliniques, dont l'accès s'avère crucial pour développer ces modèles prédictifs.

Les solutions algorithmiques

Les algorithmes en pathologie digitale sont confrontés à plusieurs défis majeurs. Le premier concerne la taille des images, qui atteint souvent plusieurs milliards de pixels. Cela pose un problème technique majeur, car les cartes graphiques et les architectures de réseaux de neurones actuelles ne sont pas adaptées à de telles tailles d'images.

La solution consiste à décomposer la lame en un grand nombre de petites images, appelées « tuiles ». Chaque tuile est ensuite convertie en vecteur par un réseau de neurones pré-entraîné, permettant d'obtenir une représentation numérique de la tuile. Ces représentations sont ensuite agglomérées, et le vecteur résultant – qui constitue donc une représentation de la lame entière – est utilisé comme entrée pour un dernier réseau de neurones chargé de prédire la variable d'intérêt. Cette approche est appelée *multiple instance learning* (MIL) et illustrée dans la Figure 1.

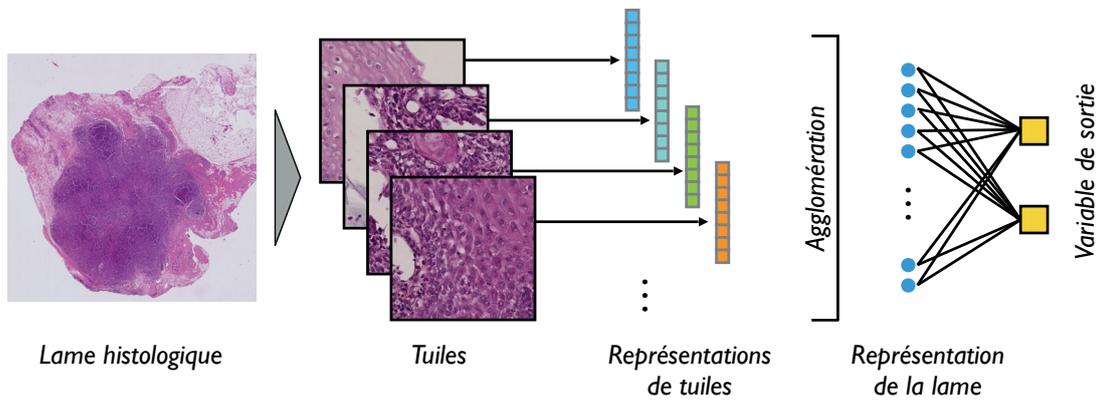


Figure 1 : *Multiple instance learning* appliqué à l'analyse automatique des lames histologiques.

L'un des principaux avantages de cette méthode est qu'elle ne nécessite pas d'annotations au niveau des tuiles, mais uniquement à l'échelle de la lame entière. Cela correspond parfaitement aux problématiques de prédiction de variables globales, telles que la survie du patient, sa réponse au traitement ou la présence d'une mutation tumorale.

Ces dernières années, la façon d'entraîner des modèles pour obtenir des représentations pertinentes a connu une évolution majeure. Avant 2020, les réseaux de neurones étaient généralement pré-entraînés sur des images naturelles sans lien avec l'histopathologie. L'introduction de l'apprentissage auto-supervisé a ensuite permis de développer des représentations mieux adaptées au domaine médical.

Cette avancée a conduit au développement des *foundation models*, qui ont eu un impact majeur en pathologie digitale. Ces modèles sont entraînés sur d'immenses bases de données d'images histopathologiques. Les publications récentes rapportent l'utilisation de milliards de tuiles pour apprendre des représentations génériques, puissantes et applicables à une grande diversité de tâches.

Ainsi, les *foundation models* représentent une avancée technologique majeure en pathologie digitale depuis 2022.

Une technique alternative que nous avons également développée au Centre de Bioinformatique de l'École des Mines (CBIO) dans le cadre de notre partenariat avec l'Institut Curie et l'Inserm, et désormais valorisée par une start-up, repose sur une approche conceptuellement différente. Plutôt que d'entraîner un modèle bout-à-bout, nous identifions et classons l'ensemble des cellules présentes dans une lame, puis calculons des descripteurs quantifiant leur abondance et leur distribution spatiale. Bien que cette approche soit plus laborieuse, elle présente l'avantage d'être interprétable par construction, offrant ainsi une meilleure compréhension des mécanismes biologiques sous-jacents.

Exemple : la prédiction du défaut de recombinaison homologue

Dans le diagnostic du cancer du sein et du cancer de l'ovaire, un aspect essentiel consiste à déterminer le

statut de la recombinaison homologue (*homologous recombination*, HR) de l'ADN, un indicateur du bon fonctionnement du processus de réparation de l'ADN. Un défaut de ce mécanisme de réparation dans les cellules tumorales entraîne une forte instabilité génomique, rendant ces cellules particulièrement sensibles aux inhibiteurs de PARP, un traitement prometteur, notamment dans les cancers du sein métastatiques.

Connaître le statut HR des cancers du sein et de l'ovaire est donc indispensable pour orienter le choix thérapeutique des patientes concernées. Cependant, les tests moléculaires permettant de le déterminer sont coûteux et leur mise en œuvre systématique chez toutes les patientes représente un véritable défi logistique. Cette approche est même inconcevable dans de nombreux pays en raison de contraintes économiques et d'infrastructures limitées.

À l'inverse, les lames histologiques sont utilisées en routine clinique et représentent un faible coût. Au CBIO et en collaboration avec l'Institut Curie, nous avons cherché à prédire le statut HR à partir des lames anatomopathologiques, en appliquant la méthode mentionnée précédemment et en corrigeant les biais induits par les différences de protocoles de production des lames. Cette approche nous a permis d'obtenir une précision élevée dans la prédiction de la HRD, avec une AUC¹ de 0,86. De plus, nous avons introduit une nouvelle méthode permettant d'identifier les régions tissulaires associées à la HRD, ce qui nous a conduits à découvrir plusieurs phénotypes morphologiques témoignant de ce défaut de réparation de l'ADN.

Nous avons ainsi démontré que la prédiction du statut HR des cancers du sein, une variable essentielle pour guider la thérapie, peut être réalisée à partir de lames histologiques tumorales. Cette étude ouvre de nouvelles perspectives, tant en recherche, pour mieux comprendre les effets biologiques de la HRD, qu'en clinique, pour évaluer l'utilité de cette approche et la rendre applicable en routine médicale.

Ces travaux sont le fruit d'une collaboration entre le CBIO et le service de pathologie de l'Institut Curie, dirigé par Anne Vincent-Salomon.

¹AUC : *Area Under the Curve* : indique la performance d'un algorithme de classification binaire. Une valeur de 1,0 indique un classifieur parfait.

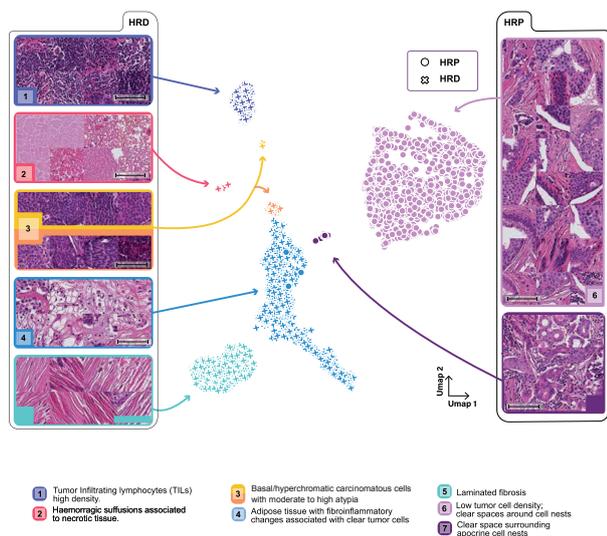


Figure 2 : Projection en deux dimensions des représentations vectorielles de tuiles sélectionnées par notre méthode [3]. Chaque tuile est représentée soit par une croix, si elle est associée à la HRD, soit par un cercle sinon. Les regroupements de tuiles correspondent à des morphologies similaires suivant un code couleur détaillé en bas de la figure (Source : [3]).

Défis actuels et intégration de données multimodales

Si les progrès en pathologie digitale ont été fulgurants ces dix dernières années, il reste encore de nombreux défis à relever.

L'un des principaux problèmes concerne les biais potentiels dans les données. En intelligence artificielle, un biais survient lorsqu'un modèle prédictif s'appuie sur des caractéristiques présentes dans les données qui ne sont corrélées à la variable de sortie que par la composition de l'ensemble d'apprentissage. Par exemple, si les lames des cas les plus sévères proviennent d'un hôpital différent, l'IA risque d'identifier des différences subtiles dans la préparation des lames, plutôt que des différences biologiques réellement liées à la gravité du cancer. Identifier et corriger ces biais est donc une priorité.

Un problème connexe est celui du *domain shift*, qui survient lorsque les données sur lesquelles un modèle a été entraîné suivent une distribution différente de celles sur lesquelles il est appliqué. En pathologie, ce problème apparaît, par exemple, lorsqu'un modèle entraîné sur des images acquises avec un type de scanner doit être appliqué dans un autre hôpital avec un scanner différent. L'adaptation de domaine est aujourd'hui un axe de recherche très dynamique, essentiel au développement de solutions robustes en pathologie digitale.

Enfin, la pathologie digitale ne couvre qu'un aspect d'une maladie. Les médecins prennent généralement leurs décisions sur la base d'un ensemble d'examen variés, couvrant différentes modalités. De la même manière, les approches en IA devraient intégrer plusieurs types de données, comme les résultats d'imagerie médicale

(scanners, IRM), le séquençage ADN ou encore les mesures d'expression génique. La fusion de ces différentes sources d'information représente un défi majeur, mais également une opportunité considérable.

En collaboration avec l'Institut Curie, nous avons publié une étude visant à prédire la réponse à l'immunothérapie dans le cancer du poumon, démontrant que la combinaison de plusieurs modalités conduit à des modèles plus performants [4]. Pour l'instant, cette approche est limitée par la taille des jeux de données, car le nombre de patients pour lesquels on dispose d'informations complètes reste faible. Cependant, cette perspective est très prometteuse, avec le potentiel d'améliorer encore davantage la prédiction de la réponse aux traitements.

Bibliographie

[1] VETA M., DIEST P. J. V., WILLEMS S. M., WANG H., MADABHUSHI A., CRUZ-ROA A., GONZALEZ F., LARSEN A. B. L., VESTERGAARD J. S., DAHL A. B., CIREŞAN D. C., SCHMIDHUBER J., GIUSTI A., & LUCA M. (2014), "Assessment of algorithms for mitosis detection in breast cancer histopathology images", *Medical Image Analysis*, pp. 1-23.

[2] EHTESHAMI BEJNORDI B., VETA M., JOHANNES VAN DIEST P., VAN GINNEKEN B., KARSSMEIJER N., LITJENS G., VAN DER LAAK J. A. W. M. & THE CAMELYON16 CONSORTIUM (2017), "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer", *JAMA*, 318(22), pp. 2199-2210, <https://doi.org/10.1001/jama.2017.14585>

[3] LAZARD T., BATAILLON G., NAYLOR P., POPOVA T., BIDARD F.-C., STOPPA-LYONNET D., STERN M.-H., DECENCIÈRE E., WALTER T., & VINCENT-SALOMON A. (2022), "Deep learning identifies morphological patterns of homologous recombination deficiency in luminal breast cancers from whole slide images", *Cell Reports Medicine*, <https://doi.org/10.1016/j.xcrm.2022.100872>

[4] CAPTIER N., LEROUSSÉAU M., ORLHAC F., HOVHANNISYAN-BAGHDASARIAN N., LUPORSI M., WOFF E., LAGHA S., SALAMOUN FEGHALI P., LONJOU C., BEAULATON C., ZINOVYEV A., SALMON H., WALTER T., BUVAT I., GIRARD N. & BARILLOT E. (2025), "Integration of clinical, pathological, radiological, and transcriptomic data improves prediction for first-line immunotherapy outcome in metastatic non-small cell lung cancer", *Nature Communications*, 16(1), article 614, <https://doi.org/10.1038/s41467-025-55847-5>

Les applications de l'IA pour l'aide à la radiologie

Par Jérôme KNOPLIOCH et Nicolas GOGIN
GE Healthcare

L'imagerie médicale est un domaine phare pour l'intelligence artificielle. Parmi les 1 016 dispositifs médicaux « intelligents » recensés par la FDA (agence américaine de certification), 777 concernent la radiologie. Les applications couvrent toute la chaîne depuis la reconstruction des images, l'assistance au diagnostic – pour des spécialistes cliniques en sous-effectif et surchargés –, jusqu'à la communication avec les patients.

Grâce aux investissements importants dans l'IA, il devient plus facile d'atteindre rapidement des performances cliniques supérieures aux méthodes traditionnelles de traitement d'image. L'entraînement des algorithmes reste cependant coûteux, entre la collecte de données et leur annotation. Pour que l'IA améliore vraiment leur productivité, les utilisateurs demandent une meilleure intégration dans le flux de travail et une fiabilité garantie. La fusion des images avec l'ensemble du dossier clinique pourra construire de nouveaux biomarqueurs, étendant la valeur prédictive de l'imagerie.

Introduction

En 2016, G. Hinton prédisait l'extinction des radiologues dans les cinq ans à venir. Cette vue est maintenant remplacée par un consensus exprimé par C. Langlotz, président de la Société de radiologie nord-américaine (RSNA) : « L'IA ne remplacera pas les radiologues, mais les radiologues utilisant l'IA remplaceront ceux qui ne l'utilisent pas ».

La radiologie représente 777 des 1 016 dispositifs médicaux incorporant de l'IA approuvés par la FDA (agence américaine de certification). Cet enthousiasme s'explique par le succès de l'imagerie médicale, l'une des premières spécialités médicales numérisées et qui crée toujours plus de données. Un scanner moderne produit des centaines d'images en quelques secondes. Chacune peut être critique pour le diagnostic. Parallèlement, le vieillissement des populations, l'amélioration de la prévention et de la prise en charge du cancer, l'émergence de nouvelles régions, entraînent une demande croissante d'imagerie.

La charge des radiologues a décollé dans les quinze dernières années selon de nombreuses métriques, entraînant un stress difficilement supportable. Ceci explique leur intérêt pour des aides à la productivité, de la lecture des images à la communication avec les patients.

Nous aborderons quelques applications et éléments techniques de l'IA en radiologie.

Améliorer l'image : reconstruction, super-résolution

L'amélioration des images a été l'une des premières applications de l'IA, prolongeant les progrès récents

de méthodes avancées de traitement du signal. Pour le scanner, l'IA facilite la lecture avec des doses faibles de radiation ou pour des patients en surpoids. Pour l'IRM, ces améliorations permettent des temps d'examen réduits et des performances meilleures pour des machines moins coûteuses.

L'IA est entraînée par l'association d'images de haute qualité et de leurs versions dégradées par l'introduction de bruit ou d'artefacts ou de conditions techniques inférieures. On utilise les signaux directs des capteurs, avant reconstruction, ce qui permet un meilleur contrôle de la qualité du modèle entraîné. La sélection des types d'images et d'artefacts ainsi que la mesure précise de l'écart entre l'image de référence et l'image reconstruite donnent un cadre rigoureux pour valider ces algorithmes.

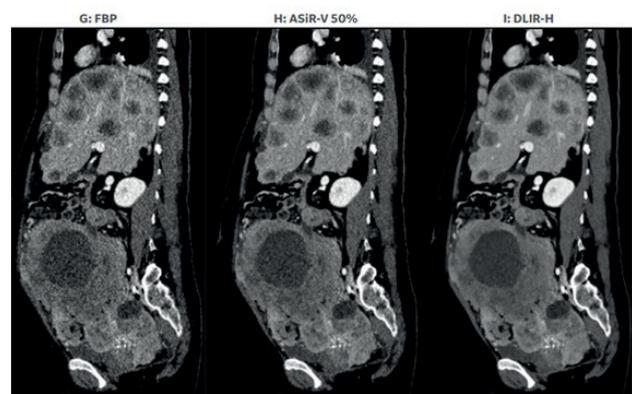


Figure 1 : Exemple d'images scanner et impact de la technologie TrueFidelity™ sur la qualité de l'image (Source : GE Healthcare White Paper).

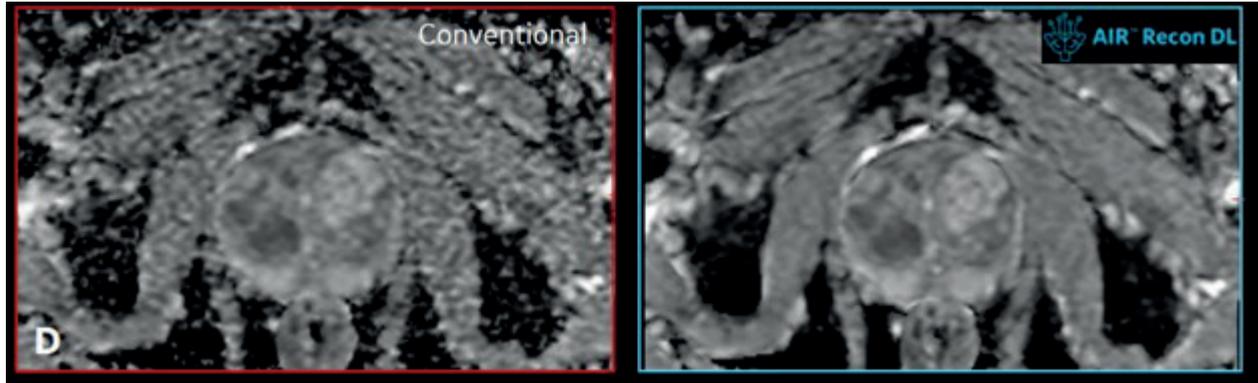


Figure 2 : Image IRM pondérée en diffusion de la prostate avec et sans la technologie AIR™ Recon DL (Source : GE Healthcare White Paper).

Assister la détection des pathologies

La détection assistée ou automatique de pathologies (*computer-assisted detection*) est l'application la plus souvent citée. Ces logiciels s'insèrent dans des pratiques de dépistage (cancer du sein, nodules pulmonaires, cancer de la prostate), ainsi que des contextes d'urgence (pneumothorax, fractures, AVC).

La détection doit être assez fiable (peu de faux négatifs / positifs) pour être acceptée par les radiologues. L'IA est sensiblement plus performante que le traitement d'image classique.

Les résultats les plus solides ont été obtenus dans le domaine du dépistage du cancer du sein en mammographie. Plusieurs études sur de nombreux patients (Hernström V., Josefsson V., Sartor H. *et al.*, 2025) ont montré que l'utilisation d'un algorithme d'IA améliore significativement la détection des cancers et réduit la charge de travail des radiologues. Dans le domaine du cancer du poumon, les premiers algorithmes de détection de nodule pulmonaire souffraient de problèmes de spécificité, car ils s'appuyaient sur des techniques de reconnaissance de forme assez basiques (Cao, 2020) sans prendre en compte des caractéristiques images plus complexes et parfois plus rares, ce que les algorithmes reposant sur l'apprentissage profond savent très bien faire aujourd'hui lorsqu'ils sont entraînés sur des milliers d'examen (Gao, 2024).

La détection doit généralement encore être confirmée par des radiologues, mais il est vraisemblable qu'un usage autonome sera possible dans le futur, notamment lorsque des radiologues qualifiés ne sont pas disponibles (zones d'accès difficile, pays émergents).

Un point important d'attention est l'évaluation du gain réel de productivité : l'IA est mal acceptée si elle induit une charge supplémentaire par des alarmes ou informations inutiles, parfois aggravée par un manque d'intégration avec les outils habituels des opérateurs.

L'apprentissage profond

Aujourd'hui, la grande majorité des algorithmes d'intelligence artificielle embarqués dans les dispositifs médicaux repose sur l'apprentissage profond (*deep learning* en anglais). Cette technique a pris son essor en 2012

avec les résultats remarquables obtenus par le réseau AlexNet (Krizhevsky, 2012) dans le problème de la classification d'image.

Le réseau AlexNet – dit convolutif – s'appuie sur des couches de convolution intercalées avec du sous-échantillonnage (*Max pooling*) permettant une analyse multi-échelle particulièrement bien adaptée à l'analyse des images. Ainsi, un réseau de classification comme AlexNet prend en entrée une image (par exemple la photo d'un chat) et retourne en sortie sa probabilité d'appartenance à différentes classes (c'est un chat à 95 %, est un chien à 4 %, etc.). Les couches de convolution appliquent des filtres à l'image à différentes résolutions afin de faire ressortir des caractéristiques qui prédiront bien la classe. Ces filtres sont optimisés durant l'apprentissage lors duquel on fait défiler des images et leur classe correspondante en utilisant la technique de descente de gradient stochastique. Il s'agit d'apprentissage supervisé dans la mesure où les données d'entraînement sont « étiquetées » ou « annotées » : la classe de chaque image utilisée lors de l'apprentissage est connue.

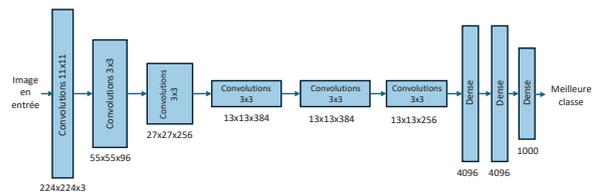


Figure 3 : Architecture AlexNet (Source : GE Healthcare).

Dans le domaine de la radiologie, l'architecture U-Net (Ronneberger, 2015), dérivée des réseaux de classification, a connu un grand succès pour réaliser des tâches dites de segmentation. La segmentation consiste à faire de la classification à l'échelle du pixel (ou voxel pour la 3D (Çiçek, 2016)) : l'architecture U-Net transforme une image en une autre de même taille dont chaque pixel correspond à une certaine classe. Prenons l'exemple de la segmentation du foie dans une image scanner : un réseau U-Net prend en entrée le scanner d'un patient (qui est une image 3D) et retourne en sortie une image de même taille dans laquelle les voxels correspondant au foie sont allumés (mis à 1), les autres étant éteints (mis à 0). Une application évidente de la segmentation

est le calcul du volume : on somme tous les pixels à 1 et on multiplie par le volume d'un voxel en mm^3 pour obtenir dans notre exemple le volume du foie en mm^3 .

L'architecture U-Net se décompose en deux parties : une partie « descendante » (encodage) et une partie « montante » (décodage) faites de couches de convolution intercalées de sur-échantillonnage. On utilise la même technique d'apprentissage supervisé pour entraîner ces réseaux. Dans l'exemple de la segmentation du foie en scanner, on donne comme exemples des centaines de scanners de patients dont la segmentation du foie est connue afin d'optimiser les poids du réseau U-Net. Après l'apprentissage, le réseau entraîné U-Net saura segmenter le foie dans une image scanner, connue ou inconnue.

Cette capacité à généraliser les exemples appris reste mal comprise aujourd'hui, si bien qu'on ne sait prédire de manière théorique la quantité d'exemples nécessaire pour une bonne capacité à généraliser. En pratique, on procède de manière empirique en augmentant progressivement la base d'apprentissage et en mesurant les performances sur une base de test indépendante. La taille de la base d'apprentissage est jugée suffisante lorsqu'on atteint les performances souhaitées sur la base de test. Cette base de test doit être représentative des conditions d'usage (techniques d'imagerie, âge, morphologie, ethnie des patients, variétés des pathologies, variantes anatomiques, etc.).

Les agences de régulation sont attentives à ce que les données de test soient bien représentatives et correctement annotées, afin de garantir une évaluation précise et non biaisée de l'algorithme. Si les images sources viennent souvent du soin courant, les annotations des

données fournies pour entraîner et évaluer le modèle demandent un travail coûteux impliquant des spécialistes rares. Un ou plusieurs radiologues experts doivent former un consensus, qui reste imparfait faute de confirmation de l'anatomie ou de la pathologie sur des patients *in vivo*. La référence peut être une structure anatomique complexe à tracer en 3D sur des dizaines d'images comme dans l'exemple de la segmentation du foie en scanner. En pratique, on utilise souvent des outils semi-automatiques pouvant reposer sur de l'IA pour créer ces annotations. Une autre stratégie utilise des « annotations faibles » : avec une très grande base de données, la simple mention d'une anomalie – sans sa localisation – peut suffire en laissant le modèle identifier la lésion par différence avec l'anatomie normale qu'il a apprise. Les données annotées peuvent être transformées ou déformées pendant l'apprentissage, ce qui favorise la généralisation des exemples. Ce processus est connu sous le nom d'augmentation de données.

L'émergence de modèles de fondation – *foundation models* comme SAM (Kirillov, 2023) – pourrait faciliter le développement de nouvelles fonctions en exploitant les connaissances acquises sur des problèmes similaires. Si ce type de méthode est adapté à l'analyse d'images photographiques pour lesquelles les corpus sont énormes, il présente des défis dans le domaine médical où les données sont plus rares. Les données pathologiques rares sont souvent sous-représentées dans les jeux de données d'entraînement des modèles de fondation, ce qui limite leur capacité de généralisation. Par conséquent, de nombreuses tâches essentielles mais spécialisées restent hors de portée de ces modèles, en raison de la disponibilité limitée de certaines données.

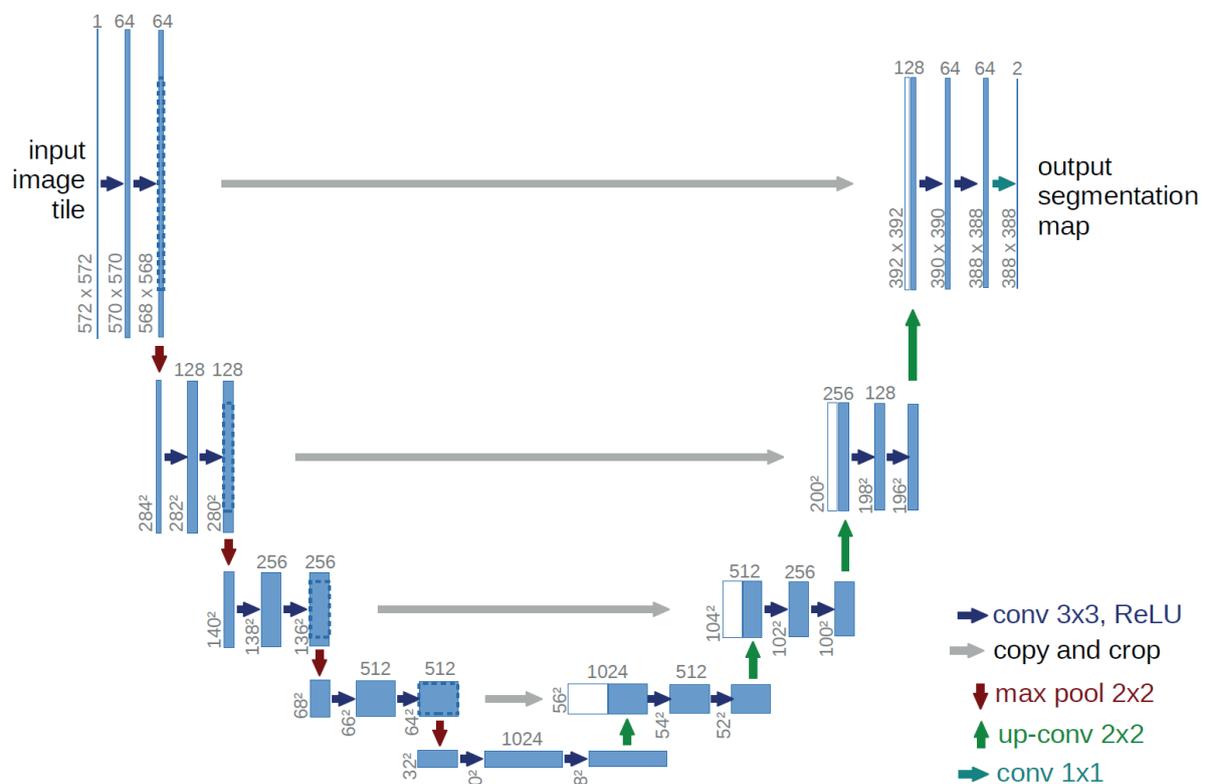


Figure 4 : L'architecture U-Net pour la segmentation d'image : les opérations de convolution sont indiquées par une flèche bleue ; les opérations de sous-échantillonnage correspondent aux flèches rouges ; celles de sur-échantillonnage aux flèches vertes (Source : U-Net: Convolutional Networks for Biomedical Image Segmentation).

Par ailleurs, si le développement du modèle est simplifié, il restera nécessaire de le tester sur des bases de cas suffisantes, ce qui exigera toujours un travail important de collecte de données annotées et de définition de la qualité.

Assister la quantification

Parmi les résultats attendus d'un examen radiologique, le travail de mesure peut représenter une charge significative, par exemple :

- la quantification de diamètres des vaisseaux (aorte pour des anévrismes, coronaires ou carotides pour diagnostiquer de possibles occlusions) ;
- les volumes ou diamètres de lésions cancéreuses : l'évolution du volume est un critère possible pour évaluer le succès d'un traitement ou la nécessité de le modifier.

Ces mesures demandent la détermination précise de la forme des organes ou des lésions. Elles sont complétées par des outils de visualisation des résultats de l'IA, éventuellement des corrections manuelles en cas d'échec. L'IA décharge ici les radiologues d'une part d'un travail répétitif et fastidieux.

Cette application qui relevait du traitement d'image traditionnel illustre les avantages de l'IA :

- prise en compte de morphologies très différentes difficiles à gérer par des règles ou des atlas ;
- besoins limités d'expertise en traitement d'image grâce à des technologies d'IA communes ;
- temps de développement réduit et plus facile à maîtriser (la part d'invention humaine est limitée).

Décision, pronostic

Parmi les possibilités nouvelles de l'IA, il est naturel d'envisager des modèles reliant les images et les dossiers de patients suivis dans la durée. On peut ainsi construire des prédicteurs associant des images précoces aux

développements futurs de pathologies probables, ce qui pourrait aider à choisir des traitements plus adaptés. Ce type de méthode rejoint la recherche autour de la « radiomique », qui recherchait des signatures similaires à l'analyse génomique dans des paramètres de texture des images de radiologie.

Parmi les domaines sur lesquels des développements actifs sont en cours, citons :

- l'évaluation du caractère malin des lésions (Massion, 2020) ;
- des pronostics sur la réponse à des traitements en oncologie (Xiang, 2025) ;
- en neurologie, la catégorisation des accidents vasculaires cérébraux (Koska, 2023).

Sur le même principe, l'âge des patients peut être évalué à partir de clichés radiologiques (Pickhardt, 2025).

Si ces méthodes sont prometteuses, leur performance n'est pas encore garantie. Si le résultat n'est pas accompagné d'éléments vérifiables par des humains, le risque d'erreur peut être inacceptable.

Communiquer

Des logiciels comme Rad AI ou PowerScribe de Microsoft associent l'IA générative et la reconnaissance vocale pour aider à rédiger les parties ouvertes des comptes rendus destinés aux médecins référents, qui représentent aujourd'hui une charge de travail significative.

Une possibilité intéressante est d'utiliser l'IA pour traduire le langage technique des comptes rendus en des termes plus compréhensibles par les patients, voire de répondre à leurs questions. Si les radiologues français ont encore le temps de parler aux patients, ce n'est pas le cas de leurs homologues américains. Des travaux ont évalué la pertinence des contenus fournis par des agents conversationnels, qui peuvent dialoguer directement avec des patients ou bien résumer

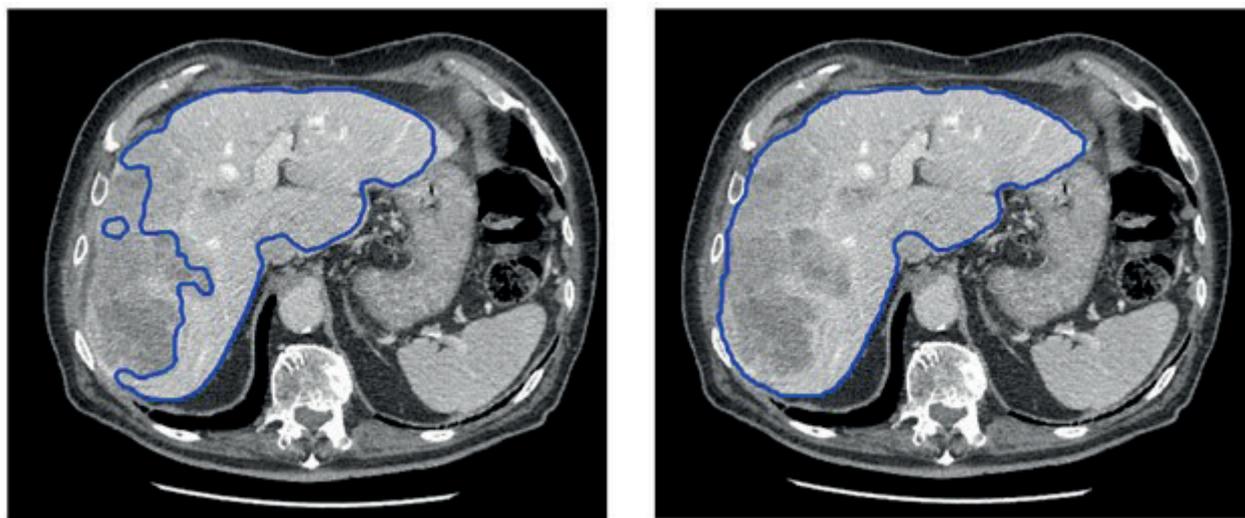


Figure 5 : Segmentation du foie en scanner : à gauche, un résultat obtenu par une méthode de traitement de l'image classique ; à droite, un résultat obtenu par un modèle d'intelligence artificielle entraîné sur des centaines d'exams – la présence de pathologies complique le travail de segmentation (Source : GE Healthcare).

les dossiers cliniques de manière structurée et systématique (Bhayana, 2025).

Le cadre réglementaire

Les autorités réglementaires se sont très tôt intéressées au domaine. Parmi les points encadrés :

- les données personnelles : consentement des patients à l'utilisation des données, protection de ces données, barrières entre régions (notamment depuis la Chine) ;
- les méthodes de développement et le respect de bonnes pratiques autour de l'IA ;
- le contrôle de la performance ;
- le remboursement de certains actes de radiologie impliquant l'IA.

Conclusion

L'intelligence artificielle est loin d'avoir remplacé les radiologues, mais elle a rapidement pris une place importante dans leur pratique, de l'acquisition des images à leur interprétation ou à la communication avec les référents. Avec des barrières au développement réduites et un financement très actif, l'innovation dans le domaine est foisonnante. La bonne intégration de ces outils, la garantie de leur performance, vont déterminer leur adoption, qui sera encore plus rapide dans les pays où l'accès aux techniciens spécialistes et aux radiologues est limité.

Avec une confiance accrue dans les modèles et une extension de l'accès aux dossiers médicaux numérisés, des applications plus ambitieuses utilisant les images comme un biomarqueur supplémentaire devraient étendre le pouvoir prédictif de l'imagerie médicale.

Bibliographie

BHAYANA R., ALWAHBI O., LADAK A. M., DENG Y., DIAS A. B., ELBANNA K., GOMEZ J. A., JAJODIA A., JHAVERI K., JOHNSON S., KAJAL D., WANG D., SOONG C. & KRISHNA S. (2025), "Leveraging Large Language Models to generate clinical histories for oncologic imaging requisitions", *Radiology*, 314(2), doi:10.1148/radiol.242134

CAO W., WU R., CAO G. & HE Z. (2020), "A comprehensive review of computer-aided diagnosis of pulmonary nodules based on computed tomography scans", *IEEE Access*, 8, pp. 78067-78080, doi:10.1109/ACCESS.2020.2990499

ÇIÇEK Ö., ABDULKADIR A., LIENKAMP S. S., BROX T. & RONNEBERGER O. (2016), "3D U-Net: Learning dense volumetric segmentation from sparse annotation", MICCAI.

GAO C., WU L., WU W., HUANG Y., WANG X., SUN Z., XU M. & GAO C. (2024), "Deep learning in pulmonary nodule detection and segmentation: A systematic review", *European Radiology*, 35, pp. 255-266, doi:10.1007/s00330-024-10907-0.

HERNSTRÖM V., JOSEFSSON V., SARTOR H. *et al.* (2025), "Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): A randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study", *The Lancet Digital Health*, doi: 10.1016/S2589-7500(24)00267-X.

HSIEH J., LIU E., NETT B., TANG J., THIBAUT JB. & SAHNEY S. (n.d.), "A new era of image reconstruction: TrueFidelity", Technical white paper on deep learning image reconstruction. Retrieved from <https://www.gehealthcare.com/-/jssmedia/files/truefidelity/truefidelity-white-paper-jb68676xx-doc2287426.pdf?rev=-1>

KIRILLOVA., MINTUN E., RAVI N., MAO H., ROLLAND C., GUSTAFSON L., XIAO T., WHITEHEAD S., BERG A. C., LO W.-Y., DOLLÁR P. & GIRSHICK R. (2023), "Segment anything", *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4015-4026.

KOSKAI. O. & SELVERA. (2023), "Artificial Intelligence in stroke imaging: A comprehensive review", *Eurasian Journal of Medicine*, 55 (Suppl. Issue: 1), S91-S97.

KRIZHEVSKY A., SUTSKEVER I. & HINTON G. E. (2012), "ImageNet Classification with deep convolutional neural networks", *Advances in Neural Information Processing Systems (NIPS)*, 25, pp. 1097-1105.

MASSION P. P., ANTIC S., ATHER S. *et al.* (2020), "Assessing the accuracy of a Deep Learning method to risk stratify indeterminate pulmonary nodules", *American Journal of Respiratory and Critical Care Medicine*, 202(2), pp. 241-249, doi:10.1164/rccm.201903-0505OC

PETERS R. D. & LAWSON S. (2020), "AIR™ Recon DL for diffusion-weighted imaging", GE Healthcare White Paper. Retrieved from https://www.gehealthcare.com/-/jssmedia/gehc/us/files/products/magnetic-resonance-imaging/air-technology/air-image-quality/mr_gbl_air-recon-dl_jb17939xx_v3.pdf?rev=-1

PICKHARDT P.J., KATTAN M.W., LEE M.H. *et al.* (2025), "Biological age model using explainable automated CT-based cardiometabolic biomarkers for phenotypic prediction of longevity", *Nat. Commun.*, 16, 1432.

RONNEBERGER O., FISCHER P. & BROX T. (2015), "U-Net: Convolutional networks for biomedical image segmentation", *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234-241.

XIANG J., WANG X., ZHANG X., XI Y., EWEJE F., CHEN Y., LI Y., BERGSTROM C., GOPAULCHAN M., KIM T., YU K.-H., WILLENS S., OLGUIN F. M., NIRSCHL J. J., NEAL J., DIEHN M., YANG S. & LI R. (2025), "A vision-language foundation model for precision oncology", *Nature*.

L'utilisation de l'IA pour mieux tirer parti d'une base documentaire technique au bénéfice des industriels de la mécanique

Par Fanny LAMBERT

Responsable de la Veille sur les IA génératives de Cetim

Pour améliorer l'exploitation des documents techniques au bénéfice des industriels de la mécanique, le Cetim a fait développer un *chatbot* interne capable d'exploiter de manière sécurisée et contrôlée des milliers de documents produits en interne. Cette initiative vise à surmonter les limites des *chatbots* accessibles sur le *web*, notamment en matière de confidentialité et de capacité de traitement.

Le projet, aujourd'hui abouti, utilise les documents de veille du Cetim pour générer des résumés, des synthèses et des comparatifs, facilitant ainsi la collecte et l'analyse rapide d'informations scientifiques et techniques. Ce projet a aussi permis de comprendre cette technologie, d'évaluer les coûts, les infrastructures IT nécessaires, et les compétences requises pour l'intégration d'une IA en interne, nous permettant de le partager avec vous aujourd'hui.

Contexte

Les *chatbots* accessibles sur le *web* peuvent permettre d'exploiter des documents, mais ils comportent leurs limites. En effet, les fichiers que nous leur fournissons peuvent être sauvegardés et exploités pour l'entraînement du prochain modèle du *chatbot* utilisé, en fonction de ses conditions d'utilisation, ce qui peut être problématique d'un point de vue de la propriété intellectuelle et de la confidentialité des données. Autre point limitant : il n'est généralement possible de fournir que quelques documents (5-10 fichiers de taille moyenne), ce qui ne correspond pas forcément au besoin quand nous avons une grande quantité de fichiers à exploiter. Pour répondre à ces problématiques, il peut donc être intéressant de déployer une IA en interne, en mesure d'exploiter plusieurs milliers de documents de manière sécurisée et contrôlée, avec la possibilité de paramétrer l'IA en fonction du type de document exploité.

C'est cette réflexion que nous avons eue au Cetim (Centre technique des industries mécaniques) et qui nous a poussés à nous lancer dans un projet de développement d'un *chatbot* interne, exploitant des documents produits par le Cetim. Pour réaliser une première preuve de concept (POC), nous avons choisi d'exploiter les documents de veille que nous produisons et mettons à disposition de nos industriels sur notre extranet, la Mécatèque¹. Ces quelques milliers de documents (plus de 6 500 à ce jour) contiennent des informations

techniques et scientifiques sur de nombreux procédés et technologies de l'industrie mécanique. Les objectifs de ce projet étaient multiples :

- Permettre à nos cotisants et aux cétimiens d'exploiter plus finement et plus rapidement cette masse d'information conséquente pour de la collecte d'informations, l'éclairage de prise de décision ou encore l'obtention d'une synthèse de l'information. L'IA serait ainsi une sorte de moteur de recherche amélioré, complémentaire au moteur de recherche classique, capable de fournir le résumé d'un document, des synthèses de plusieurs documents avec les liens vers les sources, ou encore des tableaux comparatifs ou SWOT (pour *strengths, weaknesses, opportunities* et *threats*).
- Vivre un projet d'intégration d'IA en interne, voir quelle est la démarche pour le mettre en œuvre, et comprendre le fonctionnement de cette technologie. Nous voulions connaître les possibilités d'un tel outil, les infrastructures IT et les compétences à prévoir, évaluer les coûts et délais d'un tel projet, mais aussi être conscients des limites d'un *chatbot* en interne.
- Pouvoir partager ces expériences avec nos cotisants à travers des notes de veille, webinaires et interventions.

Déroulement du projet

La première phase du projet a consisté à identifier et à sélectionner, parmi une douzaine de prestataires, quatre partenaires les plus adaptés aux besoins spécifiques du

¹ <https://www.cetim.fr/mecatèque/Toute-la-richesse-des-etudes>

projet. Cette sélection a donné lieu à un appel d'offres auprès de neuf sociétés, révélant une fourchette de budgets allant de 7 000 à 80 000 euros pour réaliser la phase 1. Les quatre prestataires retenus ont été mis en concurrence et se répartissent en deux catégories : deux intégrateurs de solutions « clé en main » (en l'occurrence Ekimetrics et The QA Company), et deux développeurs spécialisés en data science (Hurence et Cross Data). De nombreux tests rigoureux ont été effectués pour chacune des quatre solutions proposées. La seconde phase est quant à elle axée sur l'amélioration des performances du *chatbot* avec les deux prestataires que nous avons préférés (Hurence et QA Company), notamment en termes de pertinence des réponses, ainsi que sur l'optimisation de son ergonomie. Cette dernière phase visait également à préparer l'industrialisation du *chatbot* avec Hurence, que nous avons finalement retenu, afin de le déployer à grande échelle au sein de l'entreprise et de le mettre à disposition de nos cotisants.

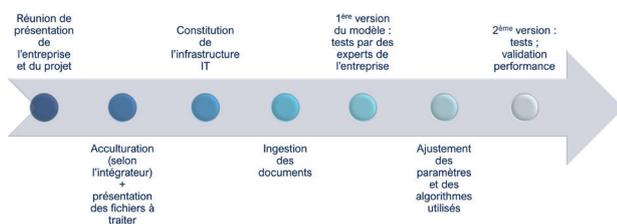


Figure 1 : Les étapes d'un projet de développement d'un *chatbot* interne (Source : Auteure).

Fonctionnement d'un modèle de langage

Les IA conversationnelles comme ChatGPT, Mistral ou Gemini par exemple sont ce qu'on appelle des grands modèles de langage, ou LLM. Ce sont des réseaux de neurones entraînés en auto-supervision sur une très grande quantité de textes. Grossièrement, cela signifie que le modèle essaye de compléter des séquences de mots tronquées issues de son corpus d'entraînement,

en calculant la probabilité que chaque mot de son vocabulaire soit celui attendu dans la phrase tronquée.

Par exemple, avec la phrase « le ciel est bleu », on va retirer le mot bleu et le modèle va essayer de retrouver la bonne proposition. Il essaye le mot qui lui semble le plus probable, il le compare ensuite avec le mot attendu, puis il met à jour ses probabilités en fonction du résultat. Il va répéter ainsi cette opération un très grand nombre de fois pour en déduire une régularité statistique, et c'est comme cela qu'il va développer ses « connaissances générales » (voir la Figure 2). Actuellement, un modèle de langage n'a donc pas de raisonnement à part entière, il fonctionne uniquement sur des statistiques pour construire ses réponses. La qualité du corpus fourni lors de l'entraînement va dès lors être déterminante dans la pertinence du modèle développé, puisque si vous lui fournissez des documents contenant des infos pour s'entraîner, cela perturbera ses statistiques et induira davantage de mauvaises réponses. De même, de nombreux biais risquent d'être transmis au modèle au cours de son entraînement. Pour vous donner un ordre d'idées, l'entraînement d'un modèle tel que GPT4 dure entre 28 et plus de 300 jours selon le nombre de machines requises. Cette procédure est donc très coûteuse financièrement et énergétiquement.

Une notion clé dans le fonctionnement des modèles de langage est la vectorisation, également appelée *embedding*. Pour que le modèle comprenne le sens des mots et des phrases, chaque mot se voit attribuer différentes valeurs qui le représentent : s'agit-il d'un verbe, quel est son genre, est-il pluriel ou singulier, est-ce un humain, un animal, une plante, etc. Toutes ces valeurs sont représentées sous forme de vecteurs, ce qui permet au modèle de déterminer si deux mots sont similaires (comme « chat » et « chaton ») ou totalement différents (comme « chien » et « maison ») (voir la Figure 3). C'est ainsi que les relations sémantiques sont établies, permettant au modèle d'exploiter plus efficacement les données.

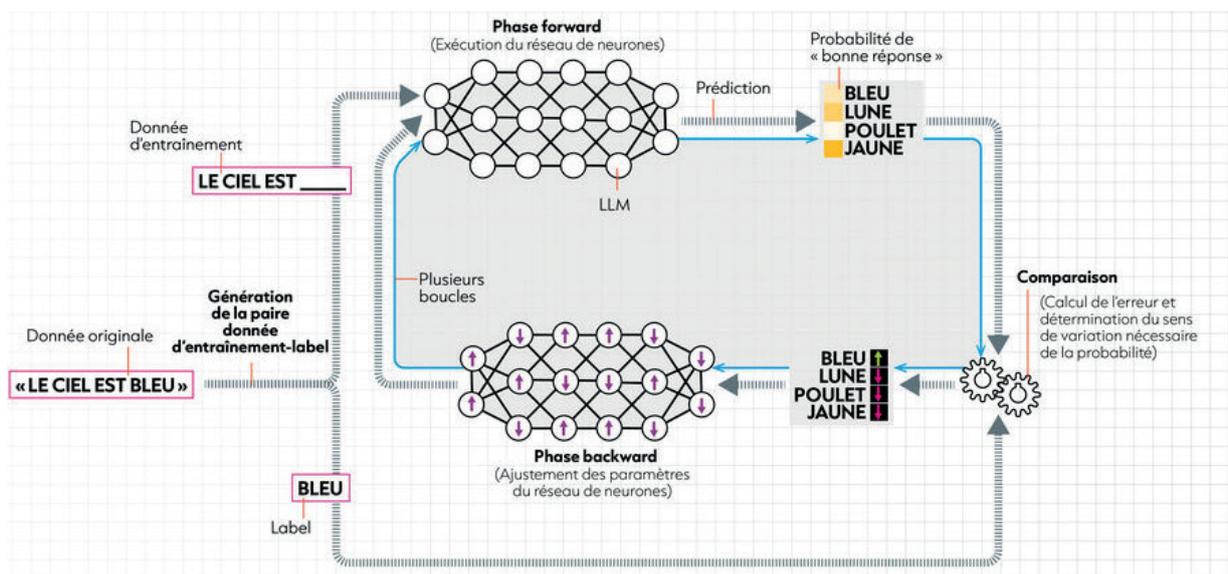


Figure 2 : Représentation schématique de l'entraînement d'un LLM par auto-supervision (Sources : Plongée dans les entrailles des grands modèles de langage qui font l'IA conversationnelle, *L'Usine Nouvelle* et Université de Genève).

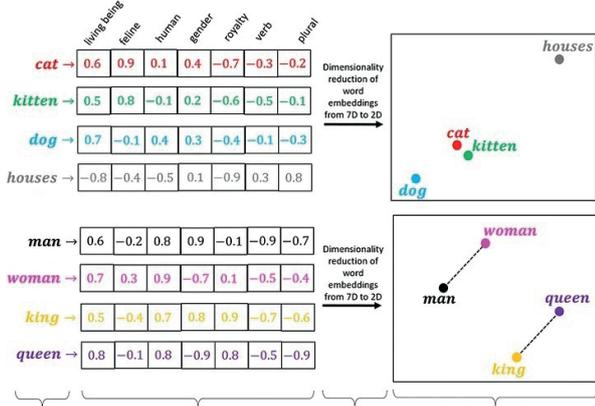


Figure 3 : Schématisation de la vectorisation des mots (Source : Medium).

Une fois ces bases posées, nous pouvons revenir à notre projet de *chatbot*. La technique retenue pour exploiter nos documents à l'aide d'une IA est appelée « RAG » (*Retrieval Augmented Generation*). En effet, cette méthode permet de connecter un LLM à une base de connaissances choisie, telle que des documents internes, une base de données ou autre, et confère ainsi des connaissances pointues au modèle sans qu'il soit nécessaire de le réentraîner. Cette méthode est assez économique et réduit les risques de réponses inventées ou fausses (que l'on appelle « hallucinations » dans le jargon). Un comparatif de cette technique par rapport à une IA utilisant uniquement ses connaissances liées à son entraînement est présenté dans le tableau ci-après.

La première étape du RAG consiste à découper vos documents en fragments (paragraphes, phrases...) appelés *chunks* par les développeurs. Ces *chunks* sont ensuite vectorisés et stockés. Lorsqu'une requête (*prompt*) est effectuée auprès de votre *chatbot*, celle-ci est également vectorisée pour être comprise par le LLM. Ensuite, les fragments de documents les plus pertinents sont recherchés dans la base de données. Enfin, ces *chunks* pré-sélectionnés sont analysés par le LLM pour générer une réponse (voir la Figure 4).

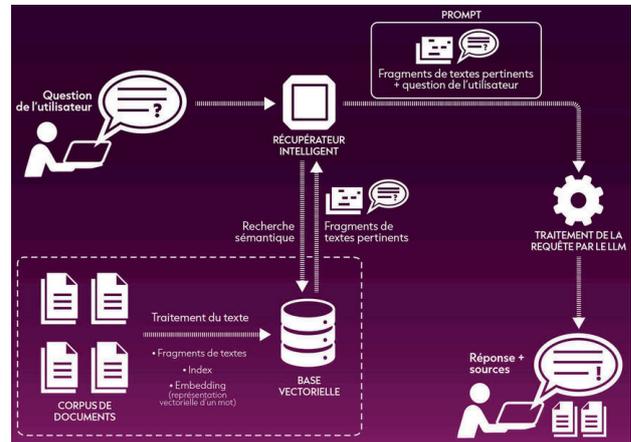


Figure 4 : Représentation simplifiée du fonctionnement du RAG (Source : L'intelligence artificielle générative se diffuse dans l'entreprise grâce au RAG).

Bien entendu, pour faire tourner ce type de technologie, vous aurez besoin d'infrastructures solides

Critère	LLM utilisant ses connaissances générales	LLM connecté à un RAG
Source des infos	Basée sur des données pré-entraînées	Accès en temps réel à des bases de données / API
Actualité des données	Fixe à la date d'entraînement	Infos à jour en temps réel
Questions spécifiques	Réponses parfois approximatives, générales	Recherche de documents précis
Contexte	Basé sur le modèle seul	Appui sur des documents externes
Données évolutives	Réentraînement nécessaire	Infos récentes sans réentraînement
Fiabilité des faits	Bonne pour les sujets courants	Plus fiable grâce aux sources récupérées
Ressources	Pas de mémoire externe requise	Nécessite des infrastructures supplémentaires
Vitesse de réponse	Très rapide	Plus lente mais acceptable (recherche externe)
Explications / Citations	Impossible de connaître la source utilisée	Peut fournir des sources et citations
Utilisation typique	Assistants virtuels, réponses générales	FAQ dynamiques, expertises techniques nécessitant des informations actualisées

incluant des GPU (cartes graphiques) professionnels. Pour cela, deux options s'offrent à vous :

- Soit vous achetez votre matériel (*hardware*) pour l'installer dans vos locaux. Avec les prix actuels, vous en aurez pour environ 50 000 € pour 2 GPU. Vous pouvez également louer le matériel que vous installez chez vous pour 1 000-2 000 €/mois. Les besoins GPU varient en fonction du nombre d'utilisateurs, mais aussi en fonction du modèle IA choisi. En effet, un modèle comme GPT4 va consommer beaucoup plus de ressources qu'un modèle de chez Mistral par exemple.
- Soit vous installez votre IA sur un *cloud*, c'est-à-dire que vous allez réserver une petite partie d'un *data center* pour vos usages (abonnement classique ou à l'usage). Il vous faudra choisir entre un *cloud* souverain (français ou européen) comme OVH, Scaleway ou vast.ai par exemple, et un *cloud* américain comme Azure (Microsoft), AWS ou Google Cloud Platform, soumis quant à eux à l'US Patriot Act².

Résultats

Le développement du *chatbot* destiné à exploiter les ressources de la MécaThèque a permis de dégager plusieurs enseignements, tant sur la qualité des réponses fournies que sur les ressources nécessaires à un tel projet.

Les performances de celui-ci ont été évaluées par les dix membres de l'équipe de veille technologique, auteurs des documents utilisés. Cette évaluation portait sur un échantillon varié de questions techniques posées au *chatbot*, couvrant les thématiques majeures de la MécaThèque. Une cinquantaine de tests ont été réalisés pour chaque version proposée par les prestataires.

Les résultats montrent qu'à la fin des tests, 55 % des réponses étaient jugées bonnes, c'est-à-dire complètes et techniquement exactes, 33 % des réponses étaient moyennes, souvent correctes mais incomplètes, nécessitant des recherches supplémentaires par l'utilisateur, et 12 % des réponses étaient mauvaises, comprenant des erreurs ou des hors-sujets.

Les erreurs étaient principalement dues à une mauvaise interprétation des termes techniques ou à une confusion dans le traitement de documents complexes (mélange d'informations par exemple), mais ne correspondaient pas à des hallucinations.

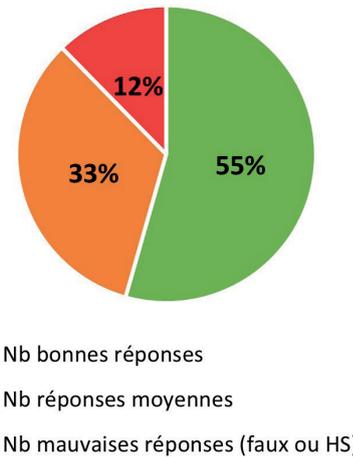


Figure 5 : Répartition de la qualité des réponses dans la dernière version du *chatbot*, à ce jour (Source : Auteurs).

Pour atteindre ces résultats, le projet a mobilisé :

- une équipe de testeurs avec une bonne connaissance des documents, permettant d'évaluer plus aisément la pertinence des réponses ;
- une implication importante des équipes SI, pour le développement et l'intégration du *chatbot* ;
- plusieurs mois de travail, pour l'entraînement de l'IA générative, la structuration des données et l'amélioration continue des réponses ;
- un budget estimé entre 40 et 70 k€ pour financer un seul prestataire jusqu'à la fin de la phase 2, sans compter le temps de travail des équipes internes et les coûts d'infrastructure.

Cette expérience nous a permis d'acquérir une meilleure compréhension du fonctionnement des IA génératives, notamment concernant l'entraînement, la vectorisation et le RAG détaillés précédemment. Nous avons également réalisé à travers nos tests l'importance de la qualité et la bonne structuration (titres, tableaux, légendes, images) des documents fournis dans le RAG pour faciliter leur exploitation par l'IA. Il faut d'ailleurs prendre en compte que les modèles de base ne sont pas capables de lire et analyser les images. Ce type de fonctionnalité peut être ajouté, mais représente un coût supplémentaire.

Cependant, le langage technique spécifique à l'industrie mécanique reste un défi, nécessitant une amélioration continue de l'IA pour mieux gérer le jargon et les subtilités des documents de veille.

² <https://shs.cairn.info/dictionnaire-du-enseignement--9782262070564-page-592?lang=fr>

Le potentiel de l'intelligence artificielle pour une entreprise technologique industrielle

Par François MARION

Directeur de la communication et membre du comité exécutif du groupe Valeo

L'utilisation de l'intelligence artificielle (IA) est un puissant moteur de transformation d'une entreprise technologique industrielle telle que Valeo. L'introduction il y a treize ans de l'IA dans les systèmes embarqués d'aide à la conduite a permis au groupe d'acquérir des compétences de premier plan en matière d'IA. Valeo déploie maintenant cette technologie au sein même de l'entreprise pour transformer les fonctions ; que ce soit en R&D, par l'optimisation du développement logiciel *via* les assistances au codage ou la démultiplication des tests et validations, ou en production avec par exemple les tests automatiques ou l'optimisation de la consommation énergétique. Plusieurs exemples montrent qu'il est possible d'introduire l'IA en appliquant une approche pragmatique de prototypage rapide, de développement et de déploiement par étape. Mais, compte tenu de l'impact attendu, son déploiement systématique nécessitera d'accompagner cette mutation.

En novembre 2023, à Paris, 120 collaborateurs de Valeo, venus du monde entier et organisés en 16 équipes, participaient à un hackathon autour de l'IA générative. Accompagnées pendant trois jours par des experts en IA de Valeo, d'Artefact et de Google, ces équipes ont pu chacune construire un prototype de système autour de leurs idées, lesquelles avaient été sélectionnées parmi plus de 600 proposées au total au sein de l'entreprise. Les trois équipes gagnantes, dont les projets portaient sur la maintenance industrielle prédictive, l'analyse de données et l'analyse de spécifications techniques en R&D, ont eu la garantie de voir leur projet développé et mené au bout. Cela a également été le cas d'autres projets, et notamment celui de l'équipe issue de la direction de la communication, visant à indexer automatiquement la bibliothèque de fichiers médias du groupe.

Le succès d'une telle démarche est, en soi, porteur de plusieurs enseignements. Tout d'abord, l'intelligence artificielle est un sujet capable de mobiliser très largement au sein de l'entreprise, impliquant des collaborateurs de métiers divers. On y voit également que l'IA, et notamment l'IA générative, se prête bien à une approche très pragmatique, fondée sur le prototypage rapide, une introduction précoce et un déploiement progressif. Les partenariats, comme ici entre trois sociétés de premier plan en matière d'IA, sont une quasi-nécessité dans ce domaine, et nous en présenterons d'autres exemples clés. Enfin et surtout, cette expérience illustre la pertinence de l'IA dans de nombreux domaines de l'entreprise, qui va ainsi être profondément transformée.

C'est essentiellement ce dernier point qui sera développé dans cet article : les perspectives de l'IA pour une entreprise technologique industrielle comme Valeo.

Les origines de l'IA chez Valeo : l'IA embarquée dans les systèmes d'aide à la conduite

Chez Valeo, l'IA est arrivée il y a treize ans avec l'introduction pour la première fois de réseaux neuronaux dans une caméra de recul intelligente. Grâce à l'IA, la caméra pouvait analyser avec précision des scènes complexes et classer ce qui était détecté : d'autres voitures, des deux-roues, des piétons, des infrastructures...

L'intérêt de l'IA dans ce cas est de passer d'une situation antérieure, où il était nécessaire d'encoder toutes les règles et toutes les situations, à un système capable d'apprendre des règles à partir du contexte, simplement en lui donnant des milliers d'exemples, en utilisant l'apprentissage profond.

Ces algorithmes ont rapidement été déployés dans d'autres types de systèmes, pour permettre aux voitures d'effectuer en autonomie des fonctions de plus en plus complexes : le freinage d'urgence automatique, le stationnement automatique, le régulateur de vitesse adaptatif, le maintien de voie, la navigation assistée ("*Navigation on Pilot*") sur autoroute, en ville... C'est également ainsi que les algorithmes de perception du capteur le plus avancé utilisé pour la conduite

autonome, le Lidar, voit, détecte, classe et interprète ce qu'un radar ou un œil ne verrait pas.

Au-delà, le développement du véhicule autonome nécessite d'apprendre à la voiture à détecter et prédire les intentions des autres utilisateurs de la route, et notamment les plus vulnérables d'entre eux que sont les piétons et les cyclistes. Connaître leur niveau d'attention ou de distraction dû, par exemple, à l'utilisation en cours du téléphone portable, est également très utile. Enfin, les véhicules devront également interpréter correctement les gestes des forces de l'ordre ou des personnels sur un chantier en bord de route. Valeo a ainsi présenté au CES (Consumer Electronics Show, un salon annuel sur la technologie, souvent présenté comme le plus important au monde), entre 2020 et 2025, plusieurs générations d'algorithmes basés sur l'IA et capables d'accomplir ces tâches, très naturelles pour des conducteurs humains, mais pas du tout pour des machines.

Compte tenu de l'importance critique de l'IA pour le développement de systèmes d'assistance à la conduite, ou de véhicules autonomes, Valeo a ouvert à Paris, en 2016, Valeo.ai, premier centre de recherche mondial en IA dédié aux applications automobiles et en interface avec la communauté scientifique mondiale. Aujourd'hui, le groupe est riche de quelque 200 experts en IA, et Valeo.ai est le fer de lance de notre recherche dans le domaine. C'est en capitalisant sur cette compétence précoce en IA que nous avons pu initier un déploiement ambitieux de l'IA dans de nombreux domaines de l'entreprise, en commençant par la R&D.

La transformation de la R&D par l'intelligence artificielle

Assez naturellement, l'IA peut être utilisée pour améliorer les logiciels d'aide à la conduite eux-mêmes équipés d'IA. La validation des logiciels d'aide à la

conduite passe en effet par la comparaison entre la perception d'une scène réelle par le logiciel et celle qu'en ferait un cerveau humain. Il faut donc disposer d'une bibliothèque de scènes annotées, ce qui nécessitait initialement à la fois des prises de vue en conditions réelles et une annotation à la main. L'IA permet non seulement d'accélérer et de faire à moindre coût l'annotation automatique d'images, mais elle permet également de multiplier le nombre de scénarios complexes étudiés : il devient possible, à partir d'une scène, d'en créer des centaines en incorporant, grâce à l'IA générative, toute sorte de variations, comme l'ajout de piétons, de véhicules, de cyclistes, ou la modification des conditions climatiques. Alliée à la puissance du *cloud*, l'IA permet ainsi d'accélérer le développement des unités de contrôle électronique jusqu'à 40 % et le développement de logiciels embarqués intégralement sur le *cloud* – c'est l'objet d'une collaboration entre Amazon Web Services et Valeo, annoncée au CES de cette année.

Au-delà des tests et de la validation des logiciels, l'IA générative peut également s'avérer un réel avantage pour la génération de code logiciel. C'est ainsi que Valeo et Google Cloud ont étendu le partenariat mentionné en introduction pour développer des outils et solutions d'IA générative basés sur le *cloud* et visant à améliorer l'efficacité et la productivité des processus de Valeo grâce à des cas d'utilisation spécifiques. Grâce à un accès anticipé aux dernières technologies d'IA générative de Google Cloud, Valeo a pu équiper, depuis mars 2024, 5 000 de ses développeurs *software* d'une boîte à outils pour la création, la correction et l'optimisation de codes. Nous constatons, partout dans le monde, une adoption très rapide de ces outils par les développeurs, qui y voient un intérêt direct pour leur performance au quotidien – on est bien ici dans le cas représentatif de « l'ingénieur augmenté par l'IA ».

Un autre exemple d'« ingénieur augmenté » en cours de déploiement est celui de la conception assistée par

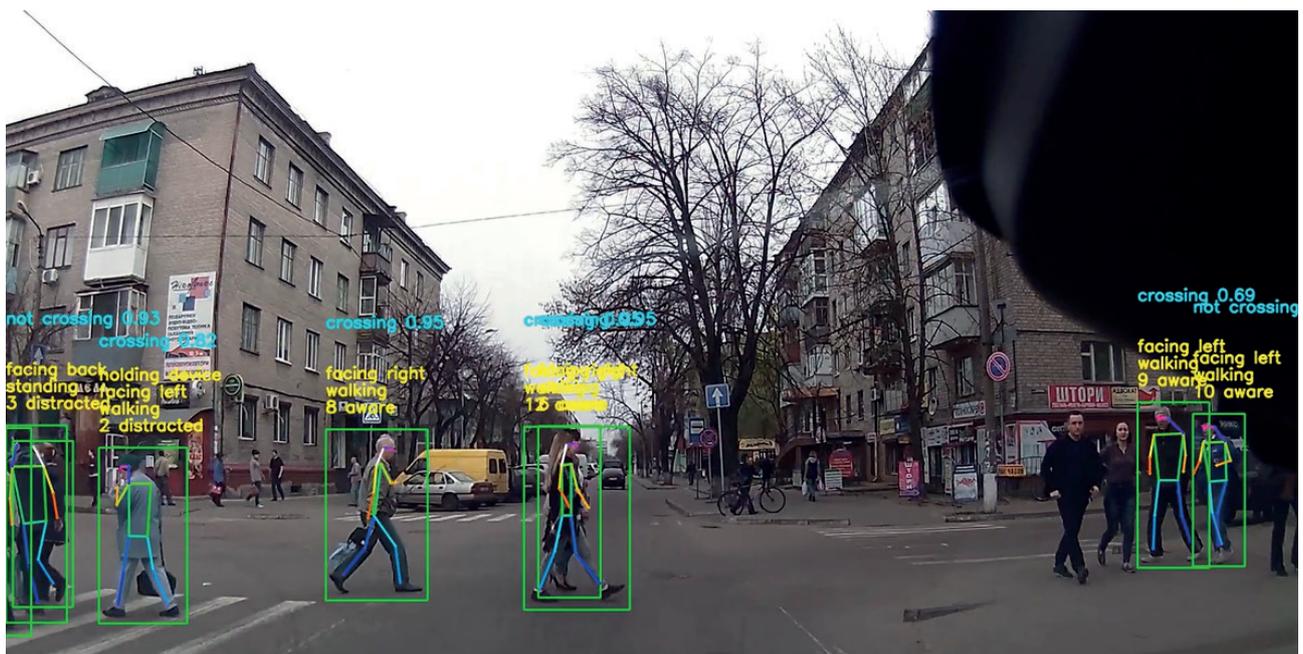


Figure 1 : Classification des piétons par l'IA et détection de leurs intentions et de leur niveau d'attention (Source : Valeo).

ordinateur, où l'IA peut permettre d'assister les ingénieurs dans la création de pièces répondant à un jeu de spécifications techniques ou de conditions aux limites.

L'intelligence artificielle est également à même de transformer d'autres pans de la R&D : aider à structurer les projets, à comprendre les spécifications techniques requises, à s'assurer de l'implémentation de règles de design... autant d'applications dont le potentiel est en cours d'évaluation.

L'IA et l'optimisation des processus industriels

La première application industrielle de l'IA chez Valeo est venue naturellement de l'expertise en IA dans les caméras : l'utilisation de la reconnaissance d'image avancée pour automatiser les contrôles qualité visuels, comme l'inspection en fin de ligne des glaces plastiques sur les phares. Dans ce cas, l'apprentissage profond a permis d'optimiser le processus, jusqu'à obtenir 10 fois moins d'erreurs de classification qu'avec l'inspection humaine. L'IA permet en outre de diminuer l'occurrence de défauts, en identifiant et optimisant la valeur des paramètres critiques des processus, dans le but de réduire le taux de rejet : une expérience menée dans ce domaine avec Telecom Paris a ainsi permis d'améliorer de 40 % le taux de rejet d'une ligne de production de machines électriques.

Un autre exemple d'usage de l'IA dans le domaine industriel concerne l'optimisation de la consommation d'énergie. Bien sûr, l'IA peut être utilisée à des fins d'optimisation de la consommation au niveau central, mais cela nécessite que l'ensemble de l'usine soit équipée de capteurs, ce qui est un coût important pour des usines existantes. Une alternative consiste à mesurer la consommation énergétique au niveau d'une ligne et de chercher ainsi à l'optimiser. Pour ce faire, un dispositif mobile, une

« valise », a été développé dans l'usine d'Abbeville : il est constitué de capteurs d'énergie connectés, sans-fil, d'un collecteur de données et d'un ordinateur portable faisant tourner une IA spécifiquement entraînée par Valeo. Les premiers tests effectués ont permis des gains énergétiques jusqu'à 20 %, et le dispositif est en cours de déploiement sur les autres sites du groupe. Cette valise, brevetée, va bientôt faire l'objet d'une commercialisation, et son inventeur, un jeune ingénieur du groupe, a reçu un prix de la part du magazine *L'Usine Nouvelle* en 2023. Ce dernier exemple illustre également à quel point l'IA se prête bien à des approches très pragmatiques d'amélioration continue.

Conclusion

Les premiers déploiements de l'IA, d'abord dans le domaine des logiciels embarqués, puis au sein de la R&D et des processus industriels, démontrent déjà l'incroyable potentiel de la technologie pour une entreprise telle que Valeo. D'autres domaines d'application semblent également prometteurs en finance, en gestion des ressources humaines, en communication ou pour optimiser divers processus administratifs.

La plupart des cas d'usage de l'IA présentés ici consistent à équiper les employés de l'entreprise d'outils leur permettant de s'acquitter de leurs fonctions mieux et plus efficacement, en faisant ainsi des « salariés augmentés ». Cela nécessitera bien sûr des formations à l'utilisation des outils, ainsi que de la pédagogie autour des avantages qu'apportent l'IA, voire à terme des changements au sein des organisations. Dès lors, le potentiel de transformation des entreprises industrielles par l'intelligence artificielle semble particulièrement important. Toutefois, réussir cette mutation présentera des enjeux majeurs en termes de conduite du changement, sur les prochaines années et sur tout le tissu industriel.



Figure 2 : Kit portable à base d'IA pour optimiser la consommation d'énergie dans les usines (Source : Valeo).

Construction : quand l'IA change la donne

Par Marie-Luce GODINOT

Directrice générale adjointe du groupe Bouygues en charge de l'innovation, du développement durable et des systèmes d'information

L'introduction de l'intelligence artificielle (IA) dans le secteur de la construction a ouvert de nombreuses opportunités. De la phase d'appel d'offres à celle de la maintenance et de l'exploitation, l'IA permet d'améliorer significativement l'efficacité et la sécurité des projets. L'intégration de technologies telles que la *computer vision* et le *machine learning* transforme les processus traditionnels, rendant possibles de nouvelles applications et innovations. Et l'arrivée de l'IA générative, bien que récente, montre déjà des perspectives prometteuses.

L'adoption de l'intelligence artificielle dans le secteur de la construction est en constante progression, et bien que les technologies soient déjà matures pour plusieurs applications, le potentiel de développement reste immense. Les avancées rapides dans ce domaine laissent présager des chantiers plus intelligents, sûrs et performants à l'avenir.

Bien avant l'émerveillement suscité par l'arrivée de ChatGPT et des IA génératives, l'intelligence artificielle avait déjà largement commencé à intégrer le monde de la construction. La complexité des projets, le volume de données s'y référant, la réalité physique des chantiers, les enjeux de sécurité ou de durabilité avaient ouvert la voie à la mise en œuvre d'algorithmes autour du *machine learning* ou de la *computer vision*. Mais aujourd'hui, avec l'essor d'outils d'IA générative, une nouvelle étape s'ouvre : plus puissante, plus créative, plus accessible, et pourtant toujours semée de défis. Jusqu'où l'IA est-elle aujourd'hui présente dans le quotidien des activités, et quelles conclusions peut-on en tirer à ce stade ?

Des attentes importantes liées aux enjeux du secteur

De façon générale, l'intelligence artificielle dans la construction a généré de formidables attentes sur toute la chaîne de valeur.

Elles sont liées aux enjeux fort spécifiques au domaine de la construction, comme :

- un secteur confronté à une pénurie de nouveaux talents, à un vieillissement de la main-d'œuvre qualifiée et qui peine à attirer, d'où des enjeux d'automatisation et de productivité ;
- des activités devant veiller aux impératifs d'optimisation, de constructibilité, de décarbonation, de consommation de ressources naturelles ou de préservation de la biodiversité, favorisant la construction à plus faible empreinte carbone et l'économie circulaire ;

- un ensemble coût / qualité / délai / sécurité qui pourrait tirer profit d'une transformation digitale incarnée par l'industrialisation, dont la standardisation autour de composants et de sous-systèmes optimisés, la préfabrication et des nouveaux modes constructifs, mais qui peine à s'opérer.

La phase d'appel d'offres

Sur la phase d'appel d'offres, c'est bien l'arrivée de l'IA générative qui joue le rôle d'accélérateur.

Que ce soit Bouygues Construction, Colas ou Equans, le constat est le même : l'IA apporte un gain de temps et d'efficacité dans les phases amont. Comment ? En analysant, pour les marchés les plus volumineux, les milliers de pièces fournies pour en détecter les incohérences ; en améliorant la sélectivité dans les activités de fonds de commerces ; enfin, en aidant à la rédaction des propositions finales.

Dans tous les cas, ces outils sont en accompagnement des équipes pour améliorer l'efficacité, respecter les plannings et réduire les erreurs. Basées sur des développements internes et / ou des solutions externes comme Altenders, ces technologies sont en cours de déploiement.

La phase de conception

Sur la partie conception, le potentiel semble immense, et l'arrivée récente de l'IA générative n'a fait qu'amplifier les attentes.

Il peut s'agir d'offrir une assistance très amont pour identifier plus rapidement des options conceptuelles, de réaliser de l'optimisation multicritères, de l'automatisation ou

de l'analyse / simulation. Les exemples chez Bouygues sont nombreux.

Generative Design

Colas exploite le *Generative Design* pour optimiser ses projets d'infrastructure et générer de manière itérative l'ensemble des solutions en respectant des règles prédéfinies. L'expert peut ainsi sélectionner la solution optimale en tenant compte de plusieurs contraintes métiers telles que le coût, le délai et l'empreinte carbone.

Un premier outil, OYA (du nom de la déesse des tempêtes), optimise le dimensionnement des bassins d'infiltration, quelle que soit leur nature (bassin ouvert, fermé, à noue, etc.).



Figure 1 : Ecran du logiciel OYA (Source : Colas).

Sur la base des données géométriques du projet (emprise, altimétrie du terrain naturel, pente en fond de bassin, nombre de canalisations entrantes, etc.) ; des caractéristiques de l'exutoire et du raccordement (pente, altitude, profondeur du fil d'eau, etc.) et des hypothèses de terrassement ; des paramètres tels que le coefficient de sécurité, l'imperméabilité du sol et la profondeur des nappes phréatiques ; des règles de calcul ; et d'une base de prix : l'outil génère des calculs itératifs pour proposer plusieurs solutions optimales en termes de dimensionnement et de coût. C'est pour les bureaux d'études un gain de temps d'étude, des économies sur le dimensionnement optimisé et une qualité de livrables améliorée. Le gain moyen observé est de 15 à 30 %.

Le second outil optimise le dimensionnement des dalots (petit canal recouvert d'une dalle, situé sous les remblais des routes ou des voies ferrées) en phase de conception. En intégrant les règles de calcul structural et géotechnique, une base de prix main-d'œuvre et matériaux, les caractéristiques du projet, l'outil génère les solutions optimales technico-économiques. Résultats très encourageants : jusqu'à 30 % d'économies et 10 % de gain de temps d'étude, avec une meilleure coordination entre métiers !

Rendus photoréalistes

Chez Bouygues Construction, la génération de rendus photoréalistes grâce à l'IA a permis aux équipes commerciales et de conception de gagner du temps et de réduire les coûts. Images de haute qualité en quelques secondes, coût raisonnable, créativité stimulée, exploration de nouveaux designs, itérations facilitées sur les projets : que d'avantages ! Toutefois, il faut vérifier la faisabilité avec un architecte et conserver une

traçabilité des données transmises et générées. Bien que ces outils ne remplacent pas totalement un infographe, surtout pour les rendus finaux, l'IA accélère la conception et inspire les équipes, nécessitant néanmoins formation et validation technique.

Optimisation de parking : un enjeu majeur en conception

Maximiser le nombre de places de parking tout en respectant les contraintes réglementaires et spatiales est un des éléments sur lequel se fait l'économie du projet.

Les solutions d'optimisation multidisciplinaire permettent des gains de matière considérables. C'est ainsi que Bouygues Construction a pu par exemple concevoir un parking de même capacité mais avec un étage en moins, ou en économisant plus de 140 tonnes d'armature sur une paroi de soutènement.

Des tests avec plusieurs outils (Parkyze, Spacio, Testfit...) ont permis de générer rapidement des ébauches de parkings, mais des progrès restent à faire en termes de précision et de conformité aux standards.

De vrais progrès dans la réalisation des métrés

Depuis des années, des solutions sont explorées, en vain, pour automatiser, à partir des maquettes BIM ou des plans 2D, la production des métrés, essentiels pour estimer les quantités de matériaux et les coûts d'un projet. Récemment, des outils comme Tocal.AI, Kreo et DatabuildR ont montré de meilleurs résultats sur la détection automatique des éléments et l'accélération du *process*. Toutefois, des erreurs de classification et un manque de fiabilité sur certains éléments nécessitent encore une validation humaine.

Bénéfices réels sur l'étude de faisabilité

Certains outils comme Autodesk Forma ou Buildrz permettent d'ores et déjà d'accélérer la conception préliminaire. Mais c'est un segment de marché en plein foisonnement, et l'évolution de l'écosystème doit être surveillée pour tirer profit des opportunités qui émergent.

Conclusion sur la phase de conception

En conclusion, même si, pour l'heure, le potentiel promis par l'IA en phase de conception ne se réalise pas pleinement, l'évolution rapide de la technologie ouvre des perspectives enthousiasmantes pour le futur de la conception augmentée par l'IA.

Compte-tenu de l'effervescence de l'écosystème dans ce domaine, il est indispensable de maintenir une veille constante pour suivre les progrès des solutions actuelles et découvrir de nouveaux acteurs entrant sur le marché avec des innovations prometteuses.

La phase de chantier

En phase chantier, la nature des enjeux à traiter conduit plutôt à utiliser des solutions intégrant du *machine learning* ou de la *computer vision*, plus matures.

Méthode brevetée de classification des particules de chantiers

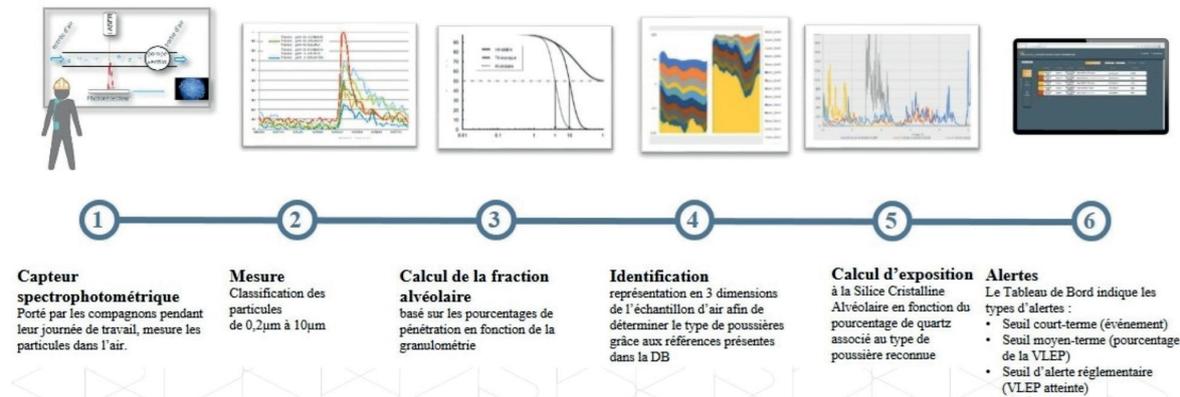


Figure 2 : Méthode brevetée de classification des particules de chantier (Source : Uby.)

Sécurité sur les chantiers, un enjeu crucial

La sécurité est la priorité sur les chantiers, où les risques sont omniprésents : interactions entre engins et piétons, chutes, détection de personnes non autorisées, port des EPI¹...

Les technologies de *computer vision*, proposées par des fournisseurs comme Vi-Act, Artefeel, KartaX ou Antron, offrent des solutions relativement matures pour détecter et alerter en temps réel. Par exemple, chez Bouygues Construction, des caméras ont été déployées sur des premiers chantiers pour surveiller les zones de levage et émettre des alarmes immédiates en cas de danger.

Ces technologies améliorent la vigilance et la sécurité, malgré quelques défis comme l'adaptation aux changements sur le site et les problèmes de luminosité, qui pourraient certainement bénéficier de l'usage de données synthétiques.

L'intégration de ces technologies dans l'entreprise marque une étape vers une maturité technologique accrue et une approche proactive en matière de sécurité.

UBY², une plateforme innovante pour le monitoring des nuisances des chantiers

Depuis plusieurs années, le groupe Bouygues a développé des solutions intégrant de l'IA pour répondre aux enjeux environnementaux et sociétaux inhérents à ses activités de chantier. La société UBY a été créée à cet effet et a mis au point des algorithmes performants sur le sujet.

En particulier, l'acceptabilité des chantiers par les riverains passe par une maîtrise des nuisances sonores. L'objectif de la solution mise en œuvre est de reconnaître le type de bruit et sa localisation pour adapter les actions à réaliser. L'algorithme actuel de reconnaissance sonore, intégré directement sur le capteur de la société Wavely,

transforme le signal sonore en image spectrale, puis reconnaît le type de bruit grâce à l'algorithme de classification basé sur un réseau de neurones convolutionnel³ et une base de données propre de plus de 10 000 signatures sonores de bruits de chantier. Cette solution va évoluer pour utiliser le modèle Yanmet de Google, pré-entraîné sur des jeux de données audio en enrichissant le réseau de neurones avec à la fois le jeu de données spécifiques et un jeu de données publiques adapté, comme ESC50.

La société UBY a également développé uDust, en partenariat avec Ellona, une méthode brevetée de classification des particules, comme le montre le schéma ci-dessus (voir la Figure 2).

Cette méthode permet d'identifier les particules présentes et de réaliser un calcul d'exposition à la silice cristalline alvéolaire. Des alertes sont alors émises en fonction de seuils paramétrés.

Un troisième cas d'usage porte sur la reconnaissance des salissures de chaussée grâce à de la vision par ordinateur. Des alertes sont générées lors de détection d'un pourcentage de salissure trop élevé. Un algorithme de *deep learning* est utilisé qui prend en compte la luminosité, l'heure, les conditions météo ou les textures de chaussée.

Ces solutions sont aujourd'hui largement utilisées, en particulier sur des chantiers d'infrastructures.

La technologie au service de la gestion des déchets de chantier

La gestion des déchets sur les chantiers, notamment leur tri, se complexifie avec le durcissement des réglementations. En France, l'objectif est de réduire la quantité de déchets enfouis au profit de leur réutilisation, recyclage ou transformation en énergie.

Pour y répondre, la société Akanthas propose une solution utilisant l'intelligence artificielle et la vision

¹ Équipements de protection individuelle.

² www.uby-group.com

³ La convolution est une opération mathématique dans les réseaux de neurones convolutionnels (CNN) qui a révolutionné le traitement d'images, améliorant beaucoup la performance des tâches de reconnaissance visuelle.

par ordinateur pour surveiller en permanence les bennes de chantier grâce à un mat télescopique et un capteur à alimentation solaire. En identifiant les déchets déposés, cette solution évalue en temps réel la composition et le poids des bennes, permettant d'en optimiser les rotations, réduisant les coûts et facilitant la gestion sur le chantier.

De plus, la solution alerte automatiquement les prestataires chargés de l'enlèvement des bennes et fournit des preuves en cas de déclassement arbitraire par les centres de tri ou les organismes de recyclage. À noter que l'apprentissage des règles de tri appropriées reste en tout état de cause essentiel pour les acteurs des chantiers.

Avec ce type de solution, l'industrie de la construction fait un pas de plus vers une gestion des déchets plus efficace et respectueuse de l'environnement.

Conclusion sur la phase chantier

En conclusion, l'utilisation de l'IA, particulièrement le *machine learning* et la *computer vision*, révolutionne la phase chantier en apportant des solutions de plus en plus matures et efficaces. Ces technologies permettent de répondre à des enjeux cruciaux tels que la sécurité, la protection de l'environnement et le suivi de l'avancement des travaux. L'IA et ses applications innovantes, comme la reconnaissance sonore ou la classification des déchets, contribuent non seulement à rendre les chantiers plus sûrs et plus respectueux de l'environnement, mais aussi à optimiser les processus et à réduire les coûts. Grâce à ces avancées, l'industrie de la construction peut désormais envisager un avenir plus durable et technologique, où chaque aspect du chantier est minutieusement contrôlé et optimisé.

Les phases d'exploitation et de maintenance des ouvrages

En phase d'exploitation et de maintenance des ouvrages, l'intelligence artificielle peut aussi se révéler très utile pour analyser les travaux à réaliser ou optimiser le fonctionnement des ouvrages.

Les cas d'usage en la matière sont très nombreux, que ce soit sur l'optimisation énergétique des bâtiments, dont Equans est un spécialiste, ou la surveillance d'infrastructures.

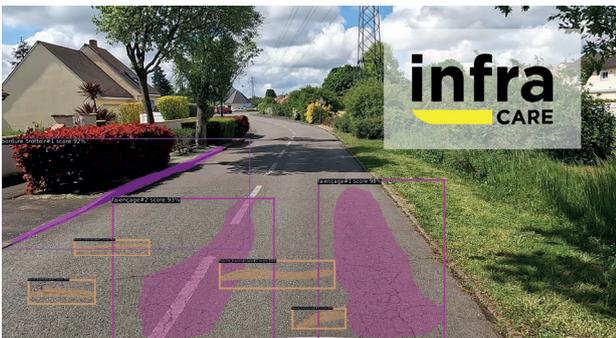


Figure 3 : détections des dégradations par le logiciel InfraCare (Source : Colas).

Colas a par exemple développé Infracare, un service intégrant l'usage de l'IA, pour aider les collectivités à mettre en place une programmation de travaux optimisée.

À partir d'un relevé de caméra GoPro, un algorithme d'IA détecte les dégradations visibles sur la chaussée et en déduit des indicateurs d'états, qui sont restitués sur le site Internet d'Infracare. La plateforme permet ensuite d'aider à l'élaboration de plusieurs scénarios d'entretien, et de suivre la performance des travaux réalisés.

Le choix a été fait par le groupe de développer son propre algorithme d'IA et de faire certifier tout le processus de développement par le LNE⁴, faisant de Colas la première entreprise à être labellisée « IA de Confiance » dans le secteur industriel. Cette certification, réalisée par un organisme d'État, permet d'attester de la qualité des processus de l'IA, de démontrer les niveaux de performance de l'algorithme et d'être en conformité au regard du règlement européen sur l'intelligence artificielle du 13 mars 2024⁵ ("AI Act").

Infracare a également été sélectionné comme un des 111 lauréats sur plus de 550 candidats dans la catégorie "AI for Efficiency", l'appel à projets lancé par le ministère français chargé de l'Économie dans le cadre du sommet de février 2025 sur l'intelligence artificielle.

Voilà un bel exemple d'usage de l'IA qui aide les collectivités à favoriser l'approche préventive de l'entretien, permettant d'allonger la durée de vie des infrastructures, d'améliorer la sécurité des usagers et de réduire leur empreinte carbone.

Conclusion

En somme, l'IA constitue déjà une pierre angulaire pour l'innovation dans le secteur de la construction. Les exemples pertinents sont déjà nombreux. Pour autant, de nombreuses opportunités restent à exploiter, dès que les solutions deviendront plus matures. Car il est clair que les potentialités de l'IA ne cessent de croître.

À l'avenir, nous pouvons imaginer des chantiers encore plus intelligents, où l'IA collaborera avec des technologies telles que l'Internet des objets (IoT), les jumeaux numériques⁶, et la robotique, permettant d'envisager des chantiers plus sûrs, plus rapides et plus performants.

L'engagement continu dans la R&D et les collaborations entre experts techniques et spécialistes de l'IA garantiront que ces visions deviennent réalité. C'est ainsi que l'industrie de la construction pourra non seulement répondre aux défis actuels, mais également anticiper et façonner un futur plus durable et résilient.

⁴Laboratoire National de Métrologie et d'Essais.

⁵ https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_FR.pdf

⁶Voir par exemple : <https://annales-des-mines.org/wp-content/uploads/2024/11/9-1.-Motivations-et-genese-des-jumeaux-numeriques-Albert-BENVENISTE-Yves-CASEAU-Nicolas-DEMASSIEUX-Patrick-JOHNSON-Catherine-LAMBERT-Jean-Luc-MOLINER-Sophie-PROUST-Gerard-ROUCAIROL.pdf>, p. 3.

GEN AI : le nouveau choc d'autonomie

Par Pierre DULON

Directeur général adjoint de CACIB, en charge de l'Informatique et des Opérations bancaires

Cet article défend une double conviction :

- La technologie GEN AI a un potentiel de transformation majeure dans les métiers de la banque, car elle va « décentraliser » les capacités de calcul et de raisonnement et fortement « autonomiser » les collaborateurs ;
- Le projet de déploiement GEN doit être d'abord abordé comme un projet humain et managérial, car c'est la mise en responsabilité des équipes qui sera le facteur clé de réussite.

Le vrai impact de l'intelligence artificielle dans le secteur bancaire est encore à venir

L'activité de la banque est faite de relations clients, de traitement de données financières et de calcul. Sans surprise, ce secteur a donc toujours été fortement consommateur de technologie, et l'intelligence artificielle y est déjà utilisée au quotidien. Dans le monde de la banque de financement et d'investissement, les réseaux de neurones sont déjà utilisés pour des calculs de prix d'instruments ou d'analyse de risques. L'IA prédictive aide déjà à anticiper le comportement des clients ou les mouvements de marchés, et les outils de détection de la fraude ou de lutte anti-blanchiment utilisent largement ces technologies. Pourtant, au-delà des premiers cas d'usage ponctuels d'utilisation de l'IA générative comme « assistant », la véritable transformation est encore devant nous.

La vraie transformation vient de l'autonomisation du plus grand nombre

Si on regarde en arrière, la dernière véritable transformation du secteur des banques d'investissement liée à la technologie remonte à quarante ans avec le déploiement des tableurs et des PC. Soudainement, chaque collaborateur a disposé sur son propre poste de travail des outils permettant de manipuler des données, de les trier et de mettre en place des modèles et des simulateurs. Cet *empowerment* massif de tous les collaborateurs et de leurs interlocuteurs clients a permis la très forte croissance des activités de fusions-acquisitions ou encore des marchés de capitaux pendant deux décennies.

Depuis, l'essor du digital et des technologies collaboratives a, certes, permis la « mise en réseau » et l'accès à l'information massive. Mais la capacité humaine de tout un chacun pour analyser, trier et synthétiser toute cette information a limité l'impact réel de ces technologies, et il est parfois difficile d'en mesurer l'apport réel de valeur.

Le banquier d'affaires reste aujourd'hui fortement dépendant du système d'information centralisé de son établissement, et sa réactivité reste souvent liée à celle de son département informatique. Contrairement au client particulier, le client entreprise de la banque n'a pas non plus vécu de bouleversement dans son quotidien, car, pour lui aussi, la quantité d'informations à traiter est bien trop lourde.

C'est ce verrou que l'essor de l'IA générative devrait faire sauter. Cette technologie va permettre à chaque individu, à chaque collaborateur, à chaque client, de disposer en autonomie d'une capacité de « digestion » de toutes les données disponibles, de génération des synthèses, de réalisation des outils de simulation ou de traitement nécessaires.

En outre, GEN AI, *via* ses capacités de « traduction », va définitivement recentrer l'interaction homme-machine du côté de l'humain. Les interactions en « langage naturel » vont se généraliser et rendre les échanges plus simples et intuitifs. Nous pourrions enfin bientôt définitivement oublier les « serveurs vocaux interactifs » qui nous obligeaient à communiquer en tapant des numéros sur notre téléphone...

Cette combinaison d'accès massif à l'information, de collaboration à distance et de capacité d'intelligence locale est inédite dans l'histoire et devrait redéfinir les standards de productivité et d'innovation.

La transformation de l'IA GEN sera d'abord humaine et managériale

Naturellement le développement GEN AI pose de nombreux problèmes techniques, environnementaux, voire éthiques, qui devront être traités.

Mais, au sein de nos entreprises, le projet de déploiement d'outils si puissants dans les mains de tous les collaborateurs est avant tout un projet riche de défis humains et managériaux.

Le premier défi le plus évident est la formation massive et l'accompagnement au changement. Comment accompagner tous les collaborateurs dans ces changements majeurs d'organisation du travail ou même parfois de métier, rassurer devant les inquiétudes et aider à acquérir les compétences nécessaires ?

Mais cette « décentralisation » massive pose également des questions sur le contrôle et la maîtrise globale du dispositif, en particulier dans un secteur si fortement régulé que la banque. Comment garantir demain la traçabilité des processus, la qualité des données échangées, la conformité aux réglementations diverses ? Il est déjà difficile dans nos entreprises de garantir la qualité, et le cycle de vie des données structurées c'est-à-dire, en théorie, « bien rangées » dans les systèmes informatiques. Alors comment obtenir la même maîtrise pour des données non structurées des multiples fichiers bureautiques, des conversations audio ou vidéo ? Des réponses à ces questions dépendront également le rôle et l'organisation des équipes informatiques dans ce nouvel environnement.

Enfin, le défi managérial : mettre à disposition d'un collaborateur un outil technologique puissant ne suffit pas à lui donner plus d'autonomie réelle. Encore faut-il mettre en place une organisation et un mode de management qui incitent et développent cette « décentralisation » du pouvoir et de la prise de décision. Le professeur Sumantra Ghosal expliquait il y a quelques années dans son discours à l'INSEAD "The Smell of the Place" toute l'importance du « climat » de travail issu de la culture d'une entreprise et de son mode de management.

C'est ce défi que le groupe Crédit Agricole a attaqué depuis cinq ans avec le lancement de son « Projet Humain ».

Les entreprises qui auront réussi à créer un climat de confiance et de mise en responsabilité de leurs collaborateurs seront certainement les mieux préparées à la transformation GEN AI.

L'intelligence artificielle en pratique pour faire face aux défis d'aujourd'hui et de demain

Par Emmanuel COX

Digital transformation officer chez SNCF Réseau

SNCF Réseau, en charge du réseau ferroviaire français, doit faire face à une demande croissante de transport en train, tout en modernisant ses infrastructures et en préservant les équilibres financiers. 50 000 collaborateurs gèrent au quotidien ce vaste réseau de 28 000 km de lignes, nécessitant de nombreuses compétences. Une transformation numérique profonde est en cours pour gagner en efficacité et en qualité de service.

L'entreprise intègre l'intelligence artificielle (IA) depuis de nombreuses années pour optimiser ses opérations : maintenance prédictive, analyse de données, gestion du trafic. Des projets utilisent l'IA pour anticiper les aléas climatiques et optimiser la gestion des infrastructures. L'IA sert aussi à améliorer la ponctualité et la réactivité.

Si l'IA générative est adoptée prudemment pour des usages industriels, elle offre des opportunités comme l'amélioration de la recherche documentaire, la génération de contenus et l'efficacité numérique. Des questions demeurent sur l'impact sur l'emploi, la souveraineté et l'environnement, faisant l'objet d'expérimentations.

Un cadre éthique encadre l'utilisation de l'IA, dont l'adoption passe aussi par l'accompagnement des salariés, leur formation et sensibilisation. Le renforcement des partenariats académiques et industriels est envisagé pour poursuivre le développement responsable de l'IA au sein de SNCF Réseau.

SNCF Réseau : au rendez-vous de nombreux défis industriels

Des challenges importants pour le gestionnaire du réseau ferré français

SNCF Réseau est la filiale du groupe SNCF en charge de l'entretien du réseau et de l'exploitation ferroviaire. Véritable colonne vertébrale du transport par le train en France, l'entreprise s'attache à garantir un traitement équitable de tous les opérateurs (historiques et nouveaux entrants) sur le territoire que ce soit pour le transport des voyageurs et du fret.

Plus de trains à faire circuler

SNCF Réseau fait face à des défis majeurs dans la gestion et le développement du réseau ferroviaire français, qui s'étend sur plus de 28 000 km de lignes. L'augmentation de la capacité du réseau constitue un enjeu crucial, alors que la demande de transport ferroviaire s'intensifie, avec plus de 5 millions de voyageurs quotidiens et 300 millions de tonnes de marchandises transportées annuellement. Cette problématique est particulièrement aiguë dans les nœuds ferroviaires des grandes métropoles et sur les axes les plus fréquentés, où transitent parfois plusieurs centaines de trains par jour.

Plus de travaux à réaliser

Le programme de modernisation représente un investissement d'environ 3 milliards d'euros annuels, avec la nécessité de rénover 1 000 km de voies par an, de remplacer 30 % des systèmes de signalisation d'ici 2030, et d'électrifier 400 km de lignes supplémentaires. Ces travaux doivent être menés dans un contexte de contraintes budgétaires fortes et avec l'impératif de minimiser les perturbations pour les usagers. La synchronisation de ces 1 500 chantiers annuels, leur financement et leur réalisation dans des fenêtres de travaux limitées à quelques heures par nuit constituent un exercice d'équilibriste permanent pour les 50 000 collaborateurs de l'entreprise.

Plus d'aléas à maîtriser

SNCF Réseau met en œuvre plusieurs mesures concrètes pour faire face aux aléas climatiques qui s'intensifient. L'entreprise a développé des systèmes de gestion en temps réel, intégrant des alertes spécifiques en partenariat avec Météo France afin d'anticiper les événements météorologiques extrêmes, tels que les vents violents et les précipitations intenses. En outre, des mesures préventives, telles que le ralentissement des trains pendant les périodes de fortes chaleurs, sont appliquées pour éviter les déformations des voies. SNCF Réseau collabore également avec les autorités

locales pour traiter des aspects techniques, notamment la gestion des eaux et l'abattage d'arbres présentant un risque pour les infrastructures ferroviaires. Ces actions sont conçues pour anticiper et atténuer les risques climatiques, tout en assurant la sécurité et la continuité du fonctionnement du réseau ferroviaire.

Ces impératifs opérationnels interviennent également dans un contexte d'engagements RSE ambitieux au service de la sécurité de tous, des citoyens, de la préservation de la biodiversité et des ressources en privilégiant l'économie circulaire et le recyclage de nos déchets.

Des collaborateurs engagés aux expertises multiples

Les 50 000 collaborateurs de SNCF Réseau interviennent en permanence sur le réseau ferré pour assurer la qualité de service, la sécurité des millions de voyageurs qui empruntent les 15 000 trains quotidiens sur le réseau.

65 métiers différents sont représentés dans l'entreprise avec des niveaux d'exigences élevés en termes de réactivité opérationnelle (7j/7 et 24h/24 pour de nombreuses activités), en termes de sécurité et d'expertises très diverses pour manipuler et assurer la maintenance de milliers de composants différents du réseau (systèmes électriques complexes, de signalisation, postes d'aiguillage qui se digitalisent, systèmes de régulation du trafic...).

À ces besoins pour nos équipes d'être formées en fonction des évolutions et de la modernisation des installations que nous gérons, s'ajoute la vague de la transformation digitale qui traverse l'ensemble de nos activités. Les smartphones sont devenus les outils de base du mainteneur sur nos voies avant la tirefonneuse ou le poste à souder. C'est ainsi un ensemble d'applications mobiles qui permettent aux équipes sur le terrain de disposer de leur programme journalier d'intervention, de pouvoir se géolocaliser en n'importe quel point du réseau, d'accéder aux consignes de sécurité ou aux instructions de maintenance...

De grandes mutations technologiques pour développer la capacité du réseau

Plusieurs grands programmes associant technologies de pointe et transformation des opérations sont en cours au sein de SNCF. Des initiatives comme NExTEO en Île-de-France, Haute Performance Marseille-Vintimille (HPMV) en PACA ou LGV+ entre Paris et Lyon visent à densifier la circulation des trains en toute sécurité. Elles mobilisent des ruptures technologiques importantes autour du Système européen de gestion du trafic ferroviaire (ERTMS en anglais), la conduite assistée, les jumeaux numériques. Ces projets structurants pour l'avenir du transport ferroviaire embarquent déjà ou embarqueront de l'intelligence artificielle, pour gérer les automatismes, faciliter la gestion des intervalles entre les trains, améliorer la ponctualité.



Figure 1 : Un réseau ferroviaire en constante mutation (Crédits : Christophe Recoura).

L'intelligence artificielle au service de la performance

Des data disponibles

SNCF Réseau s'est engagée il y a plus de dix ans dans l'application de l'intelligence artificielle pour répondre à ces enjeux. Les champs d'exploration s'avèrent en effet nombreux : maintenance prédictive, gestion de la connaissance, traitement d'images ou du langage naturel, optimisation, aide à la décision ; tous les métiers de l'entreprise sont concernés par le recours aux algorithmes et au *machine learning*.

Cette évolution a notamment été rendue possible par un volontarisme important en matière de structuration des données. Comme toute organisation historique, SNCF Réseau dispose d'un patrimoine *data* très volumineux. Depuis 2015, des efforts importants ont été mis en œuvre pour classifier les données en fonction de leurs usages (description physique du réseau, données sur son utilisation, sur le cycle de vie des *assets* ferroviaires...). La volumétrie et les traitements sont conséquents, et toute la production ferroviaire est adossée à la capacité de gérer des données soit en temps réel (suivi des trains, gestion des itinéraires) soit en temps différé (cycle de fabrication des horaires, organisation des travaux...). L'architecture de ces données est au fur et à mesure construite autour d'entrepôts de données spécialisés par grandes catégories de données et accessibles aux différentes applications consommatrices.

Des experts mobilisés

Cette organisation progressive de la *data* permet à nos équipes de *data scientists* de réaliser les traitements et de développer les algorithmes. Ces spécialistes travaillent notamment sur l'anticipation des phénomènes climatiques et les impacts sur notre réseau (risques d'inondation par exemple), sur la croissance de la végétation autour de nos emprises au moment où la réglementation concernant l'usage du glyphosate s'est durcie, mais aussi à l'anticipation des incidents ou des risques de retard.

De nombreux cas d'application de l'intelligence artificielle

Décider en temps réel

Les programmes « Train Autonome » ainsi que les nouveaux systèmes d'exploitation du trafic ferroviaire (NExTEO/ATS+) reposent sur des technologies qui embarquent de l'IA (reconnaissance d'images, algorithmes pour la gestion des intervalles). La finalité de ces programmes est d'augmenter le nombre de trains circulant par heure pour plus de confort pour les voyageurs et d'être encore plus réactifs en cas d'incident par une aide à la décision.

Intervenir avant la panne

SNCF Réseau déploie plusieurs solutions d'IA dans le cadre de la télésurveillance pour :

- éviter des incidents sur les moteurs qui actionnent les aiguillages ;
- éviter des défaillances de certains systèmes électriques de détection des circulations des trains (10 000 zones suivies) ;
- mieux cibler les visites d'intempéries en analysant les données météo.

Ces dispositifs permettent progressivement de prévenir des incidents et d'éviter concrètement des dysfonctionnements et des retards de trains préjudiciables pour nos voyageurs : prévenir 1 incident permet d'éviter plusieurs centaines de minutes de retard cumulées.

Analyser les images et les vidéos du réseau

C'est un des domaines où les expérimentations sont nombreuses et les expertises notamment au sein de filiales technologiques de SNCF Réseau existent :

- analyse des images de caméras de vidéosurveillance pour la sécurité et la gestion des flux en gares ;
- analyse des vidéos enregistrées par nos trains de mesure, qui embarquent des caméras pour analyser d'éventuels risques sur les voies ferrées ou sur les abords ;
- analyse de l'occupation des voies et de la végétation.

D'autres initiatives sont en cours pour pouvoir superviser les passages à niveau.

Traitement du langage : moteur de traduction, analyse sémantique pour les traitements des incidents, agents conversationnels

Des initiatives sont en cours en matière de traduction automatique pour les échanges vocaux aux frontières entre agents dans les postes d'aiguillage et les conducteurs de trains de pays voisins. C'est l'objet du projet ATHENA, qui s'intègre dans un projet plus large de digitalisation des échanges à bord des trains et les équipes de supervision du trafic.

L'analyse de données textuelles issues des bases de retours d'expérience fait l'objet de travaux exploratoires pour identifier des événements précurseurs d'incidents.

La direction des ressources humaines a introduit de façon très précoce un *chatbot* destiné à informer les salariés des dispositifs et services RH à leur disposition. Il a rencontré un grand succès, avec plusieurs milliers de sollicitations par mois.

La nouvelle donne de l'IA générative

La « révolution » de 2022

L'accessibilité des solutions d'IA générative, le succès très rapide d'Open.AI et le foisonnement des modèles de langage engendrent depuis le premier semestre 2023 un relais médiatique massif et une explosion des usages. De nombreux domaines ont été impactés (relation client grand public, efficacité individuelle des collaborateurs, développement IT, publicité, graphisme, culture...). Dans le secteur industriel, auquel s'apparentent les activités de SNCF Réseau, le développement de l'IA générative est plus prudent. Le secteur qui engage souvent la sécurité des personnes – nous concernant, celle des voyageurs, de nos personnels, mais aussi des marchandises que nous transportons – est moins tolérant aux hallucinations et aux inexactitudes.

Cela nous conduit à avancer à pas mesurés autour d'expérimentations concrètes pour évaluer les bénéfices de l'IA générative dans notre contexte. C'est le sens de la démarche que nous avons entreprise autour de la création d'un Lab IA et de la trentaine de projets que nous pilotons.

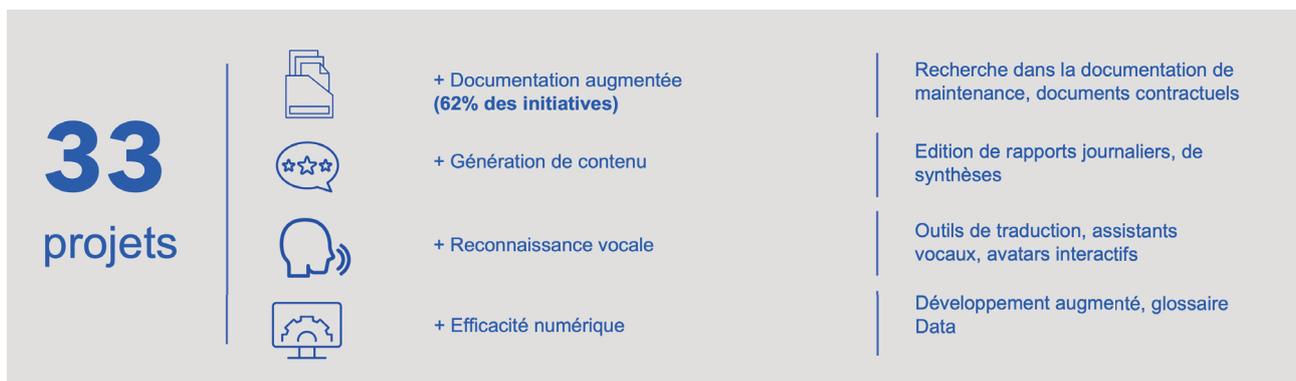


Figure 2 : Portefeuille des projets d'IA générative de SNCF Réseau en janvier 2025 (Source : DG Numérique SNCF Réseau).

Des opportunités à explorer

Comme l'ensemble du groupe SNCF, SNCF Réseau s'est progressivement emparée de ces technologies et a constitué un portefeuille de projets pour le compte de l'ensemble des métiers de l'entreprise.

Aujourd'hui, la plupart des projets (les deux tiers) s'orientent autour de solutions de recherche documentaire accélérée dans des corpus denses, parfois diffus, tels que des réglementations, des procédures ou des standards à appliquer. Les autres cas d'applications concernent la génération de synthèses ou l'exploitation de données afin d'améliorer le traitement d'incidents ou d'anticiper la maintenance d'équipement, sur la base de rapports d'inspection et de comptes-rendus d'intervention.

Des résultats qui méritent d'être consolidés

Il est encore trop tôt pour tirer des gains d'efficacité généralisables issus de nos premières expériences. Toutefois, les premières observations permettent de mettre en évidence des gains de temps représentant 6 à 7 % du temps de travail hebdomadaire. C'est pour l'instant à rapporter à des populations très limitées (quelques salariés qui testent) et en général éclairées sur l'utilisation de l'IA générative, en tout cas qui l'accueillent positivement.

Quelles ambitions et quel cadre pour la suite ?

Un cadre éthique et responsable pour le développement de l'IA

Le recours à ces technologies ne manque pas de nous questionner sur plusieurs plans.

Quels sont les impacts durables sur les métiers et l'emploi ?

L'automatisation généralisée peut aboutir à des pertes de savoir-faire, notamment dans des métiers où les expertises se transmettent entre pairs et par compagnonnage et *mentoring* au sein des équipes. Il conviendra de bien analyser sur quels périmètres l'IA doit être déployée et pour quels usages.

Comment garantir souveraineté des solutions, transparence des algorithmes et protection des données ?

C'est une question centrale en matière de recours à l'IA générative. Les solutions sont largement entre les mains d'acteurs technologiques américains et chinois, avec des engagements très relatifs en matière de respect de la propriété des données et du patrimoine intellectuel. Ces considérations nous amènent à nous doter de solutions en propre avec des espaces de données privatisés comme pour la plateforme SNCF GPT, voire des solutions hébergées localement et développées à partir de modèles *open source*.

Quels impacts environnementaux et économiques ?

Là encore, les interrogations sont persistantes quant aux conséquences de l'accroissement de la puissance

de calcul pour faire face aux multiples applications de l'IA générative. Aujourd'hui, sur les 50 000 salariés de SNCF Réseau, seuls quelques milliers ont accès aux services d'IA générative : que représenteront la consommation en énergie et le coût d'utilisation des systèmes numériques nécessaires pour les opérer (systèmes *cloud*, tarification des fournisseurs de modèles de langage...) ? Difficile encore de se projeter. C'est le sens des expérimentations que nous menons : mettre en place les systèmes d'évaluation qui permettent de mesurer les effets d'un passage à l'échelle à la fois en termes de gains et de coûts et impacts.

Le groupe SNCF s'est par ailleurs doté d'un guide éthique pour les usages de l'intelligence artificielle, permettant de fixer des principes de bonne conduite.

Un embarquement et une sensibilisation des équipes

L'ancrage de l'IA générative passera nécessairement par l'adhésion des salariés. Un véritable gain d'efficacité et de performance ne pourra intervenir que si les salariés adhèrent, voient un intérêt durable, comprennent les usages et les bénéfices de recourir à ces solutions. Sans cet accompagnement, la diffusion de l'IA générative à grande échelle ne pourra se faire.

De nombreuses actions de sensibilisation sont en cours. SNCF Réseau donne accès à la plateforme IA interne à 5 000 collaborateurs progressivement, des sessions d'information et de formation sont organisées également sur tout le territoire.

Des partenaires et de la veille

Enfin, SNCF Réseau a la conviction que le développement de l'IA se fera grâce à des échanges accrus avec les écosystèmes existants en France et en Europe : *clusters* IA, monde académique, autres groupes industriels, *start-ups* sont autant de partenaires possibles pour accélérer en commun le déploiement d'une IA plus fiable, plus responsable et en pleine complémentarité avec les expertises et talents des femmes et des hommes de SNCF Réseau.



Figure 3 : L'IA à l'aide des professionnels du ferroviaire (Crédits : Pascal Guittet).

IA de confiance, condition nécessaire pour le déploiement de l'IA dans les systèmes critiques

Par Juliette MATTIOLI et Christophe MEYER

Thales

Thales développe des solutions pour des secteurs critiques comme la défense, l'aéronautique et la cybersécurité, en mettant l'accent sur l'IA de confiance. Pour garantir la validité, la transparence, la sécurité, la fiabilité et la responsabilité des systèmes d'IA, Thales combine des approches symboliques et connexionnistes (IA hybride). L'entreprise s'assure que ses systèmes sont robustes face aux cyberattaques et transparents pour les utilisateurs (voire explicables chaque fois que cela est possible).

Thales participe à des initiatives comme ANITI et Confiance.ai pour développer des méthodologies d'ingénierie de l'IA de confiance. L'accent est également mis sur la durabilité et l'éthique, en optimisant les ressources et en réduisant l'empreinte carbone grâce à des technologies avancées dites frugales.

Introduction

Thales développe des solutions pour la défense et la sécurité, l'aéronautique et le spatial, la cybersécurité et le digital, contribuant à un monde plus sûr, plus durable et plus inclusif. Ces activités se situent dans des domaines critiques où les enjeux peuvent être vitaux, où il faut respecter des principes éthiques, de transparence, de sûreté et de sécurité. Pour s'assurer que ces solutions à base d'intelligence artificielle (IA) soient conçues de manière à limiter les risques et garantir des pratiques responsables dans leur développement et leur déploiement, il est nécessaire de démontrer les propriétés de :

- validité, pour qu'un système à base d'IA fait ce qu'il doit faire, tout ce qu'il doit faire et seulement ce qu'il doit faire ;
- transparence et d'explicabilité, pour fournir aux utilisateurs des justifications adaptées au contexte et à leur niveau de connaissance ;
- sécurité, pour assurer la robustesse et la résilience aux conditions adverses, telles que les cyber-attaques ;
- fiabilité et de sûreté, pour contrôler et réduire les risques induits par des défaillances éventuelles du système ou par des insuffisances fonctionnelles ;
- responsabilité, pour respecter les cadres éthiques, juridiques et réglementaires (y compris les règles de confidentialité des données et de protection des informations sensibles) et être conforme aux standards en vigueur.

Enjeux et approches

De nombreux verrous freinent encore l'adoption de l'IA, particulièrement dans les systèmes critiques.

Expertise

La différence entre les principaux paradigmes de l'IA réside dans le fait que, dans un système d'IA dit symbolique, les connaissances sont explicitement définies par des experts, alors que dans une approche connexionniste et statistique (par exemple les réseaux de neurones profonds – *deep learning*), les connaissances sont déduites automatiquement à partir des données. David Sadek, VP de Thales pour l'IA, explique que « l'IA connexionniste était jusqu'à récemment l'IA des sens, et l'IA symbolique celle du sens ». En combinant ces deux approches, les systèmes peuvent bénéficier de la compréhensibilité et de la fiabilité de l'IA symbolique, tout en utilisant la flexibilité et la capacité de traitement massif de données des techniques d'apprentissage automatique pour améliorer la performance des décisions. Ces systèmes d'IA hybride peuvent également offrir une plus grande précision et un temps de réponse plus rapide que l'une ou l'autre approche utilisée seule. Cependant, l'équilibre entre l'exploitation de l'expertise métier, de la connaissance du domaine et la maîtrise des technologies d'IA reste encore un défi.

Fiabilité, maintenabilité, sûreté, sécurité et rapidité

Le déploiement de l'IA industrielle et responsable nécessite la conformité à des objectifs de fiabilité, de maintenabilité, de disponibilité, de sûreté et de sécurité, voire de rapidité. Ainsi, un tel système doit reposer sur des méthodes de développement bien fondées, de sa conception à son déploiement et sa qualification. Les pratiques d'ingénierie doivent être alors enrichies par des méthodes et outils garantissant la confiance à toutes les étapes du cycle de vie d'un tel système :

1. spécification du domaine opérationnel et sa déclinaison pour la gestion des données et des connaissances ;
2. conception d'algorithmes et d'architectures de référence ;
3. caractérisation, vérification et validation ;
4. déploiement, en particulier sur une architecture embarquée ;
5. qualification, certification ;
6. et maintien en condition opérationnelle et de cybersécurité.

Robustesse

La cyber-sécurisation des systèmes est primordiale. Il faut faire la preuve de la robustesse de l'IA à toute forme de malveillance, et notamment aux cyberattaques. Aujourd'hui, la sécurité entretient une relation bilatérale avec l'IA : il s'agit, d'une part, d'utiliser l'IA pour créer des algorithmes de cybersécurité, et, d'autre part, de cyber-sécuriser les systèmes d'IA. Les algorithmes fondés sur la *machine learning* ont la capacité d'identifier des anomalies, comme la préparation d'une cyberattaque, dans un flot d'événements, même avec des signaux faibles. Pour tout ce qui relève de la malveillance ou de la cyber-fraude, les deux processus avancent au même rythme : l'IA progresse en même temps que les programmes destinés à l'attaquer. Pour déjouer les risques de l'IA, Thales a mis sur pied une équipe de *friendly hacking* (piratage bienveillant). Sa mission est d'effectuer des sortes de *crash tests* des algorithmes d'IA, notamment des architectures neuronales, pour identifier leurs vulnérabilités et proposer des contre-mesures afin de rendre nos applications les plus robustes possibles. Une suite d'outils appelée «The Battle box» est en train d'être développée. Ce laboratoire de *crash tests* fait partie du processus de qualification des systèmes à base d'IA. Notons que cette équipe développe des IA pour attaquer les IA, ce qui peut sembler paradoxal. Le but ultime étant, évidemment, surtout de s'assurer de la robustesse des IA fonctionnelles développées dans l'ensemble du groupe.

Explicabilité

La question de l'explicabilité est un enjeu majeur. Le sujet n'est pas récent mais il est redevenu d'actualité avec l'IA connexionniste. Comment justifier les décisions du système ? Certains algorithmes sont si complexes que personne – y compris les experts – n'est capable d'en détailler le fonctionnement ! C'est la fameuse notion de boîte noire. Expliquer n'est pas tracer le calcul. Un autre défi est de montrer qu'un système a bien appris ce qu'on voulait qu'il apprenne. Nous travaillons sur l'explicabilité de l'IA connexionniste et statistique, notamment dans le cadre du laboratoire industriel Sinclair (Saclay INdustrial Collaborative Laboratory for Artificial Intelligence Research), labo commun avec EDF et Total. Cependant, le problème se pose aussi dans le cas de systèmes d'IA symbolique, qui combine beaucoup de règles métiers entre elles. Dans certains cas, l'explicabilité peut être générée par

un processus à base de connaissances capables d'extraire du réseau neuronal les concepts sous-jacents à la prise de décision. Une préoccupation est d'éviter absolument par une volonté d'explicabilité de créer des IA convaincantes bien que donnant des réponses erronées (certaines réponses des IA génératives peuvent être totalement fausses mais appuyées par des éléments scientifiquement cohérents).

Ingénierie de l'IA

Pour l'aide à la conception d'un composant d'IA et plus particulièrement à base d'apprentissage (ML : *machine learning*), un processus MLOps (*machine learning operations*), fortement inspiré de l'approche DevOps (*development operations*), a pour objectif primaire de parfaire le processus de développement et de déploiement logiciel, ceci afin d'améliorer et d'unifier le développement et l'exploitation du composant. L'un des avantages est qu'il permet un déploiement plus rapide. Cependant, son application dans le contexte de systèmes critiques doit être repensée pour assurer transparence et auditabilité, afin de corriger, si besoin, les défaillances, mais aussi pour aller jusqu'à la preuve de la conformité des propriétés attendues.

Pour répondre à ce défi, la France a lancé de nombreuses initiatives auxquelles Thales participe, comme ANITI¹, les projets DEEL (*dependable and explainable learning*) et Confiance.ai² permettant de définir des méthodologies outillées de l'ingénierie de l'IA fiable.

La spécification du problème

La spécification du problème, capturée au travers des différentes exigences (fonctionnelles / non fonctionnelles) et de la couverture opérationnelle (ODD : *operational design domain*), décrit l'ensemble des conditions dans lesquelles la capacité est conçue pour fonctionner correctement. Ceci impacte la tâche de collecte des données et de modélisation des connaissances.

L'acquisition des données / connaissances

L'acquisition des données / connaissances guidée par l'ODD aboutit à une agrégation des données / des connaissances en un ensemble homogène, de taille et de qualité (compréhensibles, pertinentes, fiables, équilibrées...) suffisantes. Cependant, celles-ci sont en général nettoyées, organisées, voire labellisées. Dans certains cas, un traitement est nécessaire afin de rendre ces informations brutes exploitables. Il s'agit de la tâche de *data engineering* pouvant être complétée par du *knowledge engineering*. Cela requiert une expertise métier / domaine, qui peut manquer à des équipes uniquement spécialisées dans les technologies de l'IA.

L'aide à la conception / paramétrisation d'un algorithme d'apprentissage

Même si un algorithme statistique ou connexionniste peut être conçu ou sélectionné dans une bibliothèque d'algorithmes, une fois l'apprentissage terminé, le modèle est affiné, utilisant l'ensemble des données

¹ <https://aniti.univ-toulouse.fr/>

² <https://www.confiance.ai/>

de validation. Cela peut impliquer la modification ou l'élimination de variables, l'ajustement des paramètres spécifiques du modèle jusqu'à un niveau de précision acceptable. L'implémentation sur la plate-forme matérielle et / ou le système cible peut impacter certaines exigences techniques comme la latence, l'espace mémoire ou la consommation énergétique.

L'évaluation de la confiance

Après avoir trouvé un ensemble acceptable de paramètres et optimisé la précision du modèle, ce dernier est testé et caractérisé sur un ensemble de données, voire évalué par une approche formelle. L'évaluation peut aller au-delà de la performance fonctionnelle (par exemple la précision) et englober des métriques relatives à tout autre critère de performance attendu comme la robustesse. Après intégration du composant IA/ML dans un système critique, il faut démontrer en apportant les preuves que celui-ci possède les propriétés de confiance attendues. Il faut donc définir un cadre d'analyse d'ingénierie système de l'IA industrielle et responsable, permettant d'élaborer les stratégies de développement de systèmes et d'IVVQ (intégration, vérification, validation, qualification).

Durabilité et éthique

Les exigences éthiques ne doivent pas être vécues comme des contraintes, mais comme des opportunités de création de valeur. La mise en œuvre repose sur plusieurs étapes (capture de données et / ou connaissances, pré-traitement basé sur l'expertise métier, apprentissage, inférence) qui peuvent avoir des poids énergétiques différents. Ainsi, en fonction des données ou des connaissances, différentes étapes de traitement peuvent être nécessaires avant de réaliser la conception du composant. Certaines méthodes, comme l'IA générative, peuvent nécessiter des ressources importantes et de très grandes durées d'apprentissage. Enfin, pour la phase d'usage (inférence), même si l'application peut être active quelques fois par jour, elle est utilisée par des millions de personnes, et nécessitera plus de ressources qu'une application mono-utilisateur. L'optimisation de la phase la plus impactante en consommation de ressources doit donc être réalisée avec attention en considérant le meilleur compromis entre précision et consommation d'énergie. Il est donc important de favoriser des approches d'IA frugale en données et en énergie. De même, Thales vise à développer des modèles les plus génériques possibles, puis de les instancier application par application. Cela permet de n'effectuer qu'une fois le travail le plus énergivore et de s'assurer que toutes les applications bénéficient d'un socle commun éprouvé, voire validé et qualifié indépendamment.

Mais on peut aussi utiliser l'IA pour réduire l'empreinte carbone. Nous avons aujourd'hui une position de *leader* dans le domaine des traînées humides des aéronefs (contrails) par exemple, afin de minimiser la génération de traînées de condensations humides, qui jouent un rôle important dans la part d'effet de serre

liée à l'aviation³. L'IA permet d'optimiser la trajectoire des avions pour en produire le moins possible. On utilise des technologies très avancées d'IA, comme de l'apprentissage par renforcement (RL : *reinforcement learning*), où on n'apprend plus sur la base de données mais sur la base d'expériences, champ extrêmement prometteur. Thales a également développé une approche IA pour réduire le temps de présence des avions en approche finale d'un aéroport, en tenant compte de l'effet vortex, et de fait réduire l'empreinte carbone.

Une autre dimension vise à garantir que les systèmes d'IA prennent des décisions justes et impartiales. Cela inclut l'élimination des biais discriminatoires dans les algorithmes d'IA, la transparence dans le processus de prise de décision, et la responsabilité en cas de préjudice causé par un système d'IA. C'est pourquoi il est crucial d'établir une culture d'ouverture et de responsabilité. Les parties prenantes devraient s'engager à divulguer les informations pertinentes concernant l'impact environnemental des technologies d'IA.

Conclusion

Fin 2024, Thales a lancé cortAlx pour intégrer l'IA dans l'ensemble des secteurs d'activité de l'entreprise – défense, spatial, aéronautique, cybersécurité, identité numérique. Cet accélérateur s'articule autour de trois piliers :

- cortAlx Labs, implanté en France, au Canada, au Royaume-Uni et à Singapour, avec comme promesse pour cortAlx Labs France d'être le plus puissant laboratoire intégré dans le domaine de l'IA critique en Europe.
- Une usine de recherche technologique, la cortAlx Factory, au service de l'accélération de la qualification et l'industrialisation des outils de développement de l'IA ainsi que les cas d'usage pour les données des systèmes, comme la planification de missions, la gestion du trafic aérien, le pilotage de drones et de robots...
- cortAlx Sensors focalisé sur les capacités IA dans les capteurs, couvrant principalement les activités de défense, offrant des capacités décuplées, à la fois plus précises et efficaces, en terme de perception et d'identification de la menace et des cibles.

³L'aviation représentait environ 2,5 % des émissions mondiales de CO₂eq en 2023. Son impact climatique provient de ses émissions de CO₂ et de celles non liées au CO₂ : NOx, particules (suie), SOx mais aussi la vapeur d'eau ainsi que les effets ultérieurs de la formation de traînées de condensation-cirrus et des interactions aérosol-nuage. La dernière évaluation de l'impact climatique des émissions historiques du trafic aérien entre 1940 et 2018 estime que les effets non CO₂ sont pour plus de la moitié du réchauffement dû à l'aviation, dont 57 % liés aux traînées humides, avec toutefois des incertitudes huit fois plus grandes que celles relatives au CO₂ et des niveaux de confiance « faibles » : https://www.easa.europa.eu/sites/default/files/eaer-downloads/EASA_EAER_2025_Book_v5.pdf pp. 57-59 et étude du Pr David Simon Lee *et al.* <https://pubs.rsc.org/en/content/articlehtml/2023/ea/d3ea00091e>

Blaxtair : le défi du véhicule industriel intelligent pour une industrie plus sûre et plus efficace

Par Sabri BAYOUDH

PhD, directeur de l'innovation, Arcure-Blaxtair

Les technologies de perception intelligentes embarquées, basées sur l'IA, sont en train de révolutionner le marché du véhicule industriel. Le premier cas d'usage de ces technologies est la sécurité, principalement la détection de piétons. Depuis 2009, avec Blaxtair, Arcure développe « l'œil intelligent » du véhicule, capable d'une précision inégalée dans les environnements industriels pourtant particulièrement complexes. Avec les solutions Blaxtair®, l'IA sauve déjà des vies dans l'industrie.

Au-delà de ce premier cas d'usage, l'apprentissage profond ("*deep learning*") et la vision par ordinateur ("*computer vision*") ouvrent la voie à la lecture globale de l'environnement par le véhicule. Qui plus est, ce véhicule intelligent est désormais à même de collecter des données à travers les sites dans lesquels il évolue, pour permettre un pilotage global des installations et de leur fonctionnement.

Les constructeurs le savent et sont désormais en ordre de marche pour intégrer de telles technologies : le véhicule industriel intelligent est en route pour occuper une place centrale dans l'industrie 4.0.

Depuis 2009, l'IA bouleverse la sécurité des véhicules industriels, un enjeu évidemment crucial. Les technologies de perception intelligentes embarquées, capables d'une précision inégalée, sont rapidement apparues comme l'instrument le plus puissant pour prévenir les accidents. Arcure-Blaxtair a ainsi développé le premier cas d'usage de l'IA embarquée pour l'industrie : sauver des vies.

Sur le plan technologique, cette première application a tracé la voie du véhicule industriel « intelligent ». Doté de capacités de perception associées à l'IA, l'engin devient capable d'analyser et d'interpréter les environnements industriels complexes et exigeants. Par son rôle central dans les opérations, il s'impose en outre comme un outil de collecte de données, en temps réel, au cœur de la digitalisation des sites industriels.

La sécurité des personnes, enjeu prioritaire de la digitalisation du véhicule industriel

C'est pour répondre aux impératifs de sécurité que l'IA a d'abord pénétré le secteur du véhicule industriel. L'accidentologie de ces engins reste moins connue que celle des voitures et autres véhicules routiers. Et pourtant, les statistiques démontrent l'importance de l'enjeu pour ces engins dits *off-road*, au vu du nombre d'accidents mais aussi de leur gravité.

Dans la construction et les travaux publics, en France, on recense chaque année 10 à 20 accidents graves et mortels sur les chantiers, à la suite de collisions entre engins de construction et piétons¹.

Au sein des sites industriels, la sécurité des opérations logistiques ou de manutention est un sujet de sérieuse préoccupation pour tous les responsables HSE. Aux États-Unis, l'Occupational Safety and Health Administration (OSHA) rapporte entre 35 000 et 62 000 blessures par an liées à un accident de chariot élévateur, et une moyenne de 87 décès annuels². 11 % des chariots élévateurs sont ainsi impliqués dans un accident chaque année³, un taux supérieur à celui de l'automobile ! Le *Wall Street Journal* s'en faisait l'écho en janvier 2025, signe d'une prise de conscience outre-Atlantique⁴.

Erreurs humaines, manque de visibilité du conducteur et nombreux angles morts, coactivité entre piétons et véhicules dans ces environnements encombrés sont les principales causes d'accidents (voir la Figure 1).

¹ Base EPICEA de l'Institut National pour la Recherche sur la Sécurité (INRS) et base AT de l'Organisme Professionnel de Prévention du Bâtiment et des Travaux Publics (OPPBT), <https://stopcollision.com/>, consulté le 20 février 2025.

² www.osha.com/blog/forklift-hazards-prevention, consulté le 20 février 2025.

³ OSHA, Boston Regional Notice, 2018, p. 5, www.osha.gov/sites/default/files/enforcement/directives/CPL_04-00-023F.pdf

⁴ KEILMAN J. (2025), "Forklifts hurt thousands of workers each year. Factories are seeking alternatives", *The Wall Street Journal*, Jan. 5.



Figure 1 : Deux situations à risque, engageant la sécurité de piétons évoluant à proximité d'engins
(Source : Arcure - Blaxtair).



Figure 2 : Blaxtair®, première technologie d'IA au monde développée pour la sécurité des véhicules en milieu industriel ; ici, la dernière génération
(Source : Arcure - Blaxtair).

Les premières conséquences sont évidemment humaines. Assurer la sécurité des personnes est un impératif éthique. Mais les accidents ont aussi un impact sur la productivité. Mises à l'arrêt temporaire du site, arrêts maladie, inaptitudes partielles ou totales des victimes sont autant d'effets collatéraux sur le fonctionnement des entreprises. Aux États-Unis, on estime par exemple à 18 jours la durée médiane des arrêts maladie à la suite d'un accident avec un chariot élévateur⁵. Sur le plan financier, les accidents génèrent des coûts directs (indemnités, frais médicaux) et indirects (coût matériel, amendes le cas échéant, primes d'assurances, pénalités de retard de livraison, etc.).

Ces enjeux sont identifiés par les entreprises. La culture de la sécurité s'y développe, avec la mise en place de technologies spécifiquement dédiées.

Le premier cas d'usage de l'IA embarquée : la détection de piétons

Les solutions traditionnelles de sécurité, telles que la signalétique, les radars, les caméras de recul, les ultrasons, le LIDAR ou encore les badges à puce RFID portés par les piétons, ont pu constituer des recours pour la prévention des accidents engins / piétons sur les sites industriels.

Les études disponibles montrent toutefois que ces technologies présentent toujours des limites en termes de couverture, de réactivité ou de précision dans des environnements complexes⁶. Les radars, lasers, ultrasons et caméras simples sont certes des dispositifs de levée d'angle mort, mais ne seraient utiles que dans 60 % des situations à risque⁷.

⁵ National Safety Council, <https://injuryfacts.nsc.org/work/safety-topics/forklifts/>

⁶ STOPCOLLISION, Étude Expleo, Projet de suppression des heurts engins-piétons, mars 2024, consulté sur <https://stopcollision.com>

⁷ MARSOT J., CHARPENTIER P. & TISSOT C. (2009), « Collision engins/piétons, analyse des récits d'accidents de la base EPICEA », *Hygiène et sécurité au travail*, INRS.

C'est au tournant des années 2010 que les technologies d'IA embarquée ont changé le paradigme de la sécurité périmétrique des véhicules industriels.

Couplées à des caméras 2D ou 3D aptes à la prise d'images en milieu industriel, les technologies de vision par ordinateur ("*computer vision*") ont permis la détection spécifique de piétons dans les environnements complexes.

Apprentissage machine ("*machine learning*"), puis apprentissage profond ("*deep learning*") ont permis l'amélioration continue du niveau de performance de ces systèmes et leur intégration fluide dans les opérations de conduite, avec un taux de fausses alarmes en baisse continue.

Cette première application a supposé le développement de technologies d'IA pionnières pour l'industrie, basées sur la vision par ordinateur, qui ont donné pour la première fois aux véhicules industriels la capacité de « décrypter » leur environnement immédiat.

« L'œil intelligent » du véhicule industriel : le défi technologique Blaxtair®

La vision par ordinateur, ou "*computer vision*", est une technologie clé pour l'autonomisation des machines et des véhicules. Ce domaine de l'intelligence artificielle utilise

des algorithmes d'analyse d'images ou des vidéos, identifiant objets, visages ou actions. Parmi les algorithmes indispensables figurent ceux dits de « segmentation », entraînés pour décomposer chaque image en différentes zones d'intérêt et, *in fine*, détecter des objets prédéfinis.

Mais, en matière d'analyse d'images, le milieu industriel présente une complexité sans commune mesure : espaces encombrés, relief, multiplicités des postures possibles pour les piétons et les véhicules, occultations, lumière insuffisante, poussières, viennent considérablement compliquer la tâche (voir la Figure 3).

Le développement d'applications de vision par ordinateur en milieu industriel requiert donc à la fois une puissance de calcul bien plus élevée et une base de données d'apprentissage bien plus robuste, vaste et variée.



Figure 3 : L'identification de piétons sur des images captées sur la route et en milieu industriel ; alors que les images en milieu routier sont lisibles et standardisées, les images en milieu industriel sont bien plus complexes et d'une variabilité bien plus élevée (Source : Arcure - Blaxtair).

La puissance de calcul nécessaire aux applications industrielles est une affaire de matériel. Sur ce plan, l'évolution récente des puces permet aujourd'hui d'embarquer dans des caméras compactes la puissance des imposants calculateurs utilisés il y a dix ans (plusieurs TOPS) pour une consommation très faible de quelques Watts.

La base de données d'apprentissage en milieu industriel

Construire la base de données est en revanche une affaire de savoir-faire, qui pose d'abord la question de la collecte des images et donc de l'accès aux sites industriels privés. Une fois les images recueillies, leur administration et leur incorporation à la base d'entraînement suppose une excellente maîtrise des principes et des méthodes de l'apprentissage profond.

À titre d'exemple, la base de données Blaxtair® contient aujourd'hui des millions d'images, collectées sur des centaines de sites industriels depuis 2009 et traitées selon des pratiques rigoureuses : contrôle de la résolution, annotation, gestion de la variété et de l'équilibre des variations, séparation en sous-ensembles, etc.



Figure 4 : Un exemple de la variété des images à intégrer dans une base de données en milieu industriel : posture inclinée ou droite, occultation, couleur de casque ou de gilet figurent parmi les nombreux paramètres à prendre en compte pour la détection de piétons (Source : Arcure - Blaxtair).

Grâce à sa base de données d'apprentissage unique dans le monde, la détection de piétons Blaxtair® atteint aujourd'hui une spécificité sans équivalent, avec un taux de fausses alarmes inférieur à une seule occurrence par jour. Un tel niveau de performance atteste que la base couvre aujourd'hui presque toutes les situations possibles en milieu industriel pour ce cas d'usage, jusqu'aux piétons partiellement occultés.

De la « classification » à la « perception » : vers la lecture intelligente globale de l'environnement

Le succès de cette première application de l'IA ouvre la voie à de nouveaux développements technologiques et de nouveaux cas d'usage.

Au-delà de la détection de piétons, les technologies d'IA embarquée en milieu industriel sont aujourd'hui capables de reconnaître d'autres objets, comme les véhicules commerciaux circulant dans les sites, ou de décrypter des scènes complexes (voir la Figure 5).

Dans le champ de la sécurité, on peut aujourd'hui différencier les situations de danger de celles qui ne le sont pas et atteindre un niveau d'intelligence supérieur : la perception du risque. Le piéton détecté est-il dans la trajectoire du véhicule ; a-t-il regardé le véhicule ; est-il protégé par une barrière ; est-il situé dans un autre véhicule ? L'IA permet aujourd'hui de répondre à ces questions.

Augmenté grâce à ces technologies de perception embarquées, le véhicule industriel peut donc interagir de plus en plus largement avec son environnement, pour son propre compte. Mais il devient aussi un agent de collecte d'informations à travers le site industriel dans lequel il évolue, ouvrant la voie à l'analyse « débarquée » de données, pour un pilotage global des sites possiblement en temps réel.

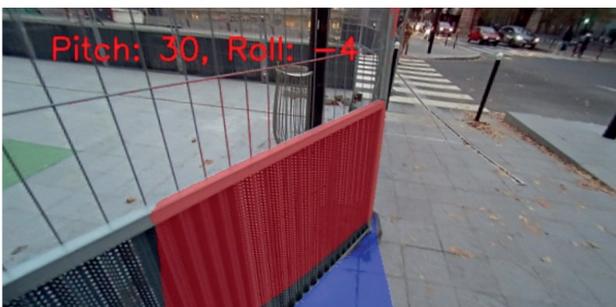


Figure 5 : La lecture intelligente de l'environnement : reconnaissance et localisation d'objets et de scènes permettent désormais l'analyse de situations complexes (Source : Arcure - Blaxtair).

IoT et données : du véhicule intelligent au site industriel intelligent

La collecte de données au sein d'un site industriel repose sur une technologie clé, intégrée depuis la fin des années 2010 dans les systèmes embarqués sur les véhicules : l'internet des objets (IoT, *Internet of Things*) et la connectivité au *cloud*, notamment permise par l'essor de la 5G.

L'intégration de ces données dans des systèmes d'aide à la décision permet aux gestionnaires de piloter efficacement leurs sites, d'optimiser les processus de maintenance et de renforcer la sécurité globale. La digitalisation des véhicules industriels ne se limite donc pas à une amélioration technique, mais s'inscrit dans une démarche de transformation opérationnelle.

Blaxtair Connect®, la première plateforme de gestion globale de la sécurité

De tels outils existent déjà dans le champ de la sécurité. Accidents, situations de presque accident, chocs avec le mobilier industriel (un rack par exemple), nombre de piétons détectés sur le parcours du véhicule sont des informations précieuses, qui, une fois traitées, offrent une vision globale du fonctionnement des installations et des risques quotidiens.

Traitées par une plateforme, ces données permettent la gestion des flottes de véhicules industriels ainsi qu'une cartographie des risques à l'échelle du site ou de l'entreprise, pour la mise en place d'actions curatives ou préventives.

L'approche préventive, illustrée par la pyramide de Bird (voir la Figure 6), permet de réduire significativement les incidents, comme le démontrent les indicateurs clés mis en œuvre par certains grands groupes, où la coactivité engins / piétons est divisée par trois.

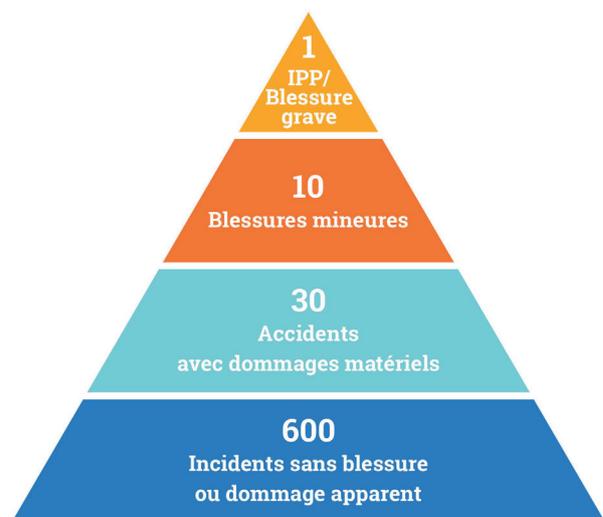


Figure 6 : La pyramide de Bird démontre l'efficacité d'actions préventives, à sa base, pour réduire les risques les plus graves : c'est ce que permet l'analyse des données à l'échelle d'un site, par la plateforme Blaxtair Connect® ci-après (p. 80) (Source : Arcure - Blaxtair).

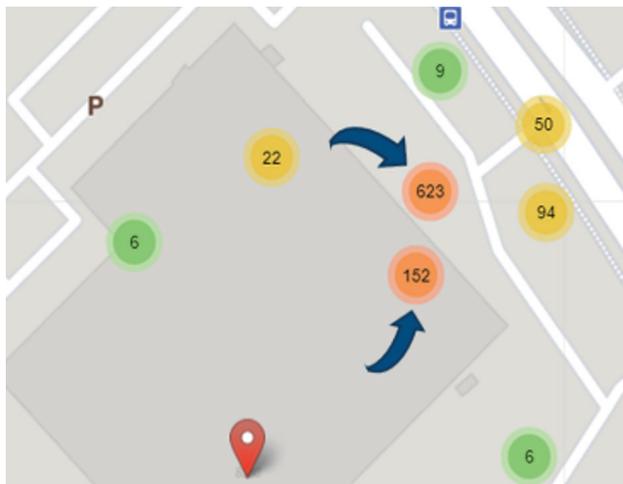


Figure 6 : La pyramide de Bird ci-avant (p. 79) démontre l'efficacité d'actions préventives, à sa base, pour réduire les risques les plus graves : c'est ce que permet l'analyse des données à l'échelle d'un site, par la plateforme Blaxtair Connect® à droite (Source : Arcure - Blaxtair).

Cet exemple d'un premier traitement global de données pour la sécurité préfigure d'autres modélisations, porteuses d'un vaste potentiel d'applications.

Les jumeaux numériques dynamiques : un gisement d'applications inestimé

Longtemps perçus comme un concept futuriste, les jumeaux numériques sont désormais devenus des réalités, par exemple dans l'industrie⁸.

Grâce aux technologies de perception intelligentes qu'ils embarquent, les véhicules industriels permettront la reproduction visuelle en 3D du site industriel dans lequel ils opèrent, de manière dynamique et en temps réel. Interfacés avec des applications d'IA générative, ces jumeaux numériques offriront des capacités d'analyse et / ou de prévision encore inégalées.

Travailleurs et gestionnaires pourront expérimenter ces nouvelles capacités et définir les cas d'usage pertinents : amélioration de l'efficacité énergétique, réduction des risques d'accidents, optimisation des flux logistiques, maintenance du site, etc.

Conclusion

L'IA et les technologies de perception intelligentes embarquées constituent une révolution en cours pour le véhicule industriel. Et demain ? Ce véhicule industriel « augmenté » sera-t-il totalement autonome ? La promesse est séduisante, mais incertaine. Dans un futur proche, parions déjà sur la généralisation d'un véhicule industriel toujours plus intelligent, au service d'un opérateur rendu toujours plus efficace.

Surtout, ce véhicule ainsi digitalisé et connecté opérera au cœur d'un site industriel qu'il aura lui-même contribué à digitaliser. Au-delà des fantasmes, la transformation des

environnements de travail par ces innovations constitue déjà, aujourd'hui, un levier considérable et bien tangible pour améliorer la sécurité des personnes, des installations, et accroître la productivité et l'efficacité globale des sites.

Les constructeurs d'engins ne s'y trompent pas. Tous ont désormais une feuille de route pour l'intégration de technologies intelligentes dans leurs véhicules neufs. La révolution du véhicule industriel intelligent est en route.

Bibliographie

1. Base EPICEA de l'Institut National pour la Recherche sur la Sécurité (INRS) et base AT de l'Organisme Professionnel de Prévention du Bâtiment et des Travaux Publics (OPPBT), <https://stopcollision.com/>, consulté le 20 février 2025.
2. www.osha.com/blog/forklift-hazards-prevention, consulté le 20 février 2025.
3. OSHA, Boston Regional Notice, 2018, p. 5, www.osha.gov/sites/default/files/enforcement/directives/CPL_04-00-023F.pdf
4. KEILMAN J. (2025), "Forklifts hurt thousands of workers each year. Factories are seeking alternatives", *The Wall Street Journal*, Jan. 5.
5. National Safety Council, <https://injuryfacts.nsc.org/work/safety-topics/forklifts/>
6. STOPCOLLISION, Étude Expleo, Projet de suppression des heurts engins-piétons, mars 2024, consulté sur <https://stopcollision.com>
7. MARSOT J., CHARPENTIER P. & TISSOT C. (2009), « Collision engins/piétons, analyse des récits d'accidents de la base EPICEA », *Hygiène et sécurité au travail*, INRS.
8. *Enjeux numériques*, décembre 2024, <https://annales-des-mines.org/2024-12-sommaire-numerique-reindustrialisation-enjeux-numeriques/>

⁸Voir par exemple les articles sur les jumeaux numériques dans *Enjeux numériques*, décembre 2024, <https://annales-des-mines.org/2024-12-sommaire-numerique-reindustrialisation-enjeux-numeriques/>

Mobilisation de l'intelligence artificielle pour les transports publics : un intérêt opérationnel évident, qui doit rester tourné vers les services aux voyageurs

Par **Hélène BRISSET**

Directrice du Numérique d'Île-de-France Mobilités

L'Île-de-France constitue l'un des deux plus grands réseaux de transport au monde. Île-de-France Mobilités est responsable de l'organisation, la modernisation et du développement de ce réseau, qui continue de se développer pour répondre aux besoins de plus de 10 millions de déplacements quotidiens des Franciliens.

L'intelligence artificielle (IA) y prend aujourd'hui une part croissante. Des applications comme Île-de-France Mobilités ou Transport Public Paris 2024 permettent de donner des infos en temps réel, d'offrir des traductions automatiques et de répondre directement aux demandes des usagers. Le numéro unifié offre un parcours simplifié pour répondre aux demandes téléphoniques des voyageurs. L'IA est aussi utilisée pour suivre l'affluence, améliorer la sécurité et optimiser le fonctionnement du réseau en cas de problème.

L'IA est un levier structurant pour la transformation des mobilités. Son déploiement doit se poursuivre avec un focus constant sur l'amélioration de l'expérience des usagers, et pas seulement avec un objectif technologique.

Les transports en Île-de-France, un enjeu quotidien majeur

L'Île-de-France constitue aujourd'hui la deuxième plaque au monde pour la mobilité. Le réseau francilien est le deuxième réseau le plus fréquenté et le plus étendu après celui de la métropole de Tokyo. Ce réseau comporte aujourd'hui 2 149 km de voies ferrées, avec 14 lignes de métros, 9 lignes de tramways et 13 lignes de trains et RER, et 1 900 lignes de bus, qui irriguent la région. Il est en forte extension et transformation, notamment avec l'ensemble des lignes du Grand Paris Express.



Figure 1 : L'Île-de-France constitue le deuxième réseau de transport au monde (Source : Île-de-France Mobilités).

Chaque jour, près de 10 millions de déplacements ont lieu en transports en commun. Les transports ont un impact majeur sur la vie des Franciliens et plus largement de tous les voyageurs en Île-de-France. Or, les besoins des Franciliens en matière de transport et de mobilité sont en constante évolution.

Autorité organisatrice des mobilités, Île-de-France Mobilités imagine, organise et finance les transports pour tous les Franciliens. Elle pilote l'ensemble des grands programmes de modernisation des transports tels que le renouvellement des trains ou des rames de métro, la mise en accessibilité des gares, le remplacement des bus diesel par des bus propres... Elle décide des prolongements et des créations des nouvelles lignes.

Le fonctionnement opérationnel des différentes lignes est directement piloté par les opérateurs, pour chacun des modes et réseaux qui leur sont confiés par contrat : bus, tram, métro, RER, vélo, parkings relais... Ils sont responsables du bon fonctionnement des lignes qui leur sont confiées, et d'atteindre les niveaux de qualité de service qui leur sont fixés (ponctualité, accessibilité, information, sécurité, propreté). Île-de-France Mobilités coordonne ainsi le bon fonctionnement des transports franciliens en lien avec l'ensemble des opérateurs, et gère les transversalités.

Les données sont cruciales pour le bon fonctionnement du réseau et la diffusion de l'information voyageur

Le réseau de transport et ses différents modes constituent bien sûr d'abord un monde physique, mais très largement piloté par le numérique, et pour lequel les données jouent naturellement un rôle central. D'une part, parce que les systèmes automatiques génèrent massivement des *data*, d'autre part, parce que les données sont cruciales pour modéliser le fonctionnement du réseau, et pour fournir en temps réel une information voyageur pertinente.

Île-de-France Mobilités propose ainsi à l'ensemble des Franciliens une application mobile¹ permettant de faciliter les déplacements en Île-de-France, avec notamment une information de l'actualité en temps réel du trafic, l'achat de titres de transport dématérialisés ou encore la préparation et la planification des déplacements *via* notamment un moteur de recherche en temps réel des meilleurs itinéraires en transport en commun, covoiturage ou vélo. L'ensemble de ces services et données sont également ouverts au public et aux entreprises de services numériques de mobilité *via* le portail régional d'information pour la mobilité (PRIM).

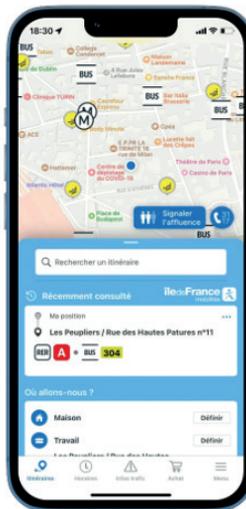


Figure 2 : L'application Île-de-France Mobilités permet la recherche d'itinéraires et l'acquisition de titres : elle est la deuxième application de mobilité la plus utilisée en Île-de-France (Source : Île-de-France Mobilités).

Pour proposer ces services, Île-de-France Mobilités s'appuie sur des référentiels et un *hub* de données en temps réel qui collecte, agrège, met en qualité et distribue les données en provenance des opérateurs de transport. Chaque jour, ce sont plus de 27 millions de notifications qui sont traitées par la plateforme.

Alors, que peut faire l'intelligence artificielle dans le champ des transports ?

Dans notre secteur à très forte vocation opérationnelle et tourné vers les voyageurs, l'IA n'est bien sûr utile que si elle permet d'améliorer le service aux usagers. Aussi,

¹ <https://www.iledefrance-mobilites.fr/>

permettez-moi d'enfoncer ici quelques portes ouvertes : on ne fait pas de l'IA juste pour faire de l'IA, mais parce qu'elle présente un intérêt serviciel ou opérationnel, parce que c'est un vrai accélérateur ou raccourci pour répondre à des problématiques complexes. Deuxième porte ouverte : on ne réinvente pas l'eau chaude, on ne redéveloppe pas ce qui existe déjà et qui fonctionne très bien et que l'on peut simplement réutiliser. Nos déploiements s'inscrivent, autant que faire se peut, dans l'état de l'art, et si des déploiements existants conviennent, c'est un excellent point d'appui pour se concentrer sur d'autres spécificités.

Plusieurs intérêts majeurs de l'IA dans le champ des transports publics se détachent d'ores et déjà. Il s'agit notamment de mieux accompagner les voyageurs dans leurs déplacements, de masquer la complexité d'un réseau très étendu et diversifié, de mieux sécuriser les déplacements, etc. Pour les autorités organisatrices des mobilités, l'IA est aussi source d'opportunités pour renforcer leur capacité à organiser les transports, au travers, par exemple, d'une meilleure maîtrise des contrats ou du patrimoine matériel, ceci dans un contexte d'ouverture à la concurrence, démultipliant les acteurs à coordonner.

L'IA permet de mieux accompagner les voyageurs dans leurs déplacements

Certains usages peuvent directement s'appuyer sur des technologies et savoir-faire existants, avec un faible besoin de *tuning*.



Figure 3 : L'application Transport Public Paris 2024 a été déployée spécifiquement pour les JOP (Source : Transport Public Paris 2024).

Par exemple, l'application Transport Public Paris 2024, déployée spécifiquement par Île-de-France Mobilités pour les spectateurs des Jeux Olympiques et Paralympiques (JOP), pour la recherche d'itinéraires vers les sites olympiques et l'acquisition de titres de transport, a nativement intégré des briques de traduction automatique par IA. Ces briques, reconnaissant la langue du téléphone pour les principales langues



Figure 4 : L'application détecte automatiquement la langue du téléphone et offre une traduction directe par IA de l'ensemble du contenu (Source : Transport Public Paris 2024).

européennes², ont nécessité peu d'entraînement spécifique, au-delà de la connaissance des termes à ne pas traduire, comme les arrêts et lignes spécifiques à l'Île-de-France, tels que par exemple Hôtel de Ville, Châtelet, École militaire, Stade de France, etc. Le déploiement de briques DeepL et d'AWS Translate a permis très rapidement de déployer une information plus ciblée, avec un coût très faible et sans développement spécifique.

Un autre exemple de déploiement de l'IA au service des usagers, *via* l'application dite « Tradivia », permet aux agents de la RATP de comprendre les attentes des usagers dans 17 langues et de leur répondre avec une traduction appuyée par l'IA. Ce déploiement a été réalisé au premier trimestre 2024 et a facilité l'accompagnement opérationnel des spectateurs des JOP.

Enfin, et surtout, un agent autonome, dont le premier prototype a été présenté à VivaTech en mai 2024, sera bientôt déployé dans l'application Île-de-France Mobilités. Cet agent conversationnel intelligent comprend la demande exprimée, et, pour construire une réponse ciblée, utilise en temps réel l'ensemble des interfaces interactives existantes (les *application programming interface* ou API) pour donner le parcours le plus adapté, l'information en temps réel du prochain passage sur la ligne, le coût du billet pour telle destination. Au-delà de la qualité formelle des réponses, déjà travaillée pour un certain nombre de *chatbots* conversationnels, et de sa pertinence dans le champ des transports, l'intérêt de cet agent est de fournir une réponse adaptée, interactive en temps réel. Cet agent autonome s'inscrira en pleine

complémentarité avec l'application, dont il pourrait à terme remplacer très directement tout ou partie des fonctionnalités, avec une interface directement interactive.

L'IA permet de développer des nouveaux services à valeur ajoutée

Le calcul de l'affluence sur le réseau et sa diffusion aux usagers en quasi temps réel leur permettent de se positionner sur les rames les moins fréquentées et d'améliorer leur confort lors des trajets. Ce dispositif, appelé Affluences, a été déployé début 2024 par la SNCF puis la RATP sur les premières lignes pilotes et a reçu un accueil très favorable des usagers. Dans ce cadre, l'IA permet de mesurer la densité des rames par des caméras positionnées en bout de quai et de transmettre une mesure statistique d'occupation des rames : vert, orange, rouge. Il s'agit, dans ce cadre, de données purement statistiques, sans impact sur la vie privée.

L'IA permet de masquer la complexité d'un réseau très diversifié et très étendu

Les interactions avec nos usagers se font certes de plus en plus au format numérique ; pour autant, il est nécessaire de conserver une capacité d'interaction téléphonique, tant pour les profils plus éloignés du numérique que pour la résolution de situations complexes.

²La traduction par IA a été déployée en allemand, espagnol, portugais, italien, en complément des deux langues olympiques officielles, qui sont également celles de l'application, le français et l'anglais.

Aujourd'hui, sur le réseau très étendu qu'est l'Île-de-France, le service client téléphonique est géré par 106 numéros d'appel, directement pilotés par les opérateurs en charge du fonctionnement du service.

Or il n'est pas souhaitable de concentrer les actions sur un service client unique, généraliste, qui serait potentiellement moins pertinent pour gérer les demandes des voyageurs, que des services spécialisés plus pertinents. La mise en place d'un serveur vocal interactif où l'on demanderait à l'utilisateur de s'orienter tout seul, en tapant 1 jusqu'à 106, relèverait d'un monde kafkaïen...

Dans ce cas très précis, l'IA a un intérêt très fort pour masquer la complexité du réseau et comprendre très directement la demande de l'utilisateur. Un numéro unifié a ainsi été déployé en pilote depuis le printemps 2024. Un service interactif d'IA comprend la demande exprimée en langage naturel, avec l'expression directe de l'utilisateur. Cette expression se fait souvent en environnement bruyant, généralement en milieu urbain. L'IA qualifie la demande et l'aiguille vers le service client compétent pour traiter la demande.

Plusieurs éléments de retour d'expérience semblent ici utiles : l'IA doit effectivement permettre de fluidifier le parcours usager ; le temps de parcours global est un des indicateurs clés de ce déploiement.

Également, il s'agit d'éviter de faire répéter des demandes, donc le transfert se fait soit après une qualification poussée quand le service client concerné le permet, soit par un transfert le plus rapide possible pour éviter des répétitions ultérieures : par exemple, on n'a pas besoin de formaliser qu'il s'agit précisément d'un parapluie bleu pour qualifier qu'il s'agit d'objets trouvés, quand la demande sera instruite par le service client pertinent.

L'IA permet de mieux sécuriser les parcours

La loi « relative aux jeux Olympiques et Paralympiques » du 19 mai 2023 a notamment autorisé et encadré une expérimentation de la vidéoprotection intelligente pour détecter les situations à risque, en appui de la sûreté.

Cette expérimentation a été conduite sous le pilotage du ministère de l'Intérieur. Autorisée par la loi sur jusqu'à huit familles de cas d'usage, elle a fait l'objet, dans le champ des transports, d'un déploiement opérationnel ciblé sur quatre cas d'usage.

Il s'agit de :

- l'intrusion, le franchissement dans une zone ;
- la densité de personnes / l'affluence dans les moyens de transport ;
- la détection d'objets délaissés ;
- et la détection des mouvements de foule.

Des expérimentations ont été réalisées avec la RATP et la SNCF lors d'événements ciblés en amont des JOP (par exemple, Roland-Garros, plusieurs concerts).

En effet, l'IA nécessite un entraînement suffisant pour être pertinente dans les situations considérées.

Plusieurs points peuvent d'ores et déjà être signalés : l'IA apporte très clairement une plus-value opérationnelle sur plusieurs cas d'usage. En effet, il est humainement impossible de surveiller manuellement l'ensemble des caméras du réseau.

L'appui par l'IA permet d'avoir une composante sûreté mobilisable par des patrouilles vidéo, en complément des équipes terrain et d'avoir une capacité de détection des événements, pas juste de réaction et d'analyse *a posteriori*.

L'IA a très clairement détecté des franchissements de ligne, ce qui correspond principalement à des personnes sur les voies ou très proches des voies, donc avec un risque physique clair d'accidents de personnes.

La détection de l'affluence aussi semble un cas d'usage pertinent et qui permet d'alerter les équipes en charge de la régulation.

D'autres cas d'usage nécessitent un entraînement plus ciblé, c'est notamment le cas des objets délaissés, où cela a peu de sens de détecter juste une valise sans pouvoir faire le lien avec son ou sa propriétaire. Cela constitue un champ d'investigation complémentaire, car les objets délaissés représentent une des principales causes des retards sur le réseau, avec de l'ordre de quatre événements par jour sur le réseau RATP (donnée RATP 2022), chacun engendrant une forte perturbation du trafic pendant une heure en moyenne³. Les algorithmes mobilisés dans le cadre de l'expérimentation ont par ailleurs montré leurs limites, avec certaines confusions entre des humains et les chaises sur lesquels ils étaient assis.

Un premier bilan a été établi *via* un comité d'évaluation institué par la loi olympique sus-citée. La prolongation de l'expérimentation au-delà du 31 mars 2025 a été par ailleurs discutée dans le cadre de la proposition de loi Tabarot.

L'IA permet de modéliser un réseau de plus en plus étendu et de plus en plus complexe

L'intelligence artificielle peut aussi servir à modéliser l'impact des dysfonctionnements sur une branche du réseau, et notamment le report de trafic. L'ensemble des connecteurs de temps réel et la masse de données d'information voyageur disponible sont mobilisés dans le cadre de cette modélisation, qui a notamment été utilisée dans le cadre des JOP.

Cette modélisation est à ses débuts, avec un intérêt opérationnel très clair, mais un besoin d'entraînement ciblé pour tenir compte de la complexité et de la spécificité du réseau francilien.

³ <https://www.ratp.fr/bagage-oublie>

Également, un travail est en cours pour modéliser l'impact du trafic sur la rémunération des opérateurs, lié au nombre de validations sur la partie du réseau qu'ils exploitent.

Conclusion

Ce court panorama permet de confirmer l'intérêt opérationnel de l'intelligence artificielle dans le champ des transports publics.

Le contexte est bien sûr favorable, avec de très nombreuses données et un potentiel d'impact très fort. Certains cas d'usage peuvent être déployés très rapidement, par exemple certaines modélisations en langage naturel NLP, d'autres nécessitent un entraînement ciblé au vu de la complexité et de la spécificité du réseau francilien.

Les potentialités de l'IA augmentent à mesure que les activités et systèmes sont en capacité de fournir des données structurées, normalisées et interopérables qui permettent d'alimenter les algorithmes de manière efficace. Le champ d'application va ainsi s'étendre progressivement à des thématiques comme la prédiction d'affluence, la maintenance prédictive des matériels et des infrastructures ou encore la détection de la fraude.

Dans ce secteur, comme dans beaucoup d'autres, nous sommes au début de la transformation par l'IA. On peut avancer résolument à condition de ne pas perdre de vue les enjeux serviciels, et notamment en gardant systématiquement l'intérêt du voyageur au cœur de nos actions.

Qu'est-ce que la "deep tech" ?

Définitions et enjeux stratégiques

Par Benjamin CABANES

Enseignant-chercheur en sciences de gestion à Mines Paris – PSL et à l'École polytechnique

Le terme "deep tech" s'impose aujourd'hui comme une notion incontournable dans l'univers des technologies, de l'innovation et des affaires. Il suscite un engouement croissant de la part des investisseurs et des acteurs de l'innovation, qui y perçoivent une formidable opportunité de transformations profondes et durables. Cet article propose d'explorer les différentes définitions du concept de "deep tech", d'en analyser les spécificités et les implications réelles, ainsi que d'en dégager les principaux enjeux stratégiques.

Introduction

En France et en Europe, la *deep tech* est au cœur de toutes les attentions, mobilisant pouvoirs publics, investisseurs privés, cabinets de conseil, universités et organismes de recherche. Considérée comme un levier stratégique de compétitivité et d'indépendance technologique, elle fait l'objet de nombreuses initiatives visant à accélérer son émergence et son développement.

En 2019, dans le cadre du plan France 2030, l'État français lance le Plan DeepTech, opéré par Bpifrance, avec trois objectifs majeurs : accélérer la création de 500 start-ups *deeptech* par an d'ici 2030, soutenir leur croissance grâce à des financements et un accompagnement adapté, et renforcer les interactions entre recherche, entrepreneuriat et industrie. Six ans plus tard, en 2025, la France s'impose comme un acteur clé de la *deep tech* en Europe. Avec 2 589 start-ups *deeptech* actives et 385 nouvelles entreprises créées en 2024, elle représente 20 % des levées de fonds *deeptech* sur le continent, affirmant ainsi son rôle de moteur de l'innovation technologique en Europe¹.

En Europe, environ 9 milliards d'euros ont été investis dans des startups *deeptech* à travers 454 opérations en 2024². Ces investissements concernent notamment l'intelligence artificielle, avec des startups comme Mistral AI (468 M€ en série B) et Aqemia (30 M€ en série A) ; l'industrie spatiale, avec The Exploration Company (150 M€ en série B) ; ainsi que le quantique, avec Quantinuum (273 M€ en série D) et Riverlane (70 M€ en série C).

Pour certains acteurs, la *deep tech* représente un véritable changement de paradigme, une « quatrième vague de l'innovation »³. Après les premières révolutions industrielles (première vague), l'essor des laboratoires industriels et grandes entreprises comme Bell Labs, IBM ou Xerox PARC (deuxième vague), plusieurs décennies dominées par les technologies de l'information, le numérique et les biotechnologies (troisième vague), cette quatrième vague se distingue par la convergence de plusieurs technologies de rupture telles que l'informatique quantique, les nanotechnologies et les biotechnologies.

Selon les acteurs, la *deep tech* peut être, tour à tour, un type d'entreprise, un secteur technologique, un projet d'innovation de rupture, une politique publique, un nouveau paradigme d'innovation ou encore une nouvelle révolution industrielle. Bien que le concept de "deep tech" puisse paraître flou et ambigu, il est néanmoins possible d'identifier au moins deux façons distinctes de le définir plus précisément.

La deep tech en tant que secteur technologique

Il est possible de définir la *deep tech* en énumérant les éléments qui en font partie, qu'il s'agisse d'innovations ou de secteurs technologiques, notamment : la biotechnologie, l'intelligence artificielle, la robotique, l'informatique quantique, les technologies propres et les sciences des matériaux. Cette approche repose sur une classification par exemples concrets ou sous-catégories, illustrant ainsi les différentes facettes du concept de *deep tech*.

¹ FONTANILLAS J. & FONTANILLAS J. (2025, March 21), « 6 ans du Plan DeepTech : une filière française structurée, moteur de la dynamique deeptech européenne », Bpifrance 2024 | Presse, <https://presse.bpifrance.fr/6-ans-du-plan-deeptech-une-filiere-francaise-structuree-moteur-de-la-dynamique-deeptech-europeenne>.

² AVOLTA & FRANCE DEEPTech (2024), European DeepTech Trends, <https://avolta.io/european-deeptech-trends-2024/>

³ BCG (2021), Riding the deep tech wave. International Business Summit, Versailles, <https://www.tresor.economie.gouv.fr/Articles/e21012fa-faf8-4fad-93b8-691e130749df/files/4f30b0cd-7c70-40d5-8c04-b69324146c93>

THE DEEP TECH MATRIX

The following matrix illustrates the scope for Deep Tech course selection, where **Deep Tech technologies** are applied within **Applications to Global Challenges**. The are not prescriptive: further technologies and sectors can be defined as needed and agreed. There may be an overlap observed between technology areas and applications, which is understandable given their rapid development. This initial matrix correlates the sectors with the main focus of the EIT KICs.

DeepTech Technologies	DeepTech Applications to Global Challenges									
	EIT Climate	EIT Culture & Creativity	EIT Digital	EIT Energy	EIT Food	EIT Health	EIT Manufacturing	EIT Raw Materials	EIT Urban Mobility	EIT Water
Advanced Computing / Quantum Computing										
Advanced Manufacturing										
Advanced Materials										
Aerospace, Automotive and Remote Sensing										
Artificial Intelligence, Machine Learning, Big Data										
Biotechnology and Life Sciences										
Communications and Networks, including 5G										
Cybersecurity and Data Protection										
Electronics and Photonics										
Internet of Things, W3C, Semantic Web										
Robotics										
Semiconductors (Microchips)										
Sustainable Energy and Clean Technologies										
Virtual Reality, Augmented Reality, Metaverse										
Web 3.0, Blockchain, Distributed Ledgers, NFTs										



Figure 1 : Deep Tech Matrix, 2023 (Source: European Institute of Innovation & Technology).

Par exemple, l'European Institute of Innovation and Technology⁴ suggère une liste de 15 technologies *deep tech*, parmi lesquels l'informatique quantique, la biotechnologie, la cybersécurité, l'électronique et la photonique, la robotique, l'énergie, etc⁵. Ces technologies *deep tech* peuvent ensuite servir à résoudre différents défis mondiaux, tels que le climat, l'énergie, la santé et la mobilité, formant ainsi une matrice *deep tech* (voir Figure 1 ci-dessus). Cette description de la *deep tech* permet de catégoriser des entités telles que des entreprises, des établissements ou des start-ups qui s'engagent à offrir des innovations technologiques de pointe pour relever les enjeux majeurs de notre époque.

Cette manière d'appréhender le concept de *deep tech* rappelle les politiques industrielles de filières et de *clusters*⁶, dont l'objectif est de soutenir et d'encourager l'innovation ainsi que la création de valeur économique. Ces politiques favorisent la concentration géographique d'entreprises, d'institutions académiques et de start-ups dans des secteurs technologiques spécifiques (biotechnologie, quantique, cybersécurité, IA, *clean tech*, etc.).

La *deep tech* en tant que projet d'innovation

Il est également possible de définir la *deep tech* en précisant les caractéristiques essentielles qu'un projet, une start-up ou une stratégie doit posséder pour être

qualifié(e) comme tel. Cela inclut l'innovation de rupture, la capacité à résoudre des problèmes fondamentaux, une forte intensité en R&D, la valorisation de découvertes scientifiques, une coopération étroite avec les écosystèmes de recherche ainsi que la protection des technologies par la propriété intellectuelle. Par exemple, pour Bpifrance, le qualificatif *deep tech* s'applique aux projets qui « mobilisent les dernières avancées de la recherche scientifique pour répondre aux enjeux de demain. Leur mission : proposer les produits qui transformeront durablement nos modes de vie. »

Dans cette perspective, la *deep tech* ne se limite pas à une simple avancée technologique, mais désigne un modèle spécifique d'innovation, ancré dans une approche scientifique et technique de long terme. Elle englobe à la fois le développement de technologies de pointe et leur commercialisation⁷, en transformant des découvertes scientifiques en solutions concrètes à fort impact économique et sociétal.

Cette approche insiste sur le rôle central et le degré de complexité de la technologie dans le projet d'innovation. L'enjeu est de distinguer le concept de *deep tech* des termes usuels de *low-tech*, *high-tech* et *shallow tech*. Dans les projets d'innovation *low-tech*, l'enjeu consiste à intégrer des technologies simples et éprouvées pour la conception de solutions durables, accessibles au plus grand nombre, écologiques et peu coûteuses⁸. À l'inverse, les projets d'innovation *high-tech* reposent sur une amélioration continue et incrémentale de technologies avancées (électronique, numérique, informatique), visant à accroître la performance de produits grand public et à encourager

⁴Agence européenne indépendante qui agit pour l'innovation et la technologie, et fait partie du programme-cadre de l'UE pour la recherche et l'innovation.

⁵EUROPEAN INSTITUTE OF INNOVATION & TECHNOLOGY (2023), EIT Deep Tech definitions, EIT Deep Tech talents for Europe Initiative (DTTI), <https://www.eitdeeptechtalent.eu/wp-content/uploads/gb/2023/02/deeptech-definitions.pdf>

⁶Kerr, W. R., & Robert-Nicoud, F. (2020). Tech clusters. *Journal of Economic Perspectives*, 34(3), 50-76.

⁷Raff, S., Murray, F. E., & Murmann, M. (2024). Why You Should Tap Innovation at Deep-Tech Startups. MIT Sloan Management Review.

⁸GRIMAUD E., TASTEVIN Y. P. & VIDAL D. (2017), « Low tech, high tech, wild tech. Réinventer la technologie ? », *Techniques & Culture. Revue semestrielle d'anthropologie des techniques*, (67), pp. 12-29.

leur renouvellement rapide (automobiles, smartphones, ordinateurs, etc.). Les innovations *shallow tech* s'appuient sur des technologies numériques standardisées et facilement accessibles (applications mobiles, logiciels, applications *web*), et leur avantage compétitif découle de modèles économiques innovants, tels que les plateformes, le *software-as-a-service* ou les *marketplaces*⁹. Enfin, les projets d'innovation *deep tech* se fondent sur la valorisation économique de découvertes scientifiques et de technologies de rupture, dans le but de résoudre des problèmes sociétaux et environnementaux complexes.

La *deep tech* est-elle un enjeu stratégique ?

Au-delà des enjeux de définition, une question plus fondamentale se pose : la *deep tech* constitue-t-elle un enjeu stratégique majeur ou une simple tendance éphémère ? Est-elle véritablement essentielle ou se résume-t-elle à un simple effet de mode ? Selon nous, l'engouement pour l'innovation *deeptech* et la popularité croissante du concept mettent en lumière quatre enjeux contemporains cruciaux pour nos sociétés.

Assurer la souveraineté technologique et industrielle

La souveraineté technologique et industrielle désigne la capacité d'un État à garantir l'accès et le développement des technologies jugées essentielles pour son bien-être, sa compétitivité et son autonomie stratégique, tout en limitant les dépendances structurelles envers des puissances étrangères¹⁰. Cet enjeu prend une résonance particulière dans le contexte actuel de tensions économiques et géopolitiques croissantes entre la Chine, l'Union européenne et les États-Unis. La rivalité technologique entre ces acteurs majeurs met en lumière l'importance stratégique de maîtriser les technologies critiques, qu'il s'agisse de l'intelligence artificielle, des semi-conducteurs ou des énergies renouvelables. Ces dynamiques accentuent l'urgence, pour les États, de renforcer leur souveraineté technologique afin de préserver leur autonomie, leur compétitivité et leur capacité d'action sur la scène internationale.

Dans ce contexte, les *deep tech*, en tant que secteurs technologiques spécifiques, apparaissent dès lors jouer un rôle déterminant dans le renforcement de la souveraineté technologique et industrielle des nations. Elles contribuent à réduire la dépendance vis-à-vis de technologies étrangères en favorisant le développement local de solutions innovantes dans les secteurs stratégiques.

⁹TEKIC Z., ABUELEZ A., & TEKIC A. (2023), "Technological synergies as antecedents of sustainable development: Deep-tech versus shallow-tech perspective", *Technology Analysis & Strategic Management*, pp. 1-15.

¹⁰EDLER J., BLIND, K., KROLL H., & SCHUBERT T. (2023), "Technology sovereignty as an emerging frame for innovation policy. Defining rationales, ends and means", *Research Policy*, 52(6), 104765.

Utiliser le changement technologique pour soutenir la croissance économique

Les technologies issues de la *deep tech*, porteuses de ruptures majeures et de progrès technique, possèdent un potentiel unique pour transformer en profondeur les industries. Elles constituent un levier stratégique pour revitaliser l'industrie nationale, moderniser les processus, créer de nouvelles chaînes de valeur et encourager l'émergence de start-ups innovantes. En favorisant des gains de productivité et l'ouverture de nouveaux marchés, elles soutiennent une croissance économique durable et renforcent la compétitivité internationale des nations dans un environnement global de plus en plus exigeant.

Ces liens entre progrès technologique, innovation et développement économique sont au cœur des économies occidentales fondées sur l'innovation. Pour les théoriciens de la croissance, de Marshall¹¹ à Aghion et Howitt¹², en passant par Schumpeter¹³, le changement technologique est le moteur principal de la croissance économique. Ce processus repose sur la dynamique de destruction créatrice : les nouvelles technologies remplacent les anciennes, entraînant la disparition d'industries établies, mais permettant l'émergence de secteurs innovants. Cette capacité à renouveler les bases de l'économie est essentielle pour maintenir la compétitivité et stimuler la croissance à long terme.

Les promesses d'innovation de rupture offertes par les technologies de la *deep tech* constituent ainsi un levier stratégique pour faire émerger de nouveaux produits, procédés, marchés et organisations.

Renforcer les collaborations université / industrie / société

Dans une économie de l'innovation fondée sur la production de connaissances, les collaborations entre universités / industries / société jouent un rôle central. Pour les entreprises, elles offrent un accès privilégié à des avancées scientifiques, favorisant ainsi la conception de produits innovants. Pour les universités, elles élargissent leur rôle au-delà de l'enseignement et de la recherche traditionnels, incarnant ce que l'on appelle la « troisième mission de l'université » ou « l'université entreprenante »¹⁴. Cette mission implique une participation accrue à l'innovation et à la croissance économique *via* la production de savoirs exploitables par les entreprises. Enfin, pour la société, ces partenariats contribuent à la croissance économique, renforcent la prospérité et peuvent participer à la réduction des inégalités¹⁵.

¹¹MARSHALL A. (1890), *Principles of economics*, Springer.

¹²AGHION P. (1990), *A model of growth through creative destruction*.

¹³SCHUMPETER J. A. (1942), *Capitalism, socialism and democracy*, Routledge.

¹⁴GUERRERO M., & URBANO D. (2012), "The development of an entrepreneurial university", *The journal of technology transfer*, 37, pp. 43-74.

¹⁵AGHION P., ANTONIN C. & BUNEL S. (2021), *The power of creative destruction: Economic upheaval and the wealth of nations*, Harvard University Press.

Élaborer des politiques publiques d'innovation adaptées aux grands défis

Si le progrès technique et la compétition mondiale par l'innovation intensive ont engendré un enrichissement sans précédent ainsi qu'une prospérité accrue pour les nations du monde occidental, ils ont également suscité de nouveaux défis économiques, sociaux et environnementaux¹⁶. En effet, la surconsommation alimentée par cette économie de l'innovation a notamment contribué à générer une pression sans précédent sur les ressources naturelles, entraînant une dégradation des écosystèmes naturels, une augmentation des émissions de gaz à effet de serre et un dépassement de plusieurs limites planétaires.

Dans un contexte marqué par ces grands bouleversements, la notion de grands défis ("*grand challenges*"¹⁷) s'est imposée comme un élément central des politiques publiques d'innovation¹⁸. Un grand défi désigne un problème complexe, d'envergure mondiale, dont les implications sont critiques pour la société. Il s'agit par exemple du changement climatique, de la transition énergétique, de la santé, de la préservation de la biodiversité, du vieillissement de la population, de la sécurité alimentaire ou encore de l'agriculture soutenable.

Pour relever ces grands défis, il est nécessaire de mobiliser et de coordonner les efforts conjoints de la recherche, de l'industrie, des gouvernements et de la société civile. Pour les États, l'enjeu réside dans l'organisation d'une mobilisation collective et la promotion de la coopération entre les différents acteurs. Cela passe notamment par l'orientation des financements, la structuration des priorités de recherche et la stimulation de l'innovation, afin de générer un impact concret et durable sur ces problématiques majeures.

La *deep tech* : symptôme d'un monde en transition

Plus qu'un nouveau paradigme d'innovation ou une énième révolution industrielle, le concept de *deep tech* s'impose avant tout comme le symptôme d'une quête intensifiée de solutions scientifiques et technologiques face à des défis sociétaux, économiques et environnementaux d'une complexité sans précédent. Il reflète la confiance placée dans la recherche scientifique et dans le rôle central du progrès technologique pour répondre à ces problématiques majeures, qu'il s'agisse du changement climatique, de la transition énergétique ou de la santé.

Cependant, cette approche repose sur une vision qui mise principalement sur les capacités de la science

et de la technologie à résoudre ces enjeux, parfois au détriment d'une analyse plus large intégrant les dimensions sociales, culturelles et politiques des problématiques. Dans ce cadre, la *deep tech* incarne à la fois un immense potentiel de transformation et un défi : garantir que ces innovations s'intègrent dans une vision systémique et durable, capable de tenir compte des limites planétaires, des inégalités sociales et des risques d'effets rebonds. En cela, la *deep tech* illustre un moment charnière dans l'histoire de l'innovation, où la technologie doit être pensée non pas comme une fin en soi, mais comme un outil au service d'un futur plus juste et résilient.

¹⁶ FORAY D. (2024), « Une économie de l'innovation pour les temps à venir », *Innovations*, 74(2), pp. 161-181, <https://doi.org/acces.bibl.ulaval.ca/10.3917/inno.pr2.0164>

¹⁷ GEORGE G., HOWARD-GRENVILLE J., JOSHI A. & TIHANYI L. (2016), "Understanding and tackling societal grand challenges through management research", *Academy of management journal*, 59(6), pp. 1880-1895.

¹⁸ MAZZUCATO M. (2017), Mission-oriented innovation policy: Challenges and opportunities, UCL Institute for Innovation and Public Purpose Working Paper, (2017-1).

Artificial intelligence in everyday life

Introduction

Serge Catoire & Arnaud Mazier.

General framework

Generative AI: A technological revolution between promises and challenges

Éric Moulines.

Generative artificial intelligence (generative AI) refers to models capable of producing original content – such as text, images, or sounds – from training data. It represents a significant shift from traditional AI due to its versatility and ability to perform various tasks without specific prior training. Recently popularized by models like GPT-4 and DALL-E, generative AI is already transforming diverse fields, including mathematical reasoning, engineering, and creative industries. However, this technology has been raising crucial challenges concerning ethics, bias, environmental impact, and reliability. Powerful yet imperfect, generative AI requires thoughtful use, closely integrating human expertise and best practices to leverage its innovation potential while minimizing associated risks entirely.

CEA's artificial intelligence program and its applications

François Terrier.

As the new electricity of the future, AI opens up a vast field of applications. Their deployment, however, raises questions of trust and frugality that could block its industrial diffusion. Europe has responded by introducing a risk-based regulatory framework designed to ensure the smooth development of applications. This raises particularly complex qualification issues, which is the subject of one of CEA's AI action programs. At the same time, deployment in industrial products and systems raises the question of embeddability and resource requirements, which could be incompatible with the intended uses and cost of this type of solution. This is why a second CEA program is focused on embedded AI solutions. At the same time, AI also offers strong potential for use in scientific activities themselves, leading to its integration into CEA's cutting-edge research. These different facets of the challenges of AI are illustrated by some use cases well representative of the issues at stake and the solutions on offer.

The stakes of responsible AI

Brigitte D'Andréa-Novel & David Ryckelynck & Georges-André Silber.

In 2022, before the explosion of generative AI, digital technology already accounted for 4.4% of France's

annual carbon footprint and 11% of its electricity consumption, *i.e.* 51.5 TWh, or 65 TWh if we take into account the electricity consumption of data centers located abroad. At the same time, the development of renewable energies, smart grids, smart cities, and digital twins means that we need to make ever greater use of AI, and especially machine learning. How can this duality be resolved?

Researchers at the Institut des Transformations Numériques of Mines Paris have been exploring and developing innovative practices for a digital world that is more respectful of the environment and resources, by means of different levers, namely: a reasoned use of data, the relevance of networks based on *a priori* knowledge, such as PINNs, and taking into account hardware/software constraints linked to implementation and the reproducibility of results.

China's AI strategy

Paul Jolie.

The USA has long been at the forefront of AI, thanks to cutting-edge research and a strong private sector, but China's relentless, state-driven approach is rapidly closing the gap.

In 2023, there were 4,482 AI companies in China, according to a census by the Chinese Academy of Information and Communication Technology.

By March 2024, China had 52 AI unicorn companies, including 13 new ones.

The 2023 Global AI Innovation Index report shows that China ranks first in the world, accounting for 36.7% of leading AI papers and 34.7% of AI patents granted among leading countries, demonstrating strong momentum in AI innovation.

One of the most striking aspects of China's AI strategy is its open source LLM ecosystem, which is progressing at a pace that has attracted the attention of global AI communities.

IA in administration, training, and law

Making a success of Post-Paris 2025: What kind of ambitions for France following the AI Action Summit?

Arno Amabile.

Between hope and fatalism, France is seeking its place in the artificial intelligence revolution. To become an "AI power", it must develop both its power over AI (technological mastery and adoption by society) and its power through AI (weight in the value chain and geopolitics of AI).

If France is well placed to be a fast follower in the development of general AI, it must still find the sectors and markets in which its ecosystem can benefit from its know-how or ecosystem, without finding itself locked into small isolated markets.

The report of the AI Commission in March 2024, the third phase of the AI strategy in February 2025 and the Summit for Action on AI in the same month have all contributed to this ambition, with no illusions about our dependencies and our lag, but without renouncing the potential of AI for France.

Generative AI in the French administration: the case of “Albert”

Ulrich Tan.

In this article, we explore the “Albert” project, the generative artificial intelligence developed by the French Interministry Digital Department (DINUM). The underlying strategy stems from an ambitious technological vision for the public sector: to take ownership of a new and promising technology, rather than being subjected to it, and to ensure the State’s digital sovereignty. Launched in response to the meteoric rise of AI, Albert aims to provide solutions tailored to the needs of the French administration, while guaranteeing data confidentiality and the necessary computing power. We will discuss the three pillars of the project, its concrete applications, and the technological and ethical challenges to be met for optimal use of AI in the public sector.

How is the French education system embracing generative artificial intelligence?

Émilie-Pauline Gallié & Erwan Paitel.

Since 2022, generative AI (GAI) has been accessible to a large proportion of the world’s population thanks to the Internet, but its use raises questions about the use and command of the tools, biases, and ethical and environmental limits. In France, the education system has encouraged experimentation with this technology. It is now necessary to develop robust strategies for integrating the use of these technologies into students’ learning curricula.

To move from individual initiatives to AI education for all, a systematic and inclusive approach is needed, guaranteeing access to the tools and training for all students. In particular, it is crucial to train teachers in the use of AI and to develop specialized skills. Finally, technical, pedagogical, and environmental challenges need to be addressed to ensure the successful integration of AI into education. This includes setting up adequate infrastructures and promoting sustainable practices to minimize the environmental impact of AI.

Law and artificial intelligence: Interactions and transformations

Yannick Meneceur.

Applications of artificial intelligence (AI) in the fields of law and justice have long since left the laboratory and

become trivial. In this context, they can be analysed from two angles: that of their legal framework, as objects of law, and that of their influence on the law, as a veritable new player.

The aim of this article is to outline the interactions and transformations underway, by describing the new European legal frameworks and the societal issues at stake with the deployment of AI in judicial systems, going beyond the classic balance between benefits and risks.

The lawyer’s office in the age of digital technology and artificial intelligence: Interview with attorney Christiane Féral-Schuhl Interview by Arnaud Mazier.

Although digital tools improve the material day-to-day life of lawyers and facilitate their relationship with clients, a lawyer cannot be content to be an ordinary user of artificial intelligence. They must scrupulously respect the essential principles governing the legal profession, as well as the fundamental principles of justice. While these principles are deeply rooted in the professional conscience of lawyers, they can be eroded by the digital environment, sometimes without the lawyer realizing it.

Artificial intelligence, particularly generative intelligence, undeniably offers new possibilities in the practice of law. However, its integration must respect continental law, while maintaining the lawyer’s full responsibility. The challenges of training are crucial to support the evolution of the profession.

Using AI to improve the quality and responsiveness of the DGFIP’s work: Moving forward and innovating within budget constraints

Esther Mac Namara & Thomas Binder.

The General Directorate of Public Finances (DGFIP for the French acronym) integrates artificial intelligence (AI) into various processes, such as optimizing tax management, combating fraud, and improving accounting operations. Since 2016, it has been using data mining to detect fraud, and has been deploying solutions like the « Foncier Innovant » project to identify undeclared constructions. While promising, generative AI (GAI) still faces technical and organizational challenges.

In 2024, the DGFIP developed its first strategic roadmap to securely integrate GAI into its processes. For 2025, it plans to establish strategic governance and training programs to support its deployment while ensuring a responsible and efficient use of resources.

Today’s business applications

AI for precision medicine in oncology

Thomas Walter.

How do you choose the best possible treatment for a cancer patient? This is the challenge of precision

medicine, which aims to offer personalized treatment to each patient. This choice must be based on data characterizing the tumor and its microenvironment.

Artificial intelligence provides us with a technique capable of creating predictive models from massive, heterogeneous data, and thus predicting the optimal treatment for each patient.

Here, we present the dazzling progress of artificial intelligence through an example of its application to digital pathology, an essential technique for oncology diagnosis. The emergence of advanced concepts, such as multiple instance learning, self-supervision, and foundation models, has contributed to an impressive number of new methods and applications, unthinkable just a few years ago. In breast and ovarian cancer, for example, AI has made it possible to predict homologous recombination defects, a defect in DNA repair, from simple histological slides. These advances pave the way for more personalized treatments, making AI an essential component of tomorrow's medicine.

AI applications for radiology assistance

Jérôme Knoplioch & Nicolas Gogin.

Medical Imaging has been at the forefront of Artificial Intelligence advances. Diagnostic Imaging products account for more than 75% of all AI-enabled radiology devices that the FDA certified so far. AI applications include better image reconstruction, assistance for detecting or quantifying lesions – for overloaded radiology staff –, and even extend to communication with patients.

With massive investment in general AI, developing new high performance clinical features has become easier and faster. Training models remains a costly endeavor, due to the difficulty of data collection and high labor costs for experts who provide ground truth. To reap the full productivity benefit of AI, users will need better workflow integration and consistently reliable performance. Fusing imaging data with full medical history records will open the way to new biomarkers, enhancing the predictive value of imaging and its impact on treatment.

Using AI to better benefit from a technical database for the mechanical engineering industry

Fanny Lambert.

To improve the use of technical documents for the benefit of mechanical engineering manufacturers, Cetim has developed an in-house chatbot capable of using thousands of documents produced in-house in a secure and controlled manner. This initiative aims to overcome the limitations of chatbots available on the web, particularly in terms of confidentiality and processing capacity.

The project, which has now been completed, uses Cetim's monitoring documents to generate summaries, syntheses, and comparisons, thus facilitating the rapid collection and analysis of scientific and technical

information. This project has also enabled us to understand this technology, to evaluate the costs, the IT infrastructures required, and the skills needed to integrate an AI in-house, enabling us to share it with you today.

The potential of artificial intelligence for an industrial technology company

François Marion.

The use of Artificial Intelligence (AI) is a powerful driver of transformation for an industrial technology company, such as Valeo. The introduction of AI 13 years ago, onboard vehicles, in driving assistance systems, allowed the group to acquire leading skills in AI. Valeo is now deploying this technology within the company itself to transform its functions; whether in R&D, through the optimization of software development via coding assistance or the multiplication of tests and validations, or in production with, for example, automatic tests or the optimization of energy consumption. Several examples show that it is possible to introduce AI by applying a pragmatic approach of rapid prototyping, development, and step-by-step deployment. However, given the expected impact, its systematic deployment will require accompanying this change.

The construction sector: When AI changes the game

Marie-Luce Godinot.

The introduction of artificial intelligence (AI) in the construction sector has opened up numerous opportunities. From the bidding phase to maintenance and operation, AI significantly enhances project efficiency and safety. The integration of technologies, such as computer vision and machine learning, transforms traditional processes, enabling new applications and innovations. The arrival of generative AI, though recent, already shows promising prospects.

The adoption of artificial intelligence in construction is constantly progressing, and although technologies are already mature for several applications, the development potential remains immense. Rapid advances in this field suggest smarter, safer, and more efficient construction sites in the future.

GEN AI: The new shock of autonomy

Pierre Dulon.

This article defends a dual conviction:

- GEN AI technology has the potential to bring about a major transformation in the banking business, as it will “decentralize” computing and reasoning capacities, and greatly “empower” employees;
- The GEN deployment project must be approached first and foremost as a human and managerial project, since it is the empowerment of teams that will be the key success factor.

Artificial intelligence in practice to meet the challenges of today and tomorrow

Emmanuel Cox

SNCF Réseau, responsible for the French rail network, must cope with a growing demand for train transportation while modernizing its infrastructure and maintaining financial balance. 50,000 employees manage this extensive network of 28,000 km of lines on a daily basis, requiring a wide range of skills. An in-depth digital transformation is underway to increase efficiency and service quality.

The company has been integrating artificial intelligence (AI) for many years to optimize its operations: predictive maintenance, data analysis, and traffic management. Projects use AI to anticipate weather-related disruptions, and optimize infrastructure management. AI also serves to improve punctuality and responsiveness.

While generative AI is being adopted cautiously for industrial uses, it offers opportunities like enhancing documentation, generating content, and boosting digital efficiency. Questions remain about its impact on employment, sovereignty, and the environment, which are being explored through experiments.

An ethical framework governs the use of AI, and its adoption also involves supporting employees through training and awareness. Strengthening academic and industrial partnerships is envisaged to continue the responsible development of AI within SNCF Réseau.

Trusted AI is a prerequisite for AI deployment in mission-critical systems

Juliette Mattioli & Christophe Meyer.

Thales develops solutions for critical sectors, such as defense, aerospace, and cybersecurity, with a focus on trusted AI. To ensure the validity, transparency, security, reliability, and accountability of AI systems, Thales combines symbolic and connectionist approaches (hybrid AI). The company ensures that its systems are robust against cyberattacks, and transparent to users (even explainable whenever possible).

Thales participates in initiatives, such as ANITI and Con fiance.ai to develop trusted AI engineering methodologies. The emphasis is also on sustainability and ethics, optimizing resources and reducing the carbon footprint through advanced so-called frugal technologies.

Blaxtair: The challenge of intelligent industrial vehicles for safer, more efficient industry

Sabri Bayouhd.

On-board intelligent perception technologies, based on AI, are revolutionizing the industrial vehicle market. The first use case for these technologies is safety, mainly pedestrian detection. Since 2009, with Blaxtair, Arcure has been developing the vehicle's "intelligent eye", capable of unrivalled precision in particularly complex

industrial environments. With Blaxtair® solutions, AI is already saving lives in industry.

Beyond this first use case, deep learning and computer vision are paving the way for the vehicle's global reading of the environment. What's more, this intelligent vehicle is now capable of collecting data across the sites in which it operates, to enable global control of installations and their operation.

Manufacturers are well aware of this, and are now ready to integrate such technologies: The intelligent industrial vehicle is poised to play a central role in Industry 4.0.

Mobilizing artificial intelligence for public transport: An obvious operational advantage, but one that must remain focused on passenger services

Hélène Brisset.

Île-de-France is one of the two largest transportation networks in the world. Île-de-France Mobilités is responsible for organizing, modernizing, and developing this network, which continues to expand to handle more than 10 million daily trips by residents of the region.

Artificial intelligence (AI) is playing an increasingly important role. Apps like Île-de-France Mobilités or Transport Public Paris 2024 provide real-time information, offer automatic translations, and respond directly to user requests. The unified phone number simplifies the process for responding to travelers' phone inquiries. AI is also used to monitor crowd density, enhance safety, and optimize network operations in case of issues.

AI is a key driver for transforming mobility. Its deployment must continue with a constant focus on improving the user experience, not just with a technological goal in mind.

Issue editors: Serge Catoire & Arnaud Mazier.

Miscellany

What is deep tech? Definitions and strategic issues

Benjamin Cabanes.

Today, the term "deep tech" has become a key concept in the world of technology, innovation, and business. It is attracting growing interest from investors and innovation players, who see in it a tremendous opportunity for profound and lasting transformation. This article explores the various definitions of the concept of "deep tech", analyzes its specific features and real implications, and identifies the main strategic issues at stake.

Ont contribué à ce numéro



Arno AMABILE

a accompagné l'envoyée spéciale du président de la République pour le Sommet de l'action sur l'IA, après des expériences dans la transformation numérique et le développement industriel. En 2023 et 2024, il a co-écrit le rapport de la Commission IA remis au président de la République. En tant qu'adjoint du DRH du Corps des mines, qui recrute et

forme des ingénieurs de haut niveau, il a participé au renforcement des compétences numériques de l'État.

De 2021 à 2022, il a animé la *task force* numérique des États généraux de la Justice, qui visaient à réformer le système judiciaire français, avant de piloter le développement de la nouvelle stratégie numérique du ministère. Il a commencé sa carrière comme *data scientist* à la mairie de New York et chez Zalando à Berlin, avant de s'installer à Strasbourg, où il a dirigé l'action de l'État pour le développement industriel régional. Dans ce rôle, il a coordonné le soutien régional aux entreprises pendant la pandémie de Covid-19 ainsi que le déploiement du plan de relance.

Diplômé en économie de l'École normale supérieure de Paris et de l'École d'Économie de Paris, il est titulaire de diplômes de l'Université de New York en informatique urbaine, et de l'École des Mines de Paris.



Sabri BAYOUDH

PhD, est directeur de l'innovation chez Arcure-Blaxtair.

Passionné par l'intelligence artificielle et les nouvelles technologies, il suit un cursus d'ingénieur électrique suivi d'un doctorat en IA à l'Université de Rennes. Après une première expérience au CEA, il rejoint Arcure en 2010 pour en devenir le directeur

technique en 2012. Après une dizaine d'années d'expérience technique, il complète sa formation par un Executive MBA à HEC en 2020. Fort de ces compétences techniques et *business*, il crée la direction innovation dont il prend la tête. Membre du comité de direction et du comité stratégique, il participe au développement et à la croissance d'Arcure par l'innovation technologique pour la création de valeur.



Thomas BINDER

ingénieur de l'École polytechnique, est responsable du département IA de la délégation à la transformation numérique de la DGFIP depuis 2023. Avant de rejoindre les finances publiques, il occupe plusieurs postes de *leadership* technique pour le développement et l'industrialisation de solutions d'intelligence artificielle

dans le secteur privé. D'abord chez IBM puis en tant qu'entrepreneur, il a à cœur les enjeux de robustesse, d'impact et de gouvernance des systèmes d'IA.



Hélène BRISSET

polytechnicienne, ingénieure générale des Mines, agit depuis près de vingt-cinq ans dans la transformation numérique de l'État et des institutions publiques.

Elle a notamment piloté la construction du « Réseau interministériel sécurisé pour les informations classifiées et sensibles » (ISIS) au secrétariat

général de la Défense nationale (SGDN) entre 2005 et 2008. De 2011 à 2015, elle a conçu et piloté la construction du « Réseau interministériel de l'État » (RIE), socle d'échanges commun entre l'ensemble des sites ministériels.

Elle a travaillé au cœur de l'État et de l'Interministériel depuis 2009 :

- Tout d'abord, comme sous-directrice des systèmes d'information sécurisés à l'Agence nationale de la Sécurité des systèmes d'information (Anssi) entre 2009 et 2011.

- Elle a ensuite été directrice adjointe du numérique et du système d'information et de communication de l'État (DINSIC) entre 2015 et 2016, puis directrice de cabinet du secrétaire d'État chargé du numérique, Mounir Mahjoubi, en 2017.

- Jusqu'en 2022, elle a porté la transformation numérique des ministères, en tant que directrice des systèmes d'information puis du numérique des ministères sociaux (ministères du Travail, de la Santé, de la Solidarité et des Sports) avec tout particulièrement la conception du Code du Travail numérique et la mise en place de la fabrique numérique des ministères.

Depuis mai 2022, Hélène Brisset est directrice du numérique d'Île-de-France Mobilités, où elle a notamment piloté la mise en place de l'application Transport Public Paris 2024 et l'activité numérique liée aux Jeux olympiques et paralympiques.



Benjamin CABANES

est enseignant-chercheur en sciences de gestion à Mines Paris - PSL. Ses recherches portent sur le management de l'innovation, le management stratégique de la R&D, l'éco-conception, l'entrepreneuriat deeptech et les collaborations université-industrie. Il est également chercheur associé à l'Institut Interdisciplinaire de l'Innovation

(UMR CNRS 9217, Mines Paris - PSL, École polytechnique, Télécom Paris). En 2022, il cofonde le MS Entrepreneuriat Deeptech et Innovation de Mines Paris - PSL, et 2024, la mineure Entrepreneuriat Deeptech de l'Université PSL. Il enseigne l'entrepreneuriat deeptech et l'innovation à Mines Paris, et l'éco-conception pour l'innovation soutenable à l'École polytechnique. Il intervient également en formation continue à HEC Paris, l'École polytechnique et Paris Dauphine - PSL.



Serge CATOIRE

est ingénieur général des Mines honoraire. Il travaille depuis ses débuts professionnels sur des sujets industriels et internationaux : tout d'abord, au sein de l'administration, en particulier en tant que conseiller technique au sein du cabinet de Hubert Curien, ministre de la Recherche et de la Technologie, puis dans l'industrie

aéronautique et spatiale. Entre autres fonctions, il a été directeur technique et industriel du groupe Aérospatiale, directeur général d'Aérospatiale missiles, puis PDG de Reims Aerospace. En 2012, il a rejoint le ministère chargé de l'Économie et des Finances pour y exercer au sein du Conseil général de l'Économie, dont il a présidé le comité de l'inspection. Depuis juin 2023, il est directeur industrie à l'UIMM.



Emmanuel COX

commence sa carrière au sein du groupe SNCF en 1991. Diplômé de Neoma Business School, il débute en tant que responsable marketing pour Fret SNCF, devenu Hexafret, puis évolue vers la branche TER, où il lance les premiers programmes de relation client (CRM).

En 2004, il a la charge d'audits de filiales et de grands projets, notamment dans le domaine des systèmes d'information. Cette expérience lui permet de développer une vision stratégique globale au niveau du groupe.

En 2008, il dirige un programme complexe de rénovation du système pour la maintenance des locomotives et voitures, touchant plus de 10 000 utilisateurs.

Après un passage humainement très riche à la fondation SNCF, dès 2014 Emmanuel Cox est chargé de déployer un programme d'innovation numérique, autour des objets et capteurs connectés (IOT), puis son engagement envers le développement technologique l'amène à être nommé *digital transformation officer* en 2017 chez SNCF Réseau. Il contribue alors avec les équipes de la direction numérique à la digitalisation de l'entreprise autour de la *data*, de programmes d'accompagnement au numérique et de la mise en place d'un accélérateur de transformations.

Fervent promoteur des approches collaboratives, il est membre-fondateur et vice-président de l'Innovation Makers Alliance (+ 100 groupes français), est à l'origine de la création du Rail Open Lab, consacré à l'*open innovation* dans le secteur ferroviaire.

En 2024, il structure le Lab IA pour animer le déploiement de l'IA générative au sein de SNCF Réseau. Convaincu de la nécessité d'imaginer le futur du numérique au-delà des intérêts propres de chaque entreprise, il œuvre au service de la filière ferroviaire pour créer de nouvelles plateformes de services numériques partagés. La prochaine d'entre elle pourrait être UbiRail, destinée à mutualiser des contenus pédagogiques immersifs pour la formation et l'attractivité des métiers.



Brigitte d'ANDREA-NOVEL

est professeure à l'École des Mines de Paris depuis 1992. Titulaire d'un diplôme d'ingénieur de l'École Supérieure d'Informatique Électronique Automatique (ESIEA) en 1984, d'un doctorat en mathématiques et automatique de l'École des Mines de Paris en 1987, elle

soutient son habilitation à diriger des recherches (HDR) à l'Université Paris-Sud en 1995.

Responsable d'un groupe de recherche sur les « systèmes de contrôle avancés » au Centre de Robotique de Mines Paris, ses domaines de recherche concernent la théorie du contrôle non linéaire orienté vers les systèmes mécaniques, avec des applications sur la voiture autonome, les systèmes mécaniques flexibles, les canaux d'irrigation et les instruments de musique.

Dans ce contexte, elle a développé des recherches sur la modélisation et le contrôle actif des instruments de musique, ce qui l'a amenée à diriger le laboratoire « Sciences et Technologies de la Musique et du Son » à l'Ircam de 2018 à 2023. De retour à Mines Paris depuis septembre 2023, elle pilote le nouvel Institut des Transformations Numériques, lequel, aux côtés de l'ingénierie numérique et de la santé, développe un axe autour des industries culturelles et créatives, prenant tout son sens au sein de l'Université Paris Sciences & Lettres, dont Mines Paris est un des membres fondateurs.

Elle développe actuellement des recherches autour des technologies de l'IA générative appliquées au domaine artistique.

Elle est auteur et co-auteur de plusieurs ouvrages, dont le livre *Acoustique, Informatique, Musique* publié aux Presses des Mines en 2012, d'une cinquantaine d'articles de revues et d'une centaine d'articles de conférences internationales.



Pierre DULON

est diplômé de l'École polytechnique (X85) et de Télécom Paris.

Il rejoint France Telecom en 1990, participe à la création de Global One, puis prend la responsabilité des secteurs de la recherche et de l'industrie pétrolière au sein de la direction commerciale des grands

comptes.

En 1998 il rejoint le Crédit Lyonnais, où il occupe des fonctions informatiques et digitales. Il est nommé directeur des Systèmes d'Information de Crédit Agricole Corporate and Investment Bank, puis devient chief operating officer des activités de gestion de fortune. Il crée, en partenariat avec Cap Gemini, la société suisse Azqore, fournisseurs de services informatiques et opérations pour les banques privées, qu'il dirige jusqu'en 2021. Il est alors nommé directeur général adjoint de CACIB, en charge de l'Informatique et des Opérations bancaires.



© Benoit Pitre

Christiane FÉRAL-SCHUHL

est avocate au Barreau de Paris et co-fondatrice du cabinet Féral, spécialisé dans le droit du numérique, des données personnelles et de la propriété intellectuelle. Elle est également médiatrice. En parallèle de son activité professionnelle, elle s'implique dans le domaine institutionnel en étant seconde

vice-présidente du Conseil national de la médiation, après avoir présidé le Conseil national des Barreaux et le Barreau de Paris. Elle a aussi co-présidé la Commission parlementaire sur le droit et les libertés à l'âge du numérique avec le député Christian Paul, et siégé au Haut Conseil à l'égalité entre les femmes et les hommes ainsi qu'au Conseil supérieur des tribunaux administratifs et des cours d'appel administratives. Investie dans la formation, elle a été marraine de promotion d'élèves avocats.

Maître Féral-Schuhl est l'auteur de plusieurs ouvrages et articles, notamment *Cyberdroit : le droit à l'épreuve de l'Internet* (Dalloz, 2020) et *La protection des données personnelles* (Dalloz, 2019). Elle a également co-écrit *La médiation, à l'épreuve de la pratique* avec Michèle Jaudel (Dalloz, édition 2025/2026), ainsi que *Cybersécurité, mode d'emploi* avec Xavier Leonetti (PUF, 2022). Plus récemment, elle a collaboré avec l'illustratrice Tiphaine Mary sur une bande dessinée intitulée *Adélaïde. Lorsque l'Intelligence Artificielle casse les codes* (Dalloz, 2024).



D.R.

Émilie-Pauline GALLIÉ

est inspectrice générale de l'éducation du sport et de la recherche spécialisée dans l'évaluation des politiques publiques en matière d'enseignement supérieur et de recherche. Elle a notamment développé une expertise dans l'organisation de la recherche, le soutien à l'innovation et l'accompagnement des établissements d'enseignement

supérieur. Elle s'intéresse au développement et aux impacts de l'intelligence artificielle, en particulier dans son champ d'action. Elle a été rapporteure de la Commission d'experts sur l'intelligence artificielle (octobre 2023-avril 2024) et est membre de la mission d'appui de la mission IA et Enseignement supérieur.



© Julien Cresp

Marie-Luce GODINOT

diplômée de l'École polytechnique (1994) et de l'École nationale supérieure des Télécommunications (1996), commence sa carrière au ministère de l'Intérieur. En 2001, elle rejoint Bouygues Construction. Après avoir occupé différentes fonctions à la DSI de Bouygues Construction, elle devient en 2015 directrice de l'Innovation et du Développement

durable puis, en 2018, directrice générale adjointe en charge de la Transformation numérique, de l'Innovation et du Développement durable de Bouygues Construction. En 2019, elle étend son périmètre de responsabilités aux systèmes d'information.

Marie-Luce Godinot est nommée en septembre 2022 directrice générale adjointe du groupe Bouygues en charge de l'innovation, du développement durable et des systèmes d'information. Elle rejoint le comité de direction générale du groupe en 2024.

Par ailleurs, elle est administratrice de CentraleSupélec et membre du conseil d'école de Télécom Paris. Elle est également membre de l'Académie des Technologies.



D.R.

Nicolas GOGIN

ingénieur X-Télécom, a fait toute sa carrière dans le domaine de l'imagerie médicale travaillant chez Philips en imagerie interventionnelle, puis chez GE Healthcare depuis 2012 dans le département visualisation avancée à Buc (78). Il a conçu et développé de nombreux algorithmes de traitement de l'image utilisés pour l'imagerie interventionnelle ainsi que pour l'imagerie

de diagnostic par scanner et IRM. Plus récemment il a mis au point le premier algorithme basé sur l'apprentissage profond à être intégré dans un produit GE Healthcare.

Il est actuellement directeur en charge des développements en intelligence artificielle pour la visualisation avancée chez GE Healthcare.

PAUL JOLIE

est ingénieur général des Mines, diplômé de l'École polytechnique et de l'École nationale supérieure des Télécommunications. Il est également titulaire du MBA Edhec.

Après une carrière passée au centre de recherche de France Télécom et dans différents services opérationnels de l'opérateur, il rejoint l'administration en 2008 pour occuper un poste de DSI adjoint au ministère des Affaires étrangères.

Après un passage à l'Inria, il occupe le poste de sous-directeur de l'informatique centrale au ministère de l'Économie et des Finances. Après deux ans en tant que conseiller pour le numérique auprès du ministre d'État de Monaco, il revient à Bercy pour occuper un poste de conseiller au SISSE. Il rejoint le Conseil général de l'Économie en 2020, en tant que référent Intelligence artificielle.



D.R.

Jerome KNOPLIOCH

ingénieur X-Télécom, auteur de plus de 35 brevets, travaille dans le domaine de l'imagerie médicale 3D depuis 1987. Il a conçu et développé de nombreux logiciels innovants de visualisation et de traitement d'images, utilisés de manière quotidienne dans le monde entier sur plus de 30 000 systèmes. Parmi ces logiciels, on peut citer le premier logiciel

clinique d'endoscopie virtuelle interactif, la première solution automatique d'analyse vasculaire volumique, ainsi que le premier logiciel automatisant l'analyse cardiaque scanner.

Jérôme Knoploch est actuellement responsable des architectures pour la visualisation avancée à GE Healthcare, dont le centre d'excellence est basé à Buc (78). Parmi les projets actuels, de nombreux développements font appel à l'intelligence artificielle et à l'apprentissage automatique.



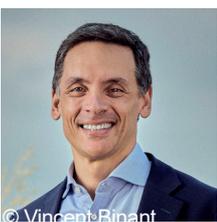
Fanny LAMBERT

à la suite de son doctorat en chimie théorique (simulation des propriétés des matériaux d'électrodes pour batteries) réalisé avec l'ADEME, a rejoint le Cetim en tant qu'ingénieure en veille technologique et stratégique. Sa principale mission est de rechercher et de synthétiser des informations techniques et scientifiques sur les technologies actuelles et futures à travers la littérature scientifique, les brevets, les salons / congrès et des interviews d'experts. Elle met ensuite ces synthèses à disposition des industriels de la mécanique, pour répondre à leurs questionnements, les aider à se positionner et à rester à jour. Sa thématique favorite : les IA génératives ! Elle réalise de nombreux tests sur de multiples modèles pour partager son expérience, elle forme des collègues en interne aux bonnes pratiques, et organise régulièrement des webinaires sur le sujet pour sensibiliser les industriels et leur présenter les cas d'usage pertinents ainsi que l'actualité foisonnante de l'IA générative.



Esther MAC NAMARA

diplômée de l'Essec, et ancienne élève de l'École nationale d'administration (Ena), a travaillé de 2006 à 2011 comme rapporteure permanente de l'Autorité de la concurrence. De 2014 à 2018, elle rejoint la Cour des comptes comme magistrat. Elle a par la suite occupé les fonctions de vice-présidente d'OpenClassrooms, une edtech spécialisée dans la formation aux nouvelles technologies. En 2023, elle devient sous-directrice en charge de la transformation numérique au sein de la DGFIP. En 2025, elle rejoint la Cour des comptes comme conseillère référendaire.



François MARION

est le directeur de la communication et membre du comité exécutif du groupe Valeo depuis le 1er juillet 2022. Entré chez Valeo en mars 2012 en tant que directeur de la stratégie et des relations extérieures, il a ensuite été nommé directeur de la ligne de produits spéciaux pour les systèmes d'éclairage en janvier 2016. Il a été président de Valeo Chine de janvier 2018 à juin 2022. François Marion avait occupé le poste de conseiller participations de l'État auprès du ministre de l'Économie, des Finances et de l'Industrie de 2010 à 2012. Il était entré au ministère en 2002, y occupant différentes fonctions successives : responsable du suivi de la Banque mondiale et de la Banque africaine de Développement, responsable des opérations de marché à l'Agence France Trésor, chef du bureau « Aide au développement et institutions multilatérales de développement » à la direction générale du Trésor, puis chef du bureau « Opérateurs de transports-secteur ferroviaire » à l'Agence des Participations de l'État.

François Marion est ingénieur des Mines et docteur en physique quantique. Il est ancien élève de l'École normale supérieure et de l'École nationale supérieure des Télécommunications.



Juliette MATTIOLI

avec dix ans dans l'enseignement des mathématiques au niveau secondaire, valide en 1993 un doctorat en mathématiques appliquées à l'intelligence artificielle (IA), abordant l'IA hybride combinant morphologie mathématique et réseaux de neurones. En 1996, elle prend en charge l'activité de recherche en optimisation combinatoire, puis de 2001 à 2016, elle dirige des laboratoires de recherche en IA pour la décision. Considérée comme une référence en IA non seulement chez Thales mais aussi en France, elle est en 2017 l'un des cinq représentants de la France à la conférence des innovateurs du G7. Depuis 2019, elle est présidente du *hub* « Data sciences & IA » du pôle de Systematic. Reconnue pour sa connaissance des enjeux industriels de l'IA, elle promeut l'IA hybride et contribue au développement d'une nouvelle discipline : « L'ingénierie de l'IA », afin d'accélérer le déploiement industriel et responsable de solutions à base d'IA. Enfin, dès 2020, consciente de la pénurie des talents mais aussi de la transformation des métiers, elle a concouru à la mise en place dans Thales d'un cycle de formation continue sur l'IA, allant de l'acculturation à l'expertise.

Arnaud MAZIER

est ingénieur général des Mines. Son parcours de directeur des systèmes d'information et d'informaticien lui a permis d'être au cœur de la transformation numérique de huit entités publiques. À la tête de la direction technique et de l'innovation de la préfecture de police, il a contribué à la préparation des Jeux olympiques et paralympiques de Paris en 2024. Il est rapporteur de l'étude remise en mars 2024 au Premier ministre sur l'IA et l'IA générative.



Yannick MENECEUR

est magistrat, inspecteur de la justice et maître de conférences associé à l'Université de Strasbourg. En tant que magistrat, il a exercé des fonctions pénales, notamment en matière de lutte contre la cybercriminalité et la contrefaçon. Il a aussi dirigé au ministère de la Justice un important projet informatique national. Détaché au Conseil de l'Europe entre 2014 et 2024, il a contribué à l'analyse comparative des systèmes judiciaires européens et a développé une expertise en matière de régulation de la transformation numérique, notamment en ce qui concerne l'intelligence artificielle. Il est l'auteur de *L'intelligence artificielle en procès* (Bruylant, 2020), prix du Cercle Montesquieu en 2021 ; et de *IA génératives et professionnels du droit* (LexisNexis, 2024).



Christophe MEYER

est un senior expert en intelligence artificielle de Thales. Il est le directeur technique du grand laboratoire IA de Thales, cortAix Labs - France, qui regroupe environ 150 ingénieurs et chercheurs en IA et technologies associées.

Il a été nommé chevalier de l'ordre national du Mérite sur recommandation du ministre des Armées.

Les sujets majeurs d'attention sur lesquels il engage ses équipes sont les liens entre la cybersécurité et l'IA, la distinction importante entre autonomie et capacité d'adaptation des systèmes « intelligents », et la collaboration homme / machine. Ces trois sujets majeurs sont déterminants pour l'introduction de technologies d'IA dans les systèmes critiques, particulièrement les systèmes militaires : Comment bénéficier des performances inouïes des systèmes apprenants les plus modernes tout en s'assurant qu'ils sont robustes, résilients, transparents et fiables ? Quel degré de prise de décision peut-on octroyer à un système autonome ? Peut-il continuer à apprendre pendant qu'il est opéré, sous quelles conditions ? Comment les IA peuvent-elles contribuer à rendre les décisions humaines meilleures et plus rapides ?

Par ailleurs, il n'a aucun sens de l'humour.



Éric MOULINES

diplômé de l'École polytechnique (promotion 1984), est ingénieur général du corps des Mines et titulaire d'un doctorat de l'École nationale supérieure des Télécommunications (1990). Actuellement professeur de statistique au Centre de Mathématiques appliquées (CMAP) de l'École polytechnique, il a auparavant enseigné à Télécom

ParisTech, se concentrant sur le traitement du signal et des images.

Ses domaines d'expertise incluent les statistiques computationnelles, les simulations Monte Carlo, l'optimisation stochastique, l'apprentissage automatique probabiliste et l'analyse des séries temporelles. Il est particulièrement reconnu pour ses contributions aux méthodes séquentielles de Monte Carlo et au filtrage non linéaire.

Ses recherches actuelles portent sur les modèles génératifs appliqués aux problèmes inverses, l'approximation stochastique dans l'apprentissage fédéré et collaboratif, ainsi que la quantification de l'incertitude par des approches bayésiennes et la prédiction conforme. Son travail scientifique lie constamment théorie et applications pratiques, notamment dans les télécommunications et le contrôle de systèmes complexes.

Chercheur prolifique, Éric Moulines est auteur de plus de 120 articles dans des revues scientifiques et de plus de 400 communications en conférences internationales. Il a occupé des fonctions éditoriales prestigieuses, notamment comme rédacteur en chef de la revue *Bernoulli* (2013-2016).

Ses contributions lui valent une reconnaissance internationale. Il est membre élu de l'Eurasip, de l'Institut de Statistiques mathématiques (IMS) et de l'Association Asie-Pacifique d'Intelligence Artificielle (AAIA). Parmi ses nombreuses distinctions figurent la médaille d'argent du CNRS (2010), le Prix Orange de l'Académie des Sciences (2011), l'Eurasip Technical Achievement Award (2020) et son élection à l'Académie des Sciences en 2017.



Erwan PAITEL

est inspecteur général de l'éducation, du sport et de la recherche depuis 2022. Toute sa carrière a été consacrée au système éducatif, au plus près des apprenants, puis dans des missions de direction, de coordination, de chef de projet et enfin de pilotage des politiques publiques aux côtés de Frédérique Vidal, ministre de l'Enseignement

supérieur, de la Recherche et de l'Innovation. Rapporteur de la commission sur l'Intelligence Artificielle (octobre 2023-avril 2024), il consacre, depuis, un temps important à analyser cette technologie et ses impacts sur le système éducatif et plus largement sur le service public, ainsi qu'à en débattre avec l'ensemble de ses acteurs.



David RYCKELYNCK

développe de l'apprentissage profond pour l'ingénierie scientifique, dans le domaine de la mécanique des matériaux. Il a réalisé des travaux novateurs sur les méthodes d'hyper-réduction dans le domaine des mathématiques appliquées et de la mécanique numérique. Cette méthode associe l'apprentissage automatique et

la réduction d'ordre de modèle basée sur la projection d'équations dans des espaces latents. Les méthodes d'hyper-réduction peuvent être considérées comme une extension de l'apprentissage automatique non supervisé pour la réduction de dimension au domaine de la réduction d'ordre de modèle pour la solution d'équations aux dérivées partielles. Il a également réalisé des travaux novateurs sur la « réduction de modèle *a priori* ». Il a dirigé plus de 30 thèses de doctorat en mécanique numérique et une équipe de recherche composée de 12 chercheurs permanents à Mines Paris de 2011 à 2015. Depuis 2017, il est responsable d'un nouveau cours sur l'ingénierie numérique des systèmes complexes (science des données pour l'ingénierie scientifique) à Mines Paris - PSL. Dans ce cours, la priorité est donnée aux approches hybrides qui combinent l'apprentissage automatique et les méthodes de modélisation habituelles en ingénierie. Son principal sujet de recherche concerne à présent les jumeaux numériques alimentés par des images *via* l'apprentissage automatique. Il développe également de nouvelles méthodes d'augmentation des données pour les applications d'ingénierie.



Georges-André SILBER

est maître-assistant au Centre de Recherche en Informatique, Mines Paris - PSL.

Titulaire d'un magistère en informatique et modélisation et d'un doctorat en informatique de l'École normale supérieure de Lyon, il est enseignant-chercheur au Centre de Recherche en Informatique (CRI) de l'École des Mines de Paris

depuis 2000, dans le domaine de la compilation, du parallélisme et de l'optimisation de code pour l'énergie. Il est titulaire de la chaire Logiciel Responsable (LoRe) de l'ITN, dont le thème est l'optimisation des logiciels libres fondamentaux pour l'énergie.

Il a également créé et dirigé une entreprise de développement logiciel entre 2009 et 2019 (Luxia, devenue Regmind), spécialisée dans les techniques avancées de traitement automatisé du langage (NLP) par apprentissage automatique, matière qu'il enseigne aux élèves ingénieurs des mines.



Ulrich TAN

est directeur adjoint d'Etalab et chef du DataLab à la direction interministérielle du Numérique (DINUM) depuis juin 2023, où il a notamment initié le programme d'IA générative souveraine de l'État, baptisé Albert.

Il commence sa carrière en 2005 dans le financement de projets d'infrastructures (PPP) avant de rejoindre l'administration financière,

où il occupe divers postes en France et à l'étranger. Il sert notamment comme commissaire contrôleur des assurances, attaché financier à New York durant la crise financière de 2008, adjoint chef de brigade à l'Autorité de Contrôle prudentiel et de Résolution (ACPR), ainsi que conseiller économique et financier et représentant de la Banque de France à Tokyo.

En 2014, il quitte le secteur public pour le privé, travaillant dans la production automobile, l'innovation, la R&D et comme directeur de la stratégie et du développement commercial. Il est également entrepreneur dans le secteur numérique, spécialisé en sciences des données et en R&D sur la sécurité et la protection des données. Avant de rejoindre la DINUM, il est CTO de la filiale de *data sciences* d'un groupe bancaire français, développant des produits d'IA pour les banques. En parallèle de ses activités professionnelles, Ulrich Tan a été enseignant vacataire en actuariat et finances pendant quatorze ans et coproducteur de films diffusés en salles en France et à l'étranger, totalisant plus de 200 000 entrées.

Ancien élève de l'École polytechnique (promotion 2002) et ingénieur en chef des Mines, il est aussi diplômé de Sciences Po en finance et stratégie, ainsi que de l'Université Pierre et Marie Curie en mathématiques.



François TERRIER

directeur de recherche au CEA, et *senior fellow* du CEA pour l'intelligence artificielle, est directeur de programme au List (Institut des systèmes numériques intelligents du CEA). Après un doctorat en intelligence artificielle et des recherches dans le domaine des systèmes experts, il a poursuivi des recherches sur l'ingénierie

des logiciels et des systèmes de confiance. Chef du département ingénierie des logiciels et des systèmes, il a développé ces activités en partenariat avec des industriels. En 2020, il prend la direction du programme stratégique du CEA sur l'intelligence artificielle, avec un focus sur les activités de recherche en IA pour un développement responsable de l'IA pour l'industrie, notamment pour faire face aux nouveaux défis critiques de la frugalité, de la confiance et de l'IA embarquée. Dans ce cadre, il contribue à la mise en place du programme de recherche national en intelligence artificielle, PEPR IA, et le dirige pour le CEA en collaboration avec le CNRS et Inria.



Thomas WALTER

est professeur de Mines Paris et directeur du Centre de Bioinformatique (CBIO), un groupe de recherche à Mines Paris. Ce groupe fait également partie de l'unité jointe entre l'Institut Curie, Mines Paris et l'Inserm : « Oncologie Computationnelle » (U1331), où il occupe le poste de directeur adjoint de l'unité. Ses recherches

portent sur le développement et l'application de méthodes d'apprentissage automatique et de vision par ordinateur, afin de répondre à des questions en biologie et en médecine.

Il a obtenu son diplôme d'ingénieur à l'Université de la Sarre en 1999 et a préparé sa thèse au Centre de Morphologie Mathématique de Mines Paris, portant sur des méthodes de diagnostic assisté par ordinateur en ophtalmologie. Il a ensuite rejoint le groupe Ellenberg à l'European Molecular Biology Laboratory (EMBL) à Heidelberg, où il a développé des méthodes de vision par ordinateur pour analyser le premier criblage génomique à l'échelle du génome entier par vidéomicroscopie, afin d'identifier les gènes impliqués dans la division cellulaire. En 2012, il a intégré le Centre de Biologie Computationnelle de Mines Paris, d'abord en tant que chercheur indépendant, puis en tant que directeur depuis 2018. Ses travaux portent sur la vision par ordinateur appliquée à la biologie et médecine, avec des applications au criblage à haut débit (*high-content screening*) et à la pathologie computationnelle. Depuis 2019, il est titulaire d'une chaire à l'Institut Prairie (PaRis Artificial Intelligence Research InstitutE - Prairie), et, en 2021, il a été nommé professeur de Mines Paris. Il est également très impliqué dans l'enseignement de la vision par ordinateur et de l'apprentissage profond appliqué à l'analyse d'images biologiques, dans divers cours à Mines Paris - PSL et au-delà.

Les Annales des mines, c'est 4 revues

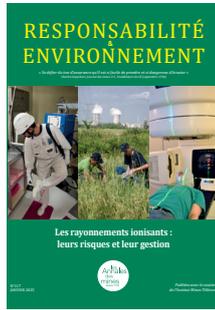


RÉALITÉS INDUSTRIELLES



Publie des dossiers thématiques
sur des enjeux importants
pour le développement industriel et économique

RESPONSABILITÉ & ENVIRONNEMENT



Publie des dossiers thématiques sur l'écologie,
le développement durable, et notamment
l'énergie et le changement climatique

GÉRER & COMPRENDRE



Revue académique pluridisciplinaire
à comité de lecture dédiée à la gestion
et à l'étude des organisations et du travail

ENJEUX NUMÉRIQUES



Publie des dossiers thématiques
sur la transition numérique en croisant les regards
technologiques, économiques et sociétaux

nous assurons une présence sur



Notre site

<https://annaes-des-mines.org/>



LinkedIn

<https://www.linkedin.com/company/page-des-annaes-des-mines/?viewAsMember=true>



Mastodon

<https://mastodon.social/@AnnaesdesMines>



Instagram

<https://www.instagram.com/annaesdesmines/>



X (ex twitter)

<https://twitter.com/AnnaesdesMines>



YouTube



Youtube

<https://www.youtube.com/@lesannaesdesmines7238>