

« ChatGPT, et après ? » Bilan et perspectives de l'intelligence artificielle



Un spectre hante le monde, le spectre de l'intelligence artificielle. Il alimente des angoisses irrationnelles autant que des attentes démesurées. Les technologies d'IA semblent à la fois omniprésentes et insaisissables, pourtant leur diffusion reste encore toute relative et leurs contours sont assez connus des spécialistes. L'IA devient un phénomène de plus en plus politique qui appelle une maîtrise démocratique. Le futur président des États-Unis, Donald Trump, a même déclaré le 28 novembre 2024 qu'il comptait nommer un « tsar de l'intelligence artificielle » qui siégerait à ses côtés à la Maison-Blanche. Aussi, faire le point sur ces technologies et leurs enjeux apparaît aujourd'hui indispensable.

Saisi en juillet 2023 par les bureaux de l'Assemblée nationale et du Sénat pour étudier les nouveaux développements de l'intelligence artificielle dans le contexte de la révolution de l'IA générative, l'Office présente un rapport qui tente, à la fois, de dresser un bilan des technologies d'intelligence artificielle et d'anticiper les tendances qui se dégagent. Ce rapport retrace les développements technologiques des différents modèles d'IA et le détail de leur fonctionnement, identifie leurs enjeux politiques, économiques, sociétaux, culturels et scientifiques, et aborde les questions de régulation, comparant la stratégie nationale française en IA à près de 20 autres, six dans l'Union européenne et onze dans le reste du monde, et en analysant une dizaine de projets de gouvernance mondiale de l'intelligence artificielle. Il propose en conclusion 18 recommandations, dont cinq à soutenir dans le cadre du Sommet pour l'action sur l'IA qui se tiendra à Paris les 10 et 11 février 2025.

Alexandre SABATOU,
député

Patrick CHAIZE,
sénateur

Corinne NARASSIGUIN,
sénatrice

Comprendre les technologies d'intelligence artificielle

I. Une brève histoire de l'intelligence artificielle

Dans le prolongement du rapport de l'Office de mars 2017, ce nouveau rapport revient sur l'histoire et même la préhistoire de l'intelligence artificielle ainsi que sur sa présence forte dans la science-fiction, qui explique pourquoi l'IA suscite à la fois des espoirs excessifs et des craintes pas toujours justifiées.

L'IA n'est pas nouvelle. Elle est liée à la naissance de l'informatique, il s'agit d'ailleurs d'un domaine de l'informatique, une sorte d'informatique avancée dont la frontière fluctue avec le temps. Sa naissance officielle a lieu en 1956 lors d'une école d'été du *Dartmouth College*.

John McCarthy y introduit le concept d'intelligence artificielle et la définit comme une science visant à simuler les fonctions cognitives humaines. Marvin Minsky précise qu'il s'agit de construire des programmes capables d'accomplir des tâches relevant habituellement de l'intelligence humaine. Depuis, l'IA a connu des périodes de grands espoirs, comme les premières découvertes des années 1960 ou les systèmes experts des années 1980, chacune de ces périodes étant suivie de phases de désillusions, les « hivers de l'IA ».

On distingue deux grandes approches : l'IA symbolique, caractérisée par l'utilisation de règles logiques pour résoudre des problèmes, simulant un raisonnement déductif, et l'IA connexionniste, qui se fonde sur une analyse probabiliste de données, simulant un raisonnement inductif.

II. Les progrès de l'apprentissage profond dans les années 2010 et l'architecture *Transformer*

Un nouvel essor de l'IA a lieu depuis les années 2010 grâce aux progrès de l'apprentissage profond ou *Deep Learning*. L'apprentissage est dit « profond » car ses algorithmes reposent sur des réseaux de neurones artificiels composés de couches multiples. Chaque neurone représente une unité de calculs mathématiques qui correspond à une fonction d'activation. L'efficacité de ces modèles d'IA se développe dans les années 2010 grâce à trois éléments : la mobilisation de ces algorithmes, des corpus de données de taille inédite grâce à Internet, et une puissance de calcul des ordinateurs en augmentation rapide.

Depuis 2017, deux innovations majeures ont accéléré la mise en œuvre de l'IA :

- l'architecture *Transformer* : cette innovation ajoute à l'apprentissage profond un meilleur traitement du contexte, grâce à un procédé appelé « mécanisme d'attention » ;

- l'application de cette architecture à la création de systèmes d'IA générative accessibles à tous. Ceci a donné naissance à des « modèles larges de langage » ou LLM comme ChatGPT mais aussi à des outils permettant de créer d'autres contenus : images, sons, vidéos, etc. Ces systèmes exploitent des milliards de données et mobilisent des centaines de milliards voire des milliers de milliards de paramètres de calcul pour générer du contenu. Certains systèmes d'IA générative se basent sur d'autres architectures mais ils ne donnent pas d'aussi bons résultats, sauf à être hybridés à l'architecture *Transformer*, par exemple selon une méthode dite du mélange d'experts (*Mixture of Experts* ou MoE). Ces systèmes donnent alors des résultats d'une précision inédite.

III. Les questions technologiques et les perspectives d'avenir

Quels sont les avantages et limites de ces technologies ? Et quelles perspectives s'ouvrent pour le futur ? Côté avantages, les applications sont innombrables, côté limites les IA génératives doivent encore relever plusieurs défis technologiques. L'entraînement des modèles nécessite d'abord des ressources considérables, en infrastructures de calcul comme de stockage des données mais pas seulement : l'IA nécessite en effet des apports considérables d'énergie tout au long de son cycle de vie. Les modèles peuvent générer des erreurs appelées « hallucinations ». Leur production est sujette à des biais présents dans les données ou introduits au stade de leur programmation humaine ; certains ont d'ailleurs dénoncé des IA *wake*. Enfin, ces modèles continuent de poser les problèmes d'opacité déjà rencontrés pour les anciennes générations de *Deep Learning* : ces systèmes fonctionnent comme des « boîtes noires », rendant leur explicabilité complexe.

Ces défis nécessitent des efforts de recherche pour améliorer la fiabilité des résultats et réduire la consommation énergétique de ces systèmes. Faire plus avec moins : l'objectif est d'aller vers des IA frugales et efficaces. L'architecture Mamba, alternative aux *Transformers* repose sur les modèles en espace d'états structurés et articule plusieurs types d'IA. C'est une piste intéressante. Les perspectives pour l'IA seront donc sa frugalité mais aussi sa multimodalité et son agentivité.

Les IA multimodales traitent déjà des données variées (texte, image, son, etc.) et en traiteront bientôt davantage encore, en entrée comme en sortie, pour des usages de plus en plus diversifiés et intégrés. En 2024, GPT-4o (avec un « o » pour omni) a ajouté aux textes et aux images le traitement d'instructions vocales et même des interactions vidéo. La société xAI a ajouté à son modèle Grok 2 la génération d'images en septembre 2024 puis la compréhension d'images en octobre 2024.

L'agentivité est la capacité des systèmes à être autonomes et proactifs. La principale innovation en 2024 est celle des *Agentic Workflows*, IA basées sur des LLM et générant une série d'actions permettant une automatisation des tâches en s'adaptant en temps réel à la complexité des flux de travail. Ces outils devraient être particulièrement utiles pour les entreprises.

Les IA vont aussi devenir des interfaces et les principales plateformes d'accès aux services numériques. Ces interfaces rendront les interactions homme-machine plus fluides et pourraient devenir le point d'appui des *smartphones* et des ordinateurs, agrégeant les fonctionnalités de l'interface du système d'exploitation, des navigateurs *web*, des moteurs de recherche, des logiciels bureautiques, des réseaux sociaux et d'autres applications. Il est probable qu'à l'avenir les systèmes d'IA deviennent les colonnes vertébrales de contrôle des ordinateurs, à partir desquelles s'articuleront plusieurs services logiciels.

Des solutions technologiques sont également attendues pour faire face à la tendance des LLM à « halluciner », c'est-à-dire à générer des propos dénués de sens ou des réponses objectivement fausses sans émettre le moindre doute. La génération augmentée de récupération (*Retrieval Augmented Generation* ou RAG) constitue par exemple un moyen d'adjoindre une base de données à un LLM, qui, à l'aide d'un récupérateur, utilise les données de cette base en lien avec l'instruction de l'utilisateur.

IV. La conjugaison entre la logique de l'IA symbolique et l'efficacité de l'IA connexionniste

Les technologies d'IA peuvent être enchâssées les unes dans les autres et conjuguées pour produire de meilleurs résultats. Combiner l'IA connexionniste avec des modèles logiques ou des modèles de représentation du monde réel sera indispensable. L'espace-temps reste en effet inconnu des meilleures IA génératives.

On peut se rapprocher de telles articulations avec les Arbres de pensées ou *Trees of Thoughts* (ToT). Sans relever directement de l'IA symbolique, cette technique s'en rapproche par son recours à des étapes formelles de raisonnement, les idées venant s'articuler logiquement les unes par rapport aux autres. L'IA devient alors neuro-symbolique car empruntant à la fois des réseaux neuronaux et des raisonnements symboliques. Cette méthode peut être appliquée directement à tous les grands modèles de langage actuels grâce à une instruction décomposée en plusieurs phases de raisonnement. On parle pour ces techniques de *prompt engineering*. À l'avenir, outre cet art du prompt, on trouvera de nombreuses façons de combiner et d'hybrider les technologies, notamment les deux branches de l'IA, afin que les systèmes se rapprochent de nos raisonnements logiques.

V. La longue et complexe chaîne de valeur de l'IA

La chaîne de valeur de l'IA comporte une dizaine d'étapes. Tout commence avec l'énergie et les matières premières : les semi-conducteurs en silicium permettent la fabrication des puces, des logiciels permettent de concevoir ces microprocesseurs et des machines lithographiques gravent le silicium à l'échelle moléculaire. Nvidia est devenu, pour le moment, l'acteur dominant de ce premier maillon de la chaîne.

Le deuxième maillon, celui des infrastructures, se subdivise en de nombreuses couches :

- la collecte et le nettoyage de données ;
- le stockage de données dans de vastes *data centers* ;
- l'informatique en nuage (*cloud*) pour les calculs ;
- lors de la phase de développement des modèles, le recours spécifique à des supercalculateurs.

La multiplication de ces infrastructures entraîne des coûts très élevés et des impacts environnementaux considérables. Nvidia devra, par exemple, réaliser un chiffre d'affaires de 600 milliards de dollars pour lui permettre un retour sur investissement.

L'étape suivante est celle de la définition des modèles d'IA, elle-même subdivisée en plusieurs phases :

- la conception de l'architecture du modèle ;
- l'entraînement du modèle de fondation à l'aide des infrastructures et d'algorithmes ;
- le réglage fin par des apprentissages supervisés et une phase d'alignement.

Les modèles de fondation s'intercalent donc entre la définition de l'architecture du modèle et le *fine tuning*, ce qui permet aux systèmes d'IA d'être déployés pour des applications spécifiques ou d'être diffusés auprès du grand public.

La dernière étape est celle des utilisateurs, les applications caractérisant l'aval. Les systèmes d'IA ne deviennent accessibles aux utilisateurs qu'à travers une couche de services applicatifs, dont les fameux LLM d'OpenAI, comme ChatGPT, conçu à partir du modèle de fondation GPT-4.

Les enjeux politiques, économiques, sociétaux, culturels et scientifiques

Trois grandes catégories d'enjeux sont présentées : les problématiques politiques de l'intelligence artificielle, les transformations socio-économiques, et les défis culturels et scientifiques. L'IA est devenue une technologie incontournable, qui transforme nos sociétés et nos économies, mais aussi les rapports de force politiques et géopolitiques.

I. Les problématiques politiques

● La géopolitique de l'IA

La recherche en IA est dominée par la recherche privée, principalement américaine. Les États-Unis, grâce à leurs géants technologiques – les GAFAM, devenus les MAAAM désignant, dans l'ordre, Microsoft, Apple, Alphabet, Amazon et Meta – dominent l'écosystème global de l'IA, avec Nvidia en sus, qui fournit toutes les entreprises en processeurs graphiques. Outre son impact économique, l'IA devient en elle-même un levier de pouvoir géopolitique.

En parallèle, la Chine aspire à devenir leader mondial d'ici 2030, investissant toute la chaîne de valeur, notamment les semi-conducteurs. Ces derniers sont devenus un enjeu de premier plan qui dépasse la seule filière de l'IA. Alors que les deux tiers des puces sont à ce jour fabriquées à Taiwan, les autorités américaines et chinoises cherchent à attirer une part croissante de cette production sur leur sol.

Pour l'Europe, et notamment la France, le défi est celui de la souveraineté numérique, afin d'éviter de devenir une pure et simple « colonie numérique ». L'Union européenne mise aujourd'hui sur la régulation de l'IA, mais cela reste insuffisant face à la taille et à l'avance des puissances américaine et chinoise. La souveraineté numérique contre la domination de la Big Tech américaine appelle au développement d'acteurs français et européens puissants.

● Les risques de manipulations politiques

L'IA génère de nouveaux risques, notamment à travers la désinformation et les hypertrucages (*deepfakes*). Les fausses informations ou les trucages ont toujours existé, mais l'IA fait changer d'échelle.

Ces technologies permettent de produire instantanément et massivement des contenus falsifiés réalistes, capables d'influer sur les élections ou de ternir la réputation d'une personnalité publique. Des régulations imposant des filigranes sur ces contenus constituent une réponse, mais elles seront difficiles à appliquer.

● Les menaces pour la sécurité et le risque existentiel

L'IA facilite les attaques à grande échelle, à travers la création de logiciels malveillants ou la capture de données sensibles. Ces risques appellent des mesures de sécurité renforcées : sécurisation des modèles, analyse des risques et formation en cybersécurité des développeurs.

L'intelligence artificielle générale (IAG) qui dépasserait les capacités humaines reste une hypothèse incertaine. Bien qu'elle suscite des débats passionnés, ses bases scientifiques comme les lois d'échelle (*scaling laws*) restent fragiles. Certains craignent que cette IAG fasse courir un « risque existentiel » à l'humanité, comme Elon Musk qui prône même une hybridation homme-machine pour « rivaliser » avec l'IA tandis que d'autres, comme Yann LeCun, estiment que l'IA actuelle demeure très loin du tournant de la Singularité, n'ayant toujours pas de sens commun ou de capacités autonomes réelles.

II. Les impacts sociétaux et-économiques

● Santé et bien-être

L'IA améliore déjà la vie, à travers une multitude d'outils pour nous aider au quotidien. Elle optimise nos parcours de transport et surveille notre rythme cardiaque grâce des applications de santé. À l'avenir, diagnostics, dépistages précoces et traitements seront optimisés grâce à l'IA, la recherche médicale sera accélérée et des capteurs permettront d'anticiper des urgences médicales, comme des crises cardiaques. En dépit de tous ces avantages pour la santé, l'utilisation massive de l'IA pourrait aussi avoir des conséquences négatives sur la santé psychologique.

● Un impact économique incertain

Les travaux qui tentent de prédire l'impact de l'intelligence artificielle sur la croissance ou son potentiel en termes de gains de productivité divergent. L'IA peut stimuler certains secteurs, mais son impact global sur la croissance reste incertain dans un contexte de faible diffusion des technologies. La direction générale du Trésor (DGT) estime qu'il est encore trop tôt pour pouvoir estimer des prévisions chiffrées. Le paradoxe de Solow (« *on voit des ordinateurs partout sauf dans les statistiques de productivité* ») pourrait se confirmer pour l'IA. Il est vrai que la diffusion de l'innovation est toujours difficile à observer et encore plus à quantifier.

Le coût énergétique de l'IA conduit à un impact environnemental considérable et grandissant : ces systèmes ont une empreinte carbone et un impact sur la ressource en eau élevés, qui ne font que croître car les besoins en énergie de l'IA explosent. Le développement de l'IA menace donc l'atteinte des objectifs climatiques.

● Une transformation du marché du travail

Il est difficile d'évaluer précisément l'impact de ces technologies sur le marché du travail. Un consensus semble cependant se dégager : plus qu'un remplacement des emplois par l'IA, on va assister à une transformation des tâches et des métiers par ces technologies. L'ampleur et les modalités de ces transformations ne sont cependant pas mesurées de la même façon par toutes les études. Même si ces effets restent encore incertains, ils appellent un dialogue social pour accompagner les transitions. L'IA pourrait augmenter la productivité et parfois exacerber les inégalités, cela appelle une certaine vigilance.

Elle nécessite la mise en place de politiques publiques de formation initiale et de formation continue, dans le but d'anticiper les évolutions et d'accompagner les travailleurs. Des politiques de requalification ambitieuses sont indispensables.

III. Les défis culturels et scientifiques

● Une uniformisation cognitive

L'IA, dominée par des acteurs anglo-saxons, risque d'accentuer fortement l'hégémonie culturelle des États-Unis. Ce phénomène d'uniformisation culturelle appauvrit la diversité culturelle et linguistique, mais crée aussi une uniformisation cognitive. Le capitalisme cognitif qui repose sur la conjugaison des écrans et de l'IA conduit à une « économie de l'attention » préoccupante, notamment car elle enferme l'utilisateur des technologies dans des bulles de filtres. Cet enfermement informationnel polarise les visions de chacun dans des croyances subjectives. Ce sont autant de prisons mentales qui se déclinent à l'échelle individuelle. Cette tendance n'est que d'apparence paradoxale : on assiste, dans le même temps, à une polarisation marquée des opinions et des identités (selon les variables de la culture américaine, avec une forte dimension émotionnelle) et à une uniformisation culturelle doublée d'une uniformisation cognitive.

Ces conséquences de l'IA, des écrans et du numérique sur la cognition doivent nous mobiliser, surtout en direction des jeunes générations et des petits enfants, particulièrement victimes de ces impacts cognitifs et ce de manière irréversible. L'éducation au numérique en général et à l'IA en particulier est une urgence impérieuse pour la cohésion de nos sociétés et la santé de chacun.

La France doit défendre sa langue et ses spécificités culturelles face à des systèmes d'IA comme les grands LLM qui privilégient l'anglais et la culture américaine sur un plan linguistique et culturel. Nous ne devons pas donner notre langue à ChatGPT ! Il faut conserver la diversité linguistique et culturelle de l'humanité. La France a besoin de modèles d'IA les plus souverains possibles, reflétant sa culture, entraînés avec des données qui la reflètent fidèlement et qui mobilisent des sources issues de son riche patrimoine culturel et linguistique.

Par ailleurs, si les raisonnements par induction, probabilistes, sont prometteurs et donnent souvent des résultats impressionnants, ils tendent à faire oublier le grand intérêt des raisonnements déductifs, sur lesquels se sont construites la plus grande partie des connaissances scientifiques. La généralisation de cas particuliers sous l'effet des données massives traitées par l'IA connexionniste est devenue la règle, or le résultat d'une inférence suivant un raisonnement inductif, même fondé sur des milliards d'exemples, peut toujours être démenti par un ou plusieurs contre-exemples.

Les deux formes de raisonnement – déductif et inductif – doivent continuer à cohabiter de manière plus équilibrée sans quoi cette ère de l'IA et du Big Data va conduire tous les habitants de la planète à penser selon le même mode, non seulement, sans le savoir, avec les mêmes cadres en termes de références culturelles, mais aussi selon les mêmes structures cognitives, tournées vers l'induction.

● **Propriété intellectuelle et création artistique**

L'IA mobilise des données protégées par le droit d'auteur ou par le *copyright*. Les artistes et les créateurs sont confrontés à des questions inédites relatives aux droits d'auteur et à leurs modèles économiques. Par exemple, les œuvres générées par IA peuvent brouiller les frontières entre originalité et imitation, remettant en question les régimes traditionnels de protection de la propriété intellectuelle. Une réflexion doit s'ouvrir sur le sujet de la propriété intellectuelle et de la création artistique à l'heure de l'intelligence artificielle. Dans ce contexte d'incertitudes, les risques contentieux sont de plus en plus grands, qu'il s'agisse de l'utilisation d'œuvres protégées pour entraîner les modèles, de la protection des œuvres générées par des systèmes d'IA ou de tout autre litige qui pourrait émerger. En l'absence de règles claires, il reviendra aux juges de trancher les litiges. Le rôle de la jurisprudence sera donc central et laisse les artistes, les entreprises et les utilisateurs dans un flou juridique anxiogène, avec des risques financiers potentiellement non négligeables. C'est pourquoi une clarification de ces enjeux et des régimes juridiques applicables est indispensable.

● **Une révolution scientifique**

Dans le domaine scientifique, l'IA fertilise les autres disciplines et ouvre des perspectives immenses, comme en témoignent les exemples de la génomique, de la modélisation du repliement des protéines ou de la création de jumeaux numériques. Ces avancées permettront de résoudre de plus en plus de problèmes complexes et d'accélérer les découvertes. Il n'est pas anodin que les prix Nobel 2024 de Physique et de Chimie soient l'un et l'autre revenus à des chercheurs en IA. Les bénéfices potentiels de ces technologies nécessitent une adaptation de nos politiques de recherche.

Si l'intelligence artificielle est porteuse d'immenses opportunités, elle implique aussi de relever des défis complexes. L'IA soulève plusieurs questions éthiques, ce qui renvoie au thème de son alignement : comment garantir que l'IA s'aligne sur nos valeurs, respecte les droits de l'homme et les principes humanistes ? Il est à cet égard crucial de partager des bonnes pratiques, d'élaborer des cadres réglementaires, de renforcer la souveraineté technologique, et surtout d'éduquer nos sociétés aux enjeux de ces technologies. Ces perspectives nécessitent une gouvernance internationale pour encadrer les développements en cours et anticiper d'éventuels risques.

La gouvernance et la régulation de l'intelligence artificielle

Plusieurs initiatives ont été prises en matière de régulation de l'IA, à l'échelle de chaque pays à travers des stratégies nationales en IA, ainsi qu'au niveau international.

I. Une stratégie nationale pour l'IA en demi-teinte

Depuis 2017, la France a tenté de structurer une stratégie nationale mais le plan « France IA » lancé en janvier 2017 a été rapidement abandonné.

Un an plus tard, sur la base d'un rapport émanant de l'ancien président de l'Office, Cédric Villani, le Président de la République a annoncé le 29 mars 2018 une « stratégie nationale et européenne pour l'intelligence artificielle », qui visait à faire de la France un leader mondial en IA.

Cette stratégie a permis la labellisation de quatre Instituts Interdisciplinaires en Intelligence artificielle (3IA), le financement de chaires et de doctorats ainsi que l'investissement dans des infrastructures de calcul comme les supercalculateurs Jean Zay, inauguré en 2019, et Adastra, inauguré en 2023, et dont les performances atteignent respectivement 36,85 pétaflops et 74 pétaflops (Jean Zay devrait toutefois atteindre 125,9 pétaflops à la fin de cette année).

À titre de comparaison, l'entreprise d'Elon Musk xAI qui développe le système Grok s'est dotée du supercalculateur Colossus développant 3,4 exaflops, composé de 100 000 processeurs Nvidia Hopper 100. Sa taille devrait doubler d'ici quelques mois pour atteindre 200 000 processeurs. Le supercalculateur Jean Zay, après son extension prévue d'ici la fin de l'année 2024, sera quant à lui doté de 1 456 puces Nvidia Hopper 100.

D'autres limites de cette stratégie nationale en IA sont notamment relevées par la Cour des comptes dans un rapport d'avril 2023 qui fait apparaître :

- des résultats insuffisants, la France ayant continué à décrocher au niveau international depuis 2018 ;
- une coordination interministérielle insuffisante.

Le pilotage de la stratégie nationale en IA reste toujours défailant en 2024 : le coordinateur rattaché initialement à la direction interministérielle du numérique et des systèmes d'information et de communication (DINSIC), puis à la direction générale des entreprises (DGE) du ministère de l'économie, est sans autorité réelle sur la stratégie et sa mise en œuvre. L'instabilité du titulaire de cette fonction et les vacances répétées du poste témoignent d'une mauvaise définition du rôle de ce coordinateur national. La stratégie demeure en réalité sans pilote, évoluant au gré des annonces de l'exécutif et des événements organisés autour de l'IA.

La Commission de l'intelligence artificielle a proposé en 2024 un investissement massif de 27 milliards d'euros sur cinq ans pour la formation, la recherche, et le développement d'un écosystème robuste en IA. La mise à disposition de sommes aussi conséquentes pour la stratégie nationale en IA semble peu probable. Lors du rassemblement des plus grands talents français de l'IA à l'Élysée le 21 mai 2024, le Président de la République a annoncé un plan d'investissement de 400 millions d'euros pour financer neuf pôles d'excellence en IA, comprenant les quatre anciens Instituts 3IA, l'objectif étant de passer de 40 000 à 100 000 personnes formées à l'IA par an. Il pourrait être judicieux de commencer par reconduire le programme Confiance.ai, peu coûteux pour les finances publiques (3,75 millions d'euros par an).

II. Les comparaisons internationales et les projets de gouvernance mondiale de l'IA

Le rapport décrit près de vingt autres stratégies nationales, six mises en place dans l'Union européenne et onze dans le reste du monde.

La réglementation de l'Union européenne, complexe et peu propice à l'innovation, est analysée en détail dans le rapport. L'*AI Act* conjugue un dispositif régulant les usages de l'IA selon leur niveau de risque et un encadrement des modèles d'IA selon leur puissance.

Le rapport recense également toutes les initiatives ayant pour objet une gouvernance mondiale de l'IA. On en compte une dizaine, ce qui n'est pas efficace : les projets de l'ONU, de l'Unesco, de l'OCDE, du Conseil de l'Europe, du G20, du G7, du Partenariat mondial sur l'intelligence artificielle (PMIA) (ou *Global partnership on artificial intelligence* – GPAI), du Conseil du commerce et des technologies (CCT) UE-États-Unis, du Forum économique mondial, des Brics, de la Chine, etc., s'accumulent.

Ceci plaide pour une convergence autour d'une régulation internationale unique.

Le rapport met en évidence des efforts nationaux inégaux pour réguler l'IA, tout en soulignant le besoin d'une approche la plus coordonnée possible.

Le prochain sommet sur l'IA qui sera organisé par la France en 2025 sera une bonne occasion de promouvoir une gouvernance mondiale cohérente pour aborder les enjeux de l'IA dans leur globalité, sans se limiter au sujet de la sécurité.

L'Office a un rôle à jouer dans cette révolution technologique, pour aider le gouvernement à structurer sa propre stratégie, mais aussi pour contribuer à cette future gouvernance mondiale de l'IA.

Les propositions de l'Office

Le rapport propose 18 recommandations, dont cinq sont consacrées à la préparation du futur sommet de l'IA qui se tiendra à Paris les 10 et 11 février prochain.

LES PROPOSITIONS À SOUTENIR DANS LE CADRE DU FUTUR SOMMET DE L'IA

1°. Faire reconnaître le principe d'une approche transversale de l'IA et renoncer à l'approche exclusivement tournée vers les risques

Cinq thèmes seront l'objet du sommet (l'IA au service de l'intérêt public avec la question des infrastructures ouvertes ; l'avenir du travail ; la culture ; l'IA de confiance ; et la gouvernance mondiale de l'IA). Il faudra que le sommet aille plus loin et permette d'inscrire solennellement le principe d'une approche transversale des enjeux de l'IA au sein d'une déclaration des participants. Les cinq thèmes retenus pour le sommet éludent deux dimensions qu'il faudrait prendre en compte de manière prioritaire :

- l'**éducation**, qui pourrait être ajoutée à la verticale culture avec pour intitulé « éducation et culture » ;

- la **souveraineté numérique**, qui pourrait être ajoutée à la verticale « l'IA au service de l'intérêt public » avec pour intitulé « souveraineté numérique et intérêt général ».

2°. Proposer de placer la gouvernance mondiale de l'IA sous l'égide d'une seule organisation internationale

Le sommet doit être l'occasion de clarifier et rationaliser la dizaine de projets visant à créer une gouvernance mondiale de l'IA. Aussi, il est proposé de placer la gouvernance mondiale de l'IA sous l'égide d'une seule organisation internationale, à savoir l'ONU, seule organisation pleinement légitime sur le plan multilatéral.

3°. Initier le cadre d'une régulation globale et multidimensionnelle de l'IA en s'inspirant des travaux de l'OCDE et de l'UE

L'approche de la régulation mondiale de l'IA doit être multidimensionnelle, afin de traiter la question de l'ensemble de la chaîne de valeur de l'IA et comme le montrent les travaux de l'OCDE et de l'UE.

4°. Annoncer un programme européen de coopération en IA

Ce programme doit associer plusieurs pays dont au moins la France, l'Allemagne, les Pays-Bas, l'Italie et l'Espagne, ces pays partageant une vision similaire de l'IA et de ses enjeux.

5°. Associer plus étroitement le Parlement à son organisation

Afin de garantir une plus grande légitimité du futur sommet, l'Office demande que le Parlement soit plus étroitement associé à son organisation. La présence d'un député et d'un sénateur au sein du comité de pilotage du sommet serait un gage de crédibilité. Elle marquerait l'attention portée par les pouvoirs publics à l'indispensable dimension démocratique d'un encadrement de l'IA à l'échelle internationale.

6°. Développer une filière française ou européenne autonome sur l'ensemble de la chaîne de valeur de l'intelligence artificielle

Un objectif doit mobiliser les pouvoirs publics nationaux et locaux, les décideurs économiques, les associations et les syndicats : viser le développement d'une filière française ou européenne autonome sur l'ensemble de la chaîne de valeur de l'intelligence artificielle, même sans chercher à rivaliser avec les puissances américaine et chinoise.

En effet, **mieux vaut une bonne IA chez soi qu'une très bonne IA chez les autres**. Que ce soit au niveau européen, dans le cadre de l'UE ou avec une coopération renforcée entre quelques pays, ou directement au niveau national, la France doit relever ce défi de construire pour elle, en toute indépendance, les nombreux maillons de la chaîne de valeur de l'intelligence artificielle.

7°. Mettre en place une politique publique de l'IA avec des objectifs, des moyens et des outils de suivi et d'évaluation

Pour être efficace, la stratégie nationale pour l'IA ne peut se passer d'objectifs clairement définis, de moyens réels dont une gouvernance digne de ce nom, et d'outils de suivi et d'évaluation. Ces éléments sont aujourd'hui cruellement absents des politiques publiques menées en France en matière d'IA. Plus largement, la politique de la *Start-up Nation* avec son bras armé la French Tech, trop élitiste et souvent inadaptée, est à abandonner au profit d'une politique de souveraineté numérique, cherchant à construire notre autonomie stratégique et à mailler les territoires.

8°. Organiser le pilotage stratégique de la politique publique de l'IA au plus haut niveau

La stratégie nationale pour l'IA ne dispose pas d'une gouvernance digne de ce nom, il faudra mieux coordonner la politique publique nationale de l'intelligence artificielle et lui donner une réelle dimension interministérielle.

9°. Former les élèves de l'école à l'Université, former les actifs et former le grand public à l'IA

Il est indispensable de lancer de grands programmes de formation à l'IA à destination des scolaires, des collégiens, des lycéens, des étudiants, des actifs et du grand public. Les politiques conduites en ce domaine par la Finlande sont des modèles à suivre. La démystification de l'IA est une première étape importante et nécessaire pour favoriser la diffusion de la technologie.

10°. Accompagner le déploiement de ces technologies dans le monde du travail et la société, notamment par la formation permanente

S'il est difficile de prévoir l'impact précis que l'IA aura sur le marché du travail, il faut tout de même accompagner le déploiement de ces technologies, notamment l'IA générative, dans le monde du travail, en particulier par des programmes de formation permanente ambitieux. Des études qualitatives et quantitatives sur l'impact de l'IA sur l'emploi, le tissu social (dont les inégalités) et les structures cognitives devront également être régulièrement menées.

11°. Lancer un grand dialogue social autour de l'intelligence artificielle et de ses enjeux

Le dialogue social par la négociation collective peut être renouvelé par l'introduction de cycles de discussions tripartites autour de l'IA et de ses enjeux. Une opération d'envergure nationale, comme un Grenelle de l'IA, pourrait également être organisée. Le dialogue social autour de l'IA devra se décliner dans les entreprises pour permettre une meilleure diffusion des outils technologiques et un rapport moins passionné à leurs conséquences.

12°. Mobiliser et animer l'écosystème français de l'IA

Tous les acteurs de l'IA, la recherche publique et privée, les grands déployeurs de systèmes mais aussi l'ensemble des filières économiques doivent s'insérer dans une grande mobilisation générale. Le rapport préconise aussi des pôles d'animation régionaux. Des expériences étrangères peuvent être des sources d'inspiration comme la structure NL AI Coalition, créée par le gouvernement néerlandais, qui rassemble depuis cinq ans l'écosystème public et privé de l'IA aux Pays-Bas, avec le concours du patronat, des universités et des grands centres de recherche. Elle s'appuie sur sept centres régionaux.

13°. Reconduire le programme « Confiance.ai » ou mettre en place un projet équivalent

Le programme « Confiance.ai » réunissait dans une logique partenariale de grands acteurs académiques et industriels français dans les domaines critiques de l'énergie, de la défense, des transports et de l'industrie et avait pour mission de permettre aux industriels d'intégrer des systèmes d'IA de confiance dans leurs process. Il ne coûtait pas cher et était efficace. Or il s'est interrompu en 2024. Il est proposé de le reconduire ou de mettre en place un projet équivalent.

14°. Soutenir la recherche publique en intelligence artificielle selon des critères de transversalité et de diversification des technologies

La recherche privée en intelligence artificielle a pris beaucoup d'avance sur la recherche publique, mais cette dernière doit revenir dans la course. La soutenir davantage est un impératif. L'Office juge pertinent de l'orienter vers des activités transdisciplinaires et, plus globalement, transversales autour de « projets de recherche » en IA.

La diversification des technologies est également fondamentale : les avancées en IA se font par la combinaison et la recombinaison de savoirs et de savoir-faire, pas par l'enfermement dans un modèle unique.

Par exemple, l'IA symbolique ne doit pas être abandonnée, elle peut s'hybrider avec les IA connexionnistes pour forger de nouvelles approches logiques, imbriquant le signifiant et le signifié, plus proches des raisonnements humains.

D'autres technologies permettant d'apporter plus de logique aux systèmes d'IA générative peuvent également inspirer de nouvelles perspectives pour la recherche, comme les modèles *Mixture of Experts* (MoE), des modèles de représentation du monde (*World Models*), la génération augmentée de récupération ou *Retrieval Augmented Generation* (RAG) ou encore les arbres de pensées ou *Trees of Thoughts* (ToT).

Pour paraphraser Rabelais qui écrivait que « *Science sans conscience n'est que ruine de l'âme* », on peut affirmer que « *l'IA sans logique n'est qu'illusion d'intelligence* ».

15°. Relever le défi de la normalisation en matière d'intelligence artificielle

Il faut permettre à la France de défendre au mieux l'intérêt national ainsi que les intérêts de ses entreprises nationales en matière de normalisation de l'IA, ce qui implique de mobiliser davantage l'Afnor et le Cofrac.

La France doit également inviter ses partenaires européens à faire preuve d'une plus grande vigilance dans le choix de leurs représentants dans les comités responsables de la normalisation en IA : s'appuyer sur des experts issus d'entreprises extra-européennes, le plus souvent américaines ou chinoises, n'est pas acceptable.

16°. S'assurer du contrôle souverain des données issues de la culture française et des cultures francophones et créer des bases de données autour des cultures francophones

Il s'agit d'un acte de résistance face à la domination linguistique et culturelle américaine, qui caractérise l'IA aujourd'hui et qui fait courir un risque grave d'uniformisation culturelle et d'appauvrissement linguistique. Les initiatives conduites par certains pays, en particulier l'Espagne, peuvent inspirer notre pays.

17°. Préparer une réforme des droits de propriété intellectuelle, dont le droit d'auteur, pour les adapter à l'IA et surtout aux usages de l'IA générative

L'objectif d'une telle réforme sera à la fois de clarifier les régimes juridiques applicables, de protéger les ayants droit des données ayant servi à l'entraînement des modèles mais aussi les créateurs d'œuvres nouvelles grâce à l'IA.

18°. Confier à l'OPECST le suivi et l'évaluation régulière de la politique publique conduite par le Gouvernement

Les aspects scientifiques et technologiques de l'intelligence artificielle ainsi que les enjeux qu'ils soulèvent appellent une **expertise à la croisée des mondes politique et scientifique**, ce qui est la mission de l'OPECST. C'est pourquoi, comme l'avait d'ailleurs proposé la commission des lois de l'Assemblée nationale dans un rapport, il semble judicieux de confier à l'OPECST le suivi permanent des questions relatives à l'intelligence artificielle.

Rapport Assemblée nationale n° 642 (XVII^e législature) – Sénat n° 170 (2024-2025)

Pour consulter le rapport :

www.senat.fr/opecest

<https://www.assemblee-nationale.fr/dyn/17/organes/delegations-comites-offices/opecest>

Assemblée nationale - 101 rue de l'Université - 75355 Paris 07 SP – Tél : 01 40 63 26 81 – Mél : secretariat-opecest@assemblee-nationale.fr
Sénat - 15 rue de Vaugirard - 75291 Paris Cedex 06 – Tél : 01 42 34 27 20 – Mél : opecest-secretariat@senat.fr