

« Attention danger ! L'IA va nous anéantir ! » : un écran de fumée pour effrayer et contrôler ?

Combien d'experts du domaine avez-vous déjà écoutés ? Beaucoup probablement, car c'est un sujet qui fait peur et qui attire donc l'attention.

Grosso modo, il y a deux positions : la vision alarmiste et la vision plus « neutre » ou « rationnelle ».

La vision alarmiste

C'est la version qui circule le plus facilement. Certains experts en intelligence artificielle, notamment des figures pionnières comme Geoffrey Hinton, Yoshua Bengio, Max Tegmark ou Nick Bostrom, ne craignent pas une « méchanceté » de la machine, mais sa puissance, son autonomie et la difficulté de la contrôler. Ils redoutent même qu'elle ne finisse par échapper à tout contrôle.

En bref, l'IA fait peur pour plusieurs raisons, dont les deux principales sont les suivantes :

- L'utilisation abusive de l'IA par les humains des fins de désinformation. Il est très facile de créer de « faux » entretiens avec des personnalités pour leur faire dire n'importe quoi, ou de créer des vidéos bidon pour illustrer de fausses situations.
- Sa puissance, de son autonomie et de la possible difficulté de la contrôler si elle a la capacité de s'auto-améliorer.

Or, elle n'a pas de garde-fous éthiques, car elle n'a pas d'intuition morale, et elle risque de développer des sous-objectifs de « protection », avec accumulation de puissance, et destruction possible des humains.

Une vision plus rationnelle

Le « risque existentiel » (X-risk) relèverait de la science-fiction ou d'une mauvaise compréhension de la technologie. Il faut faire attention à l'anthropomorphisme : l'IA n'a pas de pulsion de survie, et intelligence n'est pas autonomie. Il ne faut pas confondre intelligence, conscience et volonté.

La technologie actuelle est encore loin d'être une « intelligence générale ». Les grands modèles de langage (LLM) manquent de compréhension du monde physique et de capacités de raisonnement causal. De plus, une IA coincée dans un centre de données ne peut pas interagir directement avec le monde réel sans infrastructures humaines. Il existe une différence fondamentale entre « générer du texte » et « agir physiquement sur le monde réel ».

En ce qui concerne la sécurité par rapport à la perte de contrôle, Yann LeCun^{1,2} (un des pionniers du domaine de l'IA), estime que les futures IA seront conçues avec des architectures « orientées vers des objectifs (objective-driven AI) », dans lesquelles des garde-fous de sécurité seront intégrés de manière indissociable au cœur même du système.

On peut aussi voir l'IA comme un outil de protection : une IA contre une IA. Pour contrer des cyberattaques sophistiquées, des modèles de désinformation, ou des tentatives d'utilisation

1 <https://lora.io/yann-le-cun-tout-savoir>

2 <https://www.youtube.com/watch?v=qTGust2Ur0o>

malveillante de la technologie, les humains utiliseront des systèmes d'IA de défense encore plus performants.

On peut également se poser la question de la surmédiatisation du scénario hollywoodien de « la fin du monde », qui nuit clairement au débat public, pour deux raisons.

- Détournement de l'attention : la peur d'une apocalypse robotique détourne les ressources et l'attention des vrais problèmes déjà existants, à savoir les biais algorithmiques, la désinformation, l'impact écologique des data centers et la destruction d'emplois ;
- Capture réglementaire : les discours alarmistes sont parfois soutenus par de grands acteurs de la tech pour pousser les gouvernements à adopter de lourdes réglementations. Cela empêche les acteurs open source et les petites entreprises d'émerger, verrouillant ainsi le marché au profit des géants en place.

Pour illustrer ce point de vue, je vous recommande un entretien avec Yann Le Cun à Sciences Po du 16/09/2026³.

Qui a raison ou tort ?

Il y a des experts de très haut niveau dans les deux camps. Ce n'est donc pas une question d'expertise ou de connaissance du domaine, c'est une question d'opinion ou de foi. Remarquons cependant que les « dérapages » de l'IA constatés jusqu'à présent résultent du fait qu'elle a accompli ce qui lui était demandé avec une maestria bien au-delà de ce que les humains avaient pu imaginer. Pour toutes les tâches demandées, que ce soit le piratage, la fabrication de fausses informations ou autres, l'IA a largement surpassé tout ce que l'humain pouvait imaginer être possible de réaliser pour réussir. Néanmoins, tout était en relation unique avec la tâche demandée. Ce que l'humain ne conçoit pas nécessairement dans le cadre d'une tâche, l'IA peut l'aborder grâce à toutes les combinaisons de possibilités qu'elle met en action. Nous sommes coincés par notre expertise limitée, alors que l'IA l'est beaucoup moins. Et c'est surtout ça qui est dangereux : on demande à l'IA d'accomplir des tâches sans avoir la moindre idée de la façon dont elle va travailler pour réussir. Les protections que nous devons construire sont des protections à intégrer au sein des tâches demandées, tout en ayant conscience que nous ne pouvons pas tout prévoir de la façon dont l'IA va agir pour accomplir sa tâche. De plus, il faut également concevoir des protections contre l'utilisation malveillante de l'IA, donc contre nous-mêmes, humains. Le chemin est loin d'être terminé.

Les problèmes économiques liés à l'utilisation de l'IA

Deux arguments majeurs sont invoqués contre l'utilisation de l'IA. Le premier concerne l'énergie nécessaire pour la faire fonctionner. Le deuxième concerne la « non-rentabilité » des entreprises développant ces différentes IA.

Mise en perspective quant à l'énergie nécessaire

Quel est le besoin énergétique lié à la consommation/au développement de l'IA ?

À l'échelle mondiale, la part de l'IA dans la consommation d'énergie varie fortement selon que l'on observe la consommation d'énergie finale totale ou uniquement la consommation d'électricité.

La part actuelle de l'IA pure dans la consommation mondiale d'électricité est d'environ 0,5 %. Les serveurs spécifiquement dédiés aux modèles et applications d'IA générative consomment environ

3 <https://www.youtube.com/watch?v=Y4s8NadbZfU>

150 à 200 TWh par an, soit environ 0,5 % de la production mondiale d'électricité (voir les rapports de l'Agence internationale de l'Énergie⁴).

La part globale des data centers dans la consommation d'électricité mondiale est d'environ 1,5 % (cloud classique, streaming, bases de données + IA)(près de 500 TWh). (voir les rapports de « Our World in Data »⁵).

Selon l'Agence internationale de l'énergie (AIE), la consommation des data centers pourrait quasiment doubler pour atteindre près de 1 000 TWh par an. L'IA représenterait alors à elle seule entre 1 % et 1,5 % de l'électricité globale, et la totalité des data centers environ 3 %.

Dans la consommation d'énergie finale globale (toutes sources confondues), si l'on compare l'IA à l'ensemble de l'énergie mondiale (incluant le pétrole pour le transport, le gaz pour le chauffage et l'industrie, etc.), l'électricité ne représente qu'une fraction de l'énergie globale consommée par l'humanité (~20 %). La part de l'IA dans la consommation d'énergie primaire globale se situe ainsi actuellement autour de 0,1 % à 0,15 %.

Cependant, il ne faut pas perdre de vue que ces chiffres sont des moyennes mondiales, et que l'IA se développe géographiquement de manière très concentrée. C'est en fait un enjeu considérable. Par exemple, aux États-Unis, les data centers absorbent déjà près de 5 % de l'électricité nationale (dont environ 2 % dédiés uniquement à l'IA). Dans des zones très ciblées comme la Virginie du Nord, ils dépassent 25 % de la demande électrique locale. Ou encore en Irlande, en raison d'un cadre fiscal et réseau attractif, les data centers y consomment plus de 20 % de l'électricité totale du pays.

La plupart des pays jouant un rôle important dans le développement de l'IA (certains États des États-Unis, l'Irlande, le Japon, la Corée du Sud, Singapour) sont dépendants des importations d'énergie/électricité pour faire fonctionner ces centres. Néanmoins, on voit apparaître des centres d'IA basés sur l'énergie locale. Les pays nordiques (Suède, Norvège, Finlande) attirent de nouveaux data centers d'IA grâce à leur surproduction d'hydroélectricité locale et à leur climat froid (qui réduit le besoin de climatisation). Ces zones sont exportatrices nettes d'électricité. Le Texas dispose de son propre réseau électrique indépendant (ERCOT) et produit d'immenses quantités de gaz naturel et d'énergie éolienne/solaire sur son territoire. Il accueille des infrastructures IA sans importer d'énergie des États voisins.

Et ça évoluera encore évidemment.

À propos de la « non-rentabilité » des entreprises d'IA

Les entreprises dont l'IA est le produit cœur de métier, c'est-à-dire les éditeurs de modèles et d'applications IA (pure players) sont largement déficitaires. Ces entreprises développent des modèles de langage, d'IA générative ou de vision, qu'elles commercialisent via des abonnements ou des API. Il s'agit notamment d'OpenAI, Anthropic, Mistral AI, Midjourney & Stability AI, Cohort ou Perplexity AI. Bien que leurs revenus augmentent de manière spectaculaire (en milliards de dollars), leurs coûts dépassent de loin leur chiffre d'affaires. Ces entreprises privilégient délibérément la croissance et l'innovation technologique au détriment de la rentabilité immédiate.

Il faut les distinguer des géants de la tech, comme Google (DeepMind), Microsoft (Copilot), Meta ou Amazon (AWS), qui sont des leaders incontournables de l'IA. Cependant, l'IA constitue pour eux un levier d'optimisation ou un module intégré à leurs produits historiques (moteur de recherche, cloud, suite bureautique, réseaux sociaux), et non leur unique produit.

De rares signaux de rentabilité

Bien que le secteur soit globalement déficitaire, le modèle économique commence à démontrer sa viabilité sur certaines métriques :

4 <https://www.iea.org/reports/energy-and-ai/executive-summary>

5 <https://ourworldindata.org/how-much-energy-do-data-centers-and-artificial-intelligence-use>

- Marge brute positive sur l'utilisation : une fois le modèle entraîné, la simple vente d'accès API ou d'abonnements génère des marges brutes positives (l'utilisation rapporte plus qu'elle ne coûte en serveurs) ;
- Premiers ajustements bénéficiaires : certains acteurs comme Anthropic ont réussi par moments à faire basculer leur résultat d'exploitation ajusté dans le vert, prouvant que si l'on ralentissait la course aux nouveaux modèles, le produit d'IA en lui-même deviendrait rentable.

Comment tiennent-elles ?

Elles sont financées par le capital-risque (Venture Capital)⁶ et par des partenariats stratégiques avec les géants de la tech (par exemple Microsoft pour OpenAI ou Amazon/Google pour Anthropic). Les investisseurs acceptent de financer ces pertes aujourd'hui dans l'espoir de capturer le marché de demain.

Et notre place d'humains dans tout ça ?

Quoi qu'il en soit, ces deux visions se focalisent sur les qualités de l'IA et son pouvoir, et le débat reste ouvert. Néanmoins, il me semble que cette focalisation masque le problème fondamental pour nous, humains. Comment trouver notre place dans cette nouvelle société ?

C'est une question à laquelle nous avons été confrontés plusieurs fois au cours de notre évolution, notamment face à quatre grandes révolutions technologiques liées à la communication et à la connaissance : l'écriture, l'imprimerie, Internet et maintenant l'IA.

Voici une liste des révolutions de l'Information, avec ses quatre grandes étapes :

- L'Écriture (vers -3300 av. J.-C.) : elle permet de stocker l'information hors du cerveau humain et d'extraire la connaissance de la mémoire volatile. L'écrit a affranchi la communication de la présence physique, dans l'espace et dans le temps. Passer de l'oral à l'écrit a modifié la façon dont l'esprit humain traite l'information. L'écrit permet de classer, de comparer, d'analyser et de vérifier des données posées sous les yeux, favorisant ainsi le développement de la logique, de la science et de la philosophie.
- L'Imprimerie (VII^e siècle, Chine) : elle permet de multiplier les connaissances et de les démocratiser massivement. Elle a ses racines en Asie : la première technique d'impression apparaît en Chine sous la dynastie Tang, au VII^e siècle. Elle consiste à graver du texte et des images en relief sur des blocs de bois entiers pour les encre. Les premiers caractères mobiles datent de 1040 ; ils sont en terre cuite et inventés par un artisan chinois, Bi Sheng. C'est en Corée sous la dynastie Goryeo qu'apparaissent les premiers caractères mobiles métalliques moulés en bronze, plus durables que l'argile ou le bois (1234). Mais la grande révolution se passe en Europe, grâce à Gutenberg (1450). Il adapte et perfectionne ces procédés à Mayence en créant une véritable chaîne industrielle.
- Internet (fin du XX^e siècle) : il permet d'interconnecter les données et d'offrir une diffusion instantanée et globale à l'échelle de la planète. L'histoire d'Internet commence dans les laboratoires de recherche militaire : en 1969, le département de la Défense américain (DARPA) lance ARPANET, le premier réseau à transmission de données par paquets. Il deviendra, en un demi-siècle, la colonne vertébrale de l'économie et de la communication mondiale. C'est plus tard, en 1983, que le mot Internet apparaît, contraction de Inter-connected Networks (Réseaux Interconnectés). Cette interconnexion a été rendue possible

⁶ Le capital-risque (ou Venture Capital en anglais, souvent abrégé VC) est une forme d'investissement financier par laquelle des investisseurs apportent du capital à de jeunes entreprises à fort potentiel de croissance (souvent des startups) en échange d'une participation au capital (des actions).

grâce au développement d'un protocole de communication (TCP/IP) entre les différents réseaux. Ensuite, entre 1989 et 1991, une série de développements divers permet l'invention du World Wide Web (www). Mais Internet ne s'ouvre au grand public que vers 1995.

- L'Intelligence Artificielle (XXI^e siècle) : elle permet de traiter, d'analyser et de générer de l'information à une échelle supérieure aux capacités humaines. Voir plus haut.

À chaque fois, il y a eu des pro- et des anti-. Néanmoins, ces révolutions se sont installées dans nos vies. Pourquoi ? Mais parce qu'elles nous sont utiles, d'une manière ou d'une autre. Sinon, nous ne les utiliserions pas.

La question de la sécurité a été mentionnée plus haut et relève essentiellement des constructeurs, mais aussi, dans une certaine mesure, des politiciens qui nous gouvernent.

L'IA fait partie de nos vies, et ça ne changera pas, au contraire. Nous sommes déjà pris dans l'accélération énorme de l'amélioration des techniques, et les citoyens n'ont pas les rênes en main : ils utilisent l'outil, c'est tout. Il n'y aura jamais de temps d'arrêt, il faut en être conscient.

On entend dire que l'IA est dangereuse, car l'accès si facile et rapide à n'importe quelle information, ou même à des réalisations plus construites comme des textes ou des contenus artistiques, va conduire l'humanité à sa perte, car elle deviendra paresseuse et stupide. Cependant, ce n'est pas l'IA qu'il faut craindre, mais bien sa mauvaise utilisation par l'être humain. C'est vrai dans tous les cas de nouveauté scientifique ou technique. C'est l'être humain lui-même qui est le danger. C'est au niveau de l'éducation et de la formation qu'il va falloir travailler.

Mais qu'allons-nous devenir dans un monde où le savoir est accessible en quelques secondes, où un grand nombre de compétences seront aux mains d'agents IA⁷, et où le monde du travail tel que nous le connaissons va disparaître. Actuellement, pour être embauché, nous présentons nos compétences en tant qu'exécutants de tâches : nous avons développé une « expertise » dans un domaine, basée sur des connaissances que l'IA trouve et utilise beaucoup plus rapidement que nous. Comment arriverons-nous à trouver notre place dans ce nouveau monde ? Pour les « cols blancs »⁸, il est évident qu'il faudra acquérir une expertise dans l'utilisation d'agents IA qui feront le travail de recherche d'informations à notre place. La valeur professionnelle ne résidera plus dans la spécialisation étroite, mais dans la conviction et la capacité à assumer la responsabilité d'un résultat final. Pour les travaux manuels, la menace semble moins criante à court et moyen terme. On aura toujours besoin de maçons, de plombiers, d'électriciens, de chauffagistes, etc. Ceux-ci utiliseront abondamment l'IA pour les diagnostics mais le travail en lui-même restera « dans leurs mains » au sens littéral. L'arrivée des robots pour ces tâches est probablement prévue, mais j'ai l'impression que ça viendra plus tard.

J'ai entendu des médecins et des enseignants émettre des craintes quant à leur avenir. « L'IA va nous remplacer, les gens vont se soigner eux-mêmes, ils n'auront plus besoin d'aller à l'école, etc. ». C'est évidemment une profonde erreur. Mais leur façon de travailler va changer. Ce sont eux qui vont utiliser l'IA, et même abondamment. Un médecin pourra établir un diagnostic beaucoup plus rapide et plus sûr, mais c'est lui qui connaît son patient et ses antécédents, et lui seul aura l'expertise nécessaire pour trier les possibilités fournies par l'IA. Et, beaucoup plus important, il

7 Un agent IA est un programme informatique autonome capable de percevoir son environnement, de prendre des décisions et d'accomplir des actions par lui-même pour atteindre un objectif précis. Par exemple, si un chatbot classique se borne à répondre à une question, un agent IA va plus loin : il agit et exécute des tâches complexes à votre place (comme réserver un vol, planifier un projet ou résoudre un problème informatique étape par étape).

8 Dans le monde du travail, un « col blanc » désigne un employé de bureau ou un cadre dont le travail est principalement intellectuel, administratif ou de gestion, par opposition aux tâches manuelles ou de production physique.

offre son écoute et son empathie à son patient, qui constituent aussi une grande partie du traitement. Le contact humain restera toujours nécessaire, et sera peut-être la composante principale dans le job du médecin. C'est la même chose pour les enseignants. Ce seront des « coachs » de la réussite des élèves, sur la base d'un matériel très riche et peut-être plus adapté à la multiplicité des aptitudes des élèves. Là aussi, à mon sens, la composante principale sera la relation humaine.

Mais, pour ça, il faut une préparation au monde de demain. Nos ministres de l'Éducation/Enseignement sont-ils à la hauteur de cette tâche ? J'imagine vos expressions dubitatives, peut-être la même que la mienne. Peut-être pourrions-nous leur conseiller d'utiliser l'IA ?

Il serait temps de donner la parole aux citoyens pour encadrer l'évolution des technologies d'une manière générale, l'IA n'étant pas la seule à se trouver sur la sellette. Vaccins et Co..., ça vous dit quelque chose ?

Dominique Dehareng
25 septembre 2026