Le sommet sur la sécurité de l'IA

Contenir un nouveau risque existentiel pour l'Humanité ?

Le 1er et 2 novembre 2023 s'est tenu le premier sommet sur la sécurité de l'IA (Intelligence Artificielle) sous le patronage du Premier Ministre britannique Rishi SUNAK au Bletchley Park de Londres. 28 États y étaient représentés (Afrique, Moyen-Orient, Union Européenne, Amérique du Nord, Amérique Latine, Asie), ainsi que le secteur privé (Elon MUSK, OpenAI, Microsoft, Salesforce, Google, Meta, AWS), des institutions académiques de premier plan (Chine, États-Unis, Europe) et d'instances de l'Organisation des Nations Unies.

Par Gabriel GERVAIS

AXE PHARE "CROISSANCE & INNOVATION"

Depuis la Seconde Guerre Mondiale et l'invention de la machine Turing, l'Intelligence Artificielle (IA) a connu différents développements dus à la loi de Moore et à la miniaturisation des transistors. Après une période dite « d'hibernation » durant laquelle l'intérêt pour la recherche et développement dans l'IA, et les investissements dans l'IA ont décru, le krach de 1987 et la fin de la Guerre Froide ont marqué un nouveau tournant. Ce nouveau tourant a été franchi avec l'émergence d'IA génératives.

En outre, l'appellation « IA » reflète également un trait anthropologique hérité de la Renaissance, de PIC DE LA MIRANDOLE et de son idéal autoréférencé de l'Homme s'autoaméliorant. Le terme « IA » incarne ainsi ce rêve positiviste de reproduire voire de transcender la propre intelligence humaine. Ce dépassement verrait le développement d'une IA générale, capable de se passer de l'Homme.

Ce spectre d'une singularité technologique et les horizons qu'offre la recherche dans les IA générales catalyse une nouvelle révolution industrielle et un nouveau risque existentiel. Ainsi, l'économiste Charles I. JONES (Stanford, octobre 2023) mis en lumière le dilemme que pose l'IA : d'un côté, elle porterait une croissance économique de long terme (gains de productivité, meilleure allocation du capital et des ressources, découvertes scientifiques), d'autre part, elle pourrait mener à des catastrophes (le risque de décalage, les mésusages de l'IA). Il rejoint le lauréat du prix Nobel William D. NORDHAUS (*American Economic Journal*, Janvier 2021) qui soulignait qu'outre le risque d'extinction, l'intégration de l'IA dans l'économie avait de nombreuses externalités négatives de long terme (l'effet Baumol, la hausse des inégalités, le chômage sur les postes substituables). Il préconise donc la mise en place préventive de tests diagnostiques pour évaluer le degré de rapprochement de la singularité technologique et la mise en place d'une régulation pratique et éthique de l'intégration de l'IA dans les économies. Ce dernier point est tout l'enjeu de ce premier sommet sur la sécurité de l'IA qui inaugure une myriade d'autres dans les prochaines années.

Une réflexion sur les enjeux immédiats et de court-terme de l'IA

Le premier sommet sur la sécurité de l'IA a balayé différents risques immédiats, sans les détailler précisément, que posaient ses avancées. Trois enjeux sont ressortis des discussions :

- La rivalité géopolitique : la compétition sino-américaine exacerbe les développements d'IA innovantes, que ce soit dans le champ militaire que de la guerre informationnelle ou la cybersécurité ;
- La protection des données : avec l'augmentation des capacités de calcul de l'IA et de ses bases de données, la protection des données personnelles est un garde-fou aux biais déterministes du *data learning*. La garantie d'empêcher ses données de devenir un bien public est d'autant plus importante que la fiabilité de l'IA dans l'exploitation de ces données est encore difficile à cerner : contrairement à un algorithme, une IA est incapable d'expliquer comment elle aboutit à un résultat prédictif ;
- La cybercriminalité : les mésusages de l'IA à des fins de cyberattaques et d'usurpation d'identité sont un enjeu nécessitant une régulation immédiate.

Ces trois enjeux ont mis en exergue la nécessité de réguler multilatéralement les développements de l'IA dans un premier temps. Il en résulte plusieurs propositions :

- Elon MUSK a proposé la création d'un arbitre indépendant régulant et suivant les recherches et les développements de l'IA en entreprises. Cette régulation permettrait d'encourager les externalités positives des avancées de l'IA tout en limitant les risques associés ;
- Deux sommets sur la sécurité de l'IA ont par ailleurs été annoncés afin d'approfondir les discussions réglementaires : un sommet en

Corée du Sud puis un sommet en France courant 2024/2025.

Une déclaration commune : plus de questions que de réponses

A la fin de ce premier sommet sur la sécurité de l'IA, une déclaration commune a posé les jalons d'une réglementation future sans pour autant définir les orientations concrètes. Les participants ont toutefois reconnu la responsabilité des concepteurs de systèmes d'IA avancées dans les effets systémiques de leurs technologies.

Au-delà d'une déclaration de principe, cette déclaration révèle également la concurrence géopolitique dans la formalisation des normes internationales. En effet, les ambitions géopolitiques de la Grande-Bretagne de peser face aux États-Unis et à la Chine *« dans l'élaboration d'une régulation internationale de l'IA »*. L'approche britannique semble être unique et avant-gardiste en ce sens qu'elle n'envisage pas de nouvelle législation pour réglementer l'IA, préférant que les régulateurs existants soient responsables de l'IA dans leurs secteurs respectifs.

De plus, la déclaration omet trois enjeux de moyen-terme dans la régulation de l'IA :

- L'IA joue un rôle notable dans le domaine de la sécurité et de la défense en allant de la guerre informationnelle, la gestion des drones, l'OSINT et du cyber-renseignement à la dissuasion nucléaire. Pourtant la déclaration n'en fait pas mention ;

- Le secteur financier est également en première ligne des avancements de l'IA, de développement de la DeFi (Finance décentralisée) et la généralisation de l'usage des IA génératives. Or, il n'en est pas fait mention. Pourtant, le Fonds Monétaire International et la Banque des Règlements Internationaux ont mis en avant le potentiel de risque systémique que pourrait comporter un usage de l'IA non encadré ;

- Enfin, le risque existentiel d'une singularité technologique n'est pas mentionné.

Conclusion

Bien que saut qualitatif, ce premier sommet sur la sécurité de l'IA apporte plus de questions qu'il n'apporte de réponse. Il met en lumière une nouvelle problématique qui participe à une incertitude de plus en plus radicale. La question des prochains sommets sera de savoir si l'IA est un risque existentiel supérieur au risque de conflagration nucléaire et au changement climatique. ■

Nos recommandations

L'évaluation

Les Conférences des Parties (COP) proposées auraient pour but d'évaluer les répercussions de l'intelligence artificielle sur la sécurité et l'éthique, avec la collaboration d'organismes internationaux comme l'OIT (Organisation Internationale du Travail), l'UIT (Union Internationale des Télécommunications), l'OCDE, le G20, la Banque Mondiale et le Fonds Monétaire International. Cette évaluation pourrait était être ensuite déclinée aux régulateurs sectoriels existants aux échelons supranationaux et nationaux afin d'optimiser une gouvernance de la sécurité de l'IA.

La transparence

Instaurer une agence internationale dédiée à la sécurité de l'IA sur modèle analogue à l'AIEA (Agence Internationale à l'Énergie Atomique) permettrait de mettre en place des protocoles internationaux et des systèmes d'inspection similaires à ceux régulant le nucléaire civil et militaire pour les IA fortes et la recherche en IA générale.