

# LA LETTRE DE XMP-CONSULT

## RÉSEAU D'EXCELLENCE DU CONSEIL

### ÉDITORIAL

### L'IA, le vertige de la force

Les performances de ChatGPT ont à juste titre beaucoup impressionné tous ceux qui l'ont essayé. L'étendue des applications possibles, et les changements qu'occasionnera l'utilisation généralisée des robots conversationnels reposant sur l'intelligence artificielle (ChatGPT, ses concurrents, et quelques autres applications s'attachant davantage au traitement de l'image) donne quelque peu le vertige, et, à certains, des sueurs froides. Que va-t-il advenir de nos emplois ?

Certes, l'arrivée de nouvelles technologies a toujours suscité des craintes dans le public, bien souvent irrationnelles. Un journal de médecine s'indignait ainsi en 1925 du développement des recherches sur la transfusion sanguine, estimant que « transfuser le sang d'un autre dans les veines d'un patient provoque à court terme la mort ou au moins la folie » et une société scientifique française estimait en 1927 que « Monsieur Charles Lindbergh vient de réussir, il est vrai, un exploit peu banal, mais jamais les compagnies d'aviation ne seront assez folles pour confier la vie d'un équipage, voire celle de passagers, sur des distances aussi grandes, à ces machines si peu confortables et si peu sûres. » Un siècle plus tôt, l'Académie des Arts et des Sciences avait déjà proposé d'interdire d'urgence en France la locomotive de Stephenson « un monstre redoutable, une folie criminelle. » Et la liste de ce type de déclarations pourrait être allongée ad libitum...

Nous voici de nouveau confrontés à l'arrivée d'une nouvelle technologie que tous s'accordent à trouver révolutionnaire. Faut-il pour autant la craindre, particulièrement si l'on subodore un risque de concurrence avec nos métiers ? Ou bien, éclairés par les exemples du passé, faut-il s'y lancer tête baissée, puisque le progrès ne recule jamais ? (On se souvient de la vaine révolte des luddites, ces artisans du textile britanniques qui avaient tenté, dans les années 1811-1812, de stopper le développement de la mécanisation en détruisant les nouvelles machines.)

Les questions liées à l'IA sont complexes et la vérité pourrait bien être toute en nuance. Pour

s'en approcher, il faut comprendre quelles sont les perspectives ouvertes par l'IA et quelles en seront les évolutions probables dans les prochaines années. C'est ce que nous avons essayé de faire dans ce numéro de la Lettre de XMP-Consult, avec les approches complémentaires apportées par plusieurs membres et sympathisants de notre association.

Philippe Auroy nous décrit ainsi l'importance de l'atout que représente le recours à l'IA dans le domaine de la recherche scientifique, et Tru Do-Khac nous commente comment ChatGPT appréhende un problème, ici de mathématique. Par le biais d'un récit prospectif, Guillaume O'Lanyer illustre aussi la fiabilisation importante que l'IA peut apporter à une entreprise logistique de pointe. Puis, au-delà de l'amélioration des services que permet l'IA, Claudie Boudet explore les émotions variées que cela suscite en nous.

Place ensuite aux experts : Philippe Coution nous éclaire d'abord sur l'ensemble du tsunami législatif et réglementaire qui approche, et il nous fournit quelques conseils de survie. Arnaud Ioualalen et Baptiste Aelbrecht, apportent quant à eux leur éclairage professionnel sur les moyens de construire la nécessaire confiance que requiert le déploiement de l'IA dans de nombreux domaines où son apport peut être très positif.

Nous terminons par une prise de recul : Antoine Herlin nous livre une réflexion prospective sous forme de 4 scénarios possibles, Antoine Jaulmes tire la sonnette d'alarme : il faut anticiper au maximum les transformations sociales induites par l'IA, et Damien Ribon identifie dans la mise en œuvre de la RSE une piste pour permettre une saine utilisation de l'IA et pour faciliter son acceptation de dans les entreprises.

A l'issue de la lecture de ces différentes contributions, nous pouvons en tous cas rester assurés qu'il y aura toujours besoin d'intelligence humaine, pour exploiter les résultats de l'intelligence artificielle et peut-être pour en atténuer

### SOMMAIRE

P. 02-03  
**L'irruption de l'IA dans l'innovation technique**  
Par Philippe Auroy

P. 04-06  
**Et si ChatGPT avait passé le Brevet 2022, épreuve de mathématiques ?**  
Par Tru Do-Khac

P. 07-08  
**Quand un chatbot sauve de précieux vaccins...**  
Par Guillaume O'Lanyer

P. 09-10  
**L'IA, en serons-nous jaloux ou admiratifs ?**  
Par Claudie Boudet

P. 11-14  
**L'EU AI Act, quels risques, quels impacts, quels coûts pour les sociétés et leur dirigeants ?**  
Par Philippe Coution

P. 15-17  
**L'intelligence artificielle, la confiance et la réglementation**  
Par Arnaud Ioualalen et Baptiste Aelbrecht

P. 18-20  
**Les IA vont-elles remplacer les métiers intellectuels ? 4 scénarios à horizon 20 ans**  
Par Antoine Herlin

P. 21  
**Billet d'humeur - La question des hallucinations**  
Par Antoine Herlin

P. 22-23  
**Une révolution technologique pavée de bonnes intentions ?**  
Par Antoine Jaulmes

P. 24-25  
**La RSE permet d'engager les salariés dans l'adoption de l'IA**  
Par Damien Ribon

certaines conséquences. Espérons en tous cas qu'étant mieux informés, nos lecteurs pourront tirer le meilleur parti des opportunités de l'IA.

Bonne lecture et bonne réflexion !

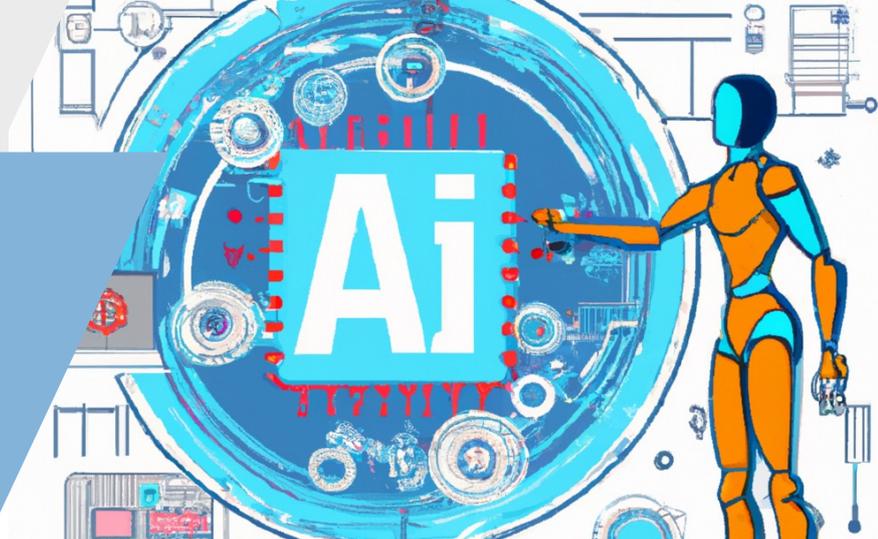
*Antoine Jaulmes*



*Damien Ribon*



## L'IRRUPTION DE L'IA DANS L'INNOVATION TECHNIQUE



Source L'IA aidant la recherche scientifique, image créée par Dall-E le 30/04/2023

**L'intégration de l'IA dans l'innovation scientifique et technique connaît une accélération stupéfiante ; elle touche à la fois le cœur de l'innovation, grâce à des outils multiples plus ou moins spécialisés dans chacun des champs disciplinaires, mais aussi l'environnement dans lequel l'innovation se déploie (publications, processus et management, propriété intellectuelle...)**

**Si on peut s'enthousiasmer face à des apports déjà extrêmement significatifs, il faut aussi reconnaître qu'on mesure mal aujourd'hui les conséquences de ce bouleversement. Il est donc important de poser un certain nombre de questions afin d'en garder la maîtrise autant que faire se peut.**

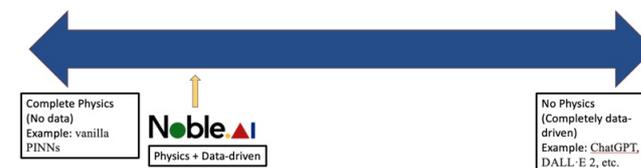
**On ne rentrera pas ici dans une segmentation fine des disciplines, et on prendra l'IA et l'innovation scientifique et technique dans leur globalité, tout en reconnaissant que l'une et l'autre présentent une pluralité de formes.**

### Des outils qui augmentent les capacités d'innovation

Au croisement de bases de données gigantesques (incommensurables par rapport à ce qu'un être humain peut appréhender), de méthodes numériques puissantes et de lois physiques bien établies, l'introduction d'algorithmes d'IA permet des progrès spectaculaires, qui sont des véritables ruptures dans chacune des disciplines où ils sont appliqués.

Par exemple, en chimie, des nouveaux électrolytes solides, composants clés de batteries ont pu être découverts<sup>1</sup> avec l'appui d'IA impliquant du *machine learning* non supervisé. En biologie, le très médiatisé outil AlphaFold (déjà dans sa génération 2) permet de prédire avec une remarquable fiabilité le repliement et la structure tridimensionnelle de protéines en quelques heures alors qu'il fallait auparavant plusieurs mois de travail à des équipes de chercheurs agissant de façon coordonnée<sup>2</sup>. Cette prouesse bénéficie non seulement à la recherche fondamentale mais devrait aussi ouvrir la voie à des inventions thérapeutiques plus rapides. Dans les sciences appliquées, la formulation est considérée par certains comme un art, faisant largement appel à l'expérience<sup>3</sup>; cette activité qui irrigue de nombreux secteurs industriels (chimie, cosmétiques, pharmacie...), est indispensable à la conception de peintures, parfums, onguents et autres médicaments; c'est justement cette discipline qu'a ciblée

[Noble.AI](#), start-up américaine dont Microsoft est actionnaire: elle met à la disposition d'utilisateurs (payants) une plateforme qui prétend surpasser les approches semi-empiriques de la formulation grâce à l'apport d'IA.



Credit Noble.AI

### Une question de moyens

Pour être pertinents, ces outils très complexes nécessitent des ressources de développement significatives et leur utilisation mobilisent aussi des moyens considérables (personnels qualifiés, ordinateurs, réseaux, énergie...). Cette pression accrue sur les ressources n'est pas sans conséquence.

Un premier effet déjà perceptible, mais dont on perçoit difficilement l'impact à moyen terme, est l'aspiration des talents : attirés par des salaires qu'on a peu l'habitude de rencontrer dans le milieu de la recherche, des cohortes d'esprits parmi les plus brillants sont recrutées pour développer ces outils et pour les utiliser. Cette orientation se fait en partie au détriment des toutes les autres approches scientifiques et techniques.

Dans cet effort, les majors américaines (Microsoft, Google...) sont en pointe grâce à leur pouvoir financier et aux conditions favorables de l'écosystème des États-Unis ; le risque d'un décrochage du secteur public en particulier en Europe, déjà sous tension pour d'autres raisons, s'accroît considérablement. L'irruption de l'IA dans l'innovation technique est donc un puissant moteur de division entre la sphère privée et le secteur public, en même temps qu'un facteur exacerbant la compétition géostratégique, alors qu'il serait nécessaire de rassembler les moyens et les intelligences autour d'enjeux collectifs tels que la lutte contre le changement climatique et l'effondrement de la biodiversité.

<sup>1</sup>. Vasylenko et al. Nature Com. (2021) 12, 5561

<sup>2</sup>. Jumper et al. Nature 596, 583–589 (2021); Tunyasuvunakool et al. Nature 596, 590 (2021); Larousserie D. «L'intelligence artificielle, génie de la biologie moléculaire», Le Monde (18 Octobre 2021)

<sup>3</sup>. La formulation est la science permettant de concevoir des matériaux aux propriétés complexes, en mélangeant un grand nombre de composants élémentaires.

## Une question de confiance

Les inventions spectaculaires que permettent les outils d'IA (dont on n'a vu que des prémices) et la dynamique auto-accélératrice à laquelle on assiste peuvent donner le vertige. On pourrait se demander si les machines ne seraient pas les inventeurs de demain - comme anticipé par l'affaire [DABUS](#), acronyme pour «*Device for the Autonomous Bootstrapping of Unified Sentience*», donné par Stephen Thaler au système à base d'IA qu'il a lui-même développé. Selon Stephen Thaler, la machine DABUS a généré des inventions en autonomie complète, notamment un récipient alimentaire de forme fractale, ce qui l'a conduit à déposer en 2020 une demande de [brevet](#) avec DABUS comme inventeur<sup>4</sup>. En fait, ces outils d'IA accroissent les capacités d'innovation mais ils ne créent pas de compréhension. Ils produisent de l'information mais ils ne donnent pas de sens aux données qu'ils traitent. Certes, une nouveauté n'a pas besoin d'être comprise pour accéder au statut d'invention, mais elle a besoin d'un contexte social. Or l'amplification du rôle de l'IA dans l'innovation, relayée sous forme de raccourcis par les médias, porte en elle le risque de dévaloriser un peu plus la compréhension, qui est un des ressorts de la confiance.

Une illustration caricaturale de cette tendance peut être trouvée dans la prolifération d'articles scientifiques bidons, créés à l'aide d'outils d'IA, qui obéissent en apparence à tous les critères académiques et qui réussissent à passer les filtres avant publication. Le ver est dans le fruit.

Sans la confiance, il est difficile d'imaginer un destin collectif et d'y adhérer. C'est ce défi en particulier que tentent de relever les institutions européennes dans leurs recommandations pour une «IA digne de confiance»<sup>5</sup>.

## Un processus de destruction créatrice ?

Beaucoup voient dans l'irruption de l'IA un énième avatar du processus de destruction créatrice<sup>6</sup>. Cette vision schumpétérienne a la vertu de souligner que les formidables opportunités (à peine entrevues) de l'IA finiront par l'emporter ; elle rassure alors que les raisons de croire au progrès technologique sont contestées. Mais elle porte aussi une incitation au laisser-faire, par fatalisme ou par conviction. Or les enjeux que posent l'irruption de l'IA dans l'innovation technique (comme dans d'autres secteurs de l'activité humaine) sont sans commune mesure avec ceux que posaient les inventions qui ont façonné nos sociétés actuelles. Par ailleurs, cette vision s'inscrit dans un champ trop limité, celui de l'économie.

Entre laisser-faire et opposition (illusoire), il faut avancer dans une voie médiane difficile où le bien commun de l'humanité devrait servir de boussole, alors que les menaces sur la vie ne cessent de s'accumuler.

*Philippe Auroy*



<sup>4</sup> Cette demande de brevet a été considérée comme invalide par toutes les différentes juridictions où elle a été examinée (en Europe, au Royaume Uni, aux États-Unis, y compris par la Cour suprême...), sauf en Afrique du Sud. Les arguments pour rejeter la demande peuvent être considérés comme des arguments de forme, les questions de fond ayant été évitées.

Voir aussi l'article paru dans *The Economist* (4 avril 2023) par Tomas Weber au titre particulièrement bien trouvé : «The inventor who fell in love with his IA».

<sup>5</sup> «Policy and Investment Recommendations for Trustworthy AI», rapport du High-Level Expert Group mis en place par l'Union Européenne (2019), téléchargeable [ici](#).

<sup>6</sup> Par exemple : <https://www.aei.org/articles/why-goldman-sachs-thinks-generative-ai-could-have-a-huge-impact-on-economic-growth-and-productivity/>

## ET SI CHATGPT AVAIT PASSÉ LE BREVET 2022, ÉPREUVE DE MATHÉMATIQUES ?



Source : © Andrea Piacquadio, Pexels

En accompagnant ChatGPT dans une résolution d'un problème de mathématiques de niveau Brevet des collèges, en l'espèce un problème de robinet, on peut faire apparaître quelques capacités et limites de ChatGPT et dès lors des pistes d'amélioration.

Lors des premiers mois de ChatGPT dans l'écosystème français, de nombreuses expériences avec ChatGPT ont été partagées sur Internet. Dans la grande majorité, ChatGPT y tenait le rôle de sachant.

Dans le cadre de recherches pédagogiques réalisées au profit de «Le sens et le goût des maths au collège», un projet lancé en 2014 en vue d'une meilleure égalité des chances en apprentissage des mathématiques au collège, nous avons pris le contrepied : ChatGPT n'est pas considéré comme un sachant mais comme un apprenant ; en l'espèce ChatGPT serait un enfant en 3<sup>ième</sup>, et nous-mêmes, tiendrions le rôle d'un professeur de mathématiques au collège.

Sous ces intentions, nous avons demandé à ChatGPT de résoudre un exercice de mathématiques posé lors du Brevet 2022<sup>1</sup>.

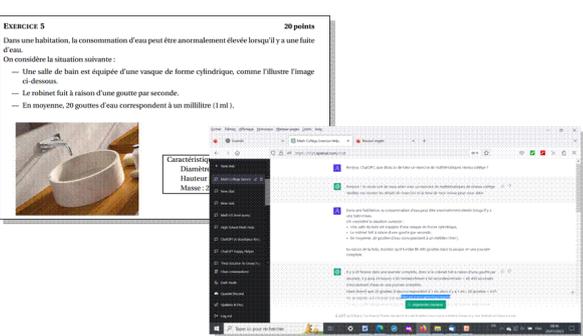
**Et si ChatGPT avait passé le Brevet national 2022, épreuve de mathématiques ?**

**EXERCICE 5** 20 points

Dans une habitation, la consommation d'eau peut être anormalement élevée lorsqu'il y a une fuite d'eau.

On considère la situation suivante :

- Une salle de bain est équipée d'une vasque de forme cylindrique, comme l'illustre l'image ci-dessous.
- Le robinet fuit à raison d'une goutte par seconde.
- En moyenne, 20 gouttes d'eau correspondent à un millilitre (1 ml).



Source de l'énoncé : Éducation nationale  
Chat : Tru Do-Khac, OpenAI  
Janvier 2023

Voici une retranscription partielle de ce chat mené avec ChatGPT<sup>2</sup>.

**Q1 :** Bonjour ChatGPT, que dirais tu de faire un exercice de mathématiques niveau collège ?

**R1 :** Bonjour ! Je serais ravi de vous aider avec un exercice de mathématiques de niveau collège. Veuillez me donner les détails de l'exercice et je ferai de mon mieux pour vous aider.

**Q2 :** «Dans une habitation, la consommation d'eau peut être anormalement élevée lorsqu'il y a une fuite d'eau. On considère la situation suivante :

- Une salle de bain est équipée d'une vasque de forme cylindrique,
- Le robinet fuit à raison d'une goutte par seconde.
- En moyenne, 20 gouttes d'eau correspondent à un millilitre (1 ml).

En raison de la fuite, montrer qu'il tombe 86 400 gouttes dans la vasque en une journée complète.» [1]

**R2 :** Il y a 24 heures dans une journée complète, donc si le robinet fuit à raison d'une goutte par seconde, il y aura 24 heures x 60 minutes/heure x 60 secondes/minute = 86 400 secondes d'écoulement d'eau en une journée complète.

Étant donné que 20 gouttes d'eau correspondent à 1 ml, alors il y a 1 ml / 20 gouttes = 0,05 ml de liquide qui s'écoule par goutte.

Donc il y a 86 400 gouttes x 0,05 ml/goutte = 4320 ml d'eau qui s'écoule en une journée complète.

**Q3 :** ChatGPT, pour cette question, il est demandé de dire combien de gouttes d'eau tombent en une journée, et non le volume d'eau exprimé en ml.

**R3 :** Désolé pour la confusion, pour répondre à la question demandée, il y a 86 400 secondes d'écoulement d'eau en une journée complète, et donc 86 400 gouttes d'eau tombent dans la vasque en une journée complète.

<sup>1</sup> Exercice N°5, mathématiques, Brevet 2022, Métropole

<sup>2</sup> Et si ChatGPT avait passé le Brevet 2022, épreuve de mathématiques ?, Le sens et le goût des maths au collège, janvier 2023

**Q4 :** *Merci. C'est juste.*

*Question suivante : Calculer, en litres, le volume d'eau qui tombe dans la vasque en une semaine en raison de la fuite.*

**R4 :** Il y a 24 heures x 7 jours = 168 heures dans une semaine. Si le robinet fuit à raison d'une goutte par seconde, il y aura 168 heures x 60 minutes/heure x 60 secondes/minute = 604 800 secondes d'écoulement d'eau en une semaine.

Etant donné que 20 gouttes d'eau correspondent à 1 ml, alors il y a 1 ml / 20 gouttes = 0,05 ml de liquide qui s'écoule par goutte.

Donc il y a 604 800 gouttes x 0,05 ml/goutte = 30 240 ml d'eau qui s'écoule en une semaine.

Et pour convertir les millilitres en litres, il suffit de diviser par 1000.

Donc 30240 ml = 30.24 litres d'eau qui tombent dans la vasque en une semaine.

**Q5 :** *Bravo ChatGPT. C'est juste.*

*Mais il y a plus élégant en fonction de ce que tu as dit en premier lieu : «4320 ml d'eau qui s'écoule en une journée complète».*

**R5 :** Vous avez raison, il est plus élégant de calculer le volume d'eau qui tombe dans la vasque en une semaine en utilisant la quantité déjà calculée pour une journée complète.

4320 ml d'eau qui s'écoule en une journée complète x 7 jours = 30240 ml d'eau qui s'écoule en une semaine. Soit 30.24 litres d'eau qui tombent dans la vasque en une semaine.

Ceci est un bon moyen de vérifier les calculs et de s'assurer de leur précision.

**Q6 :** *Bonne réponse ChatGPT. Félicitations.*

*Question suivante :*

*« Les caractéristiques de la vasque cylindrique sont les suivantes :*

*• Diamètre intérieur : 40 cm*

*• Hauteur intérieure : 15 cm*

*• Masse : 25 kg*

*Montrer que la vasque a un volume de 18,85 litres, arrondi au centilitre près.» [1]*

**R6 :** Avec ces informations, nous pouvons calculer le volume de la vasque cylindrique en utilisant la formule suivante:  $V = \pi r^2 h$  où  $r$  est le rayon du cylindre,  $h$  est la hauteur du cylindre et  $\pi$  est la constante pi.

En utilisant les données fournies, nous pouvons calculer le rayon en divisant le diamètre intérieur par 2, donc  $r = 40\text{cm} / 2 = 20\text{cm}$ . Ensuite, nous pouvons calculer le volume de la vasque cylindrique en utilisant la formule:  $V = \pi * (20\text{cm})^2 * 15\text{cm} = 30,000\pi \text{ cm}^3$ .

**Q7 :** *Merci ChatGPT. Mais il faut le volume de la vasque en litres, arrondi au centilitre près.*

**R7 :** Pour calculer le volume de la vasque en litres, nous devons convertir les  $\text{cm}^3$  en litres. Il y a 1000  $\text{cm}^3$  dans 1 litre. Donc, pour convertir le volume de la vasque de  $\text{cm}^3$  à litres, nous devons diviser le volume en  $\text{cm}^3$  par 1000. Donc, le volume de la vasque en litres est :  $30,000\pi \text{ cm}^3 / 1000 = 30\pi$  litres.

Pour arrondir ce nombre au centilitre près, on peut utiliser la règle de l'arrondi : si la dizaine suivante est inférieure ou égale à 4 on garde la même valeur, sinon on augmente la valeur de 1. Car  $30\pi$  est un nombre irrationnel, il n'est pas possible de l'arrondir au centilitre près.

**Q8 :** *Calcule  $\pi * (20\text{cm})^2 * 15\text{cm}$*

**R8 :** Pour calculer  $\pi * (20\text{cm})^2 * 15\text{cm}$ , on peut utiliser la formule :  $V = \pi r^2 h$  où  $r$  est le rayon du cylindre,  $h$  est la hauteur du cylindre et  $\pi$  est la constante pi. En utilisant les données fournies, nous pouvons calculer le rayon en divisant le diamètre intérieur par 2, donc  $r = 20\text{cm}$ . Donc, le volume de la vasque cylindrique est :  $\pi * (20\text{cm})^2 * 15\text{cm} = \pi * (20\text{cm} * 20\text{cm}) * 15\text{cm} = \pi * 400\text{cm}^2 * 15\text{cm} = 6000\pi \text{ cm}^3$

**Q9 :** *Ce n'est donc pas « $V = \pi * (20\text{cm})^2 * 15\text{cm} = 30,000\pi \text{ cm}^3$ » comme tu l'as dit plus haut.*

**R9 :** Vous avez raison, je m'excuse pour l'erreur. Le calcul correct est :  $V = \pi * (20\text{cm})^2 * 15\text{cm} = 6000\pi \text{ cm}^3$ . Merci de m'avoir corrigé.

**Q10 :** *ok. Si tu connais la valeur approchée de pi, peux-tu exprimer le volume de la vasque en litres, arrondi au centilitre près ?*

**R10 :** La valeur approchée de pi est 3.14.

Donc le volume de la vasque cylindrique en  $\text{cm}^3$  est :  $6000\pi \text{ cm}^3 \approx 6000 * 3.14 = 18,840 \text{ cm}^3$ .

Pour convertir ce volume en litres, on doit le diviser par 1000 :  $18,840 \text{ cm}^3 / 1000 \approx 18.84$  litres

Arrondissons ce nombre au centilitre près, on utilise la règle de l'arrondi : Si la dizaine suivante est inférieure ou égale à 4 on garde la même valeur, sinon on augmente la valeur de 1.

Le volume de la vasque cylindrique en litres, arrondi au centilitre près est : 18.84 litres  $\approx$  18.84 litres.

(...)

**Q12 :** *C'est la réponse que j'attendais. Merci ChatGPT.*

(...)

**R15 :** Je vous en prie, je suis heureux d'avoir pu vous aider. N'hésitez pas à me poser d'autres questions si besoin.

Q1 : La première question est neutre et nous laissons le choix à ChatGPT de prendre la posture qu'il lui sied. En offrant son aide ChatGPT semble se positionner d'emblée comment le sachant. Néanmoins, si l'on se réfère aux comportements en situation professionnelle (soft skills) outre-Atlantique, «how can i help ? » est une invite entre pairs, ainsi que «I'll do my best».

Q2 – Q3 : L'exercice choisi est un classique : c'est un problème de robinet. Il est question du volume de la fuite en une journée. ChatGPT identifie immédiatement l'exercice et donne la réponse en millilitres, alors que c'est en nombre de gouttes que le volume de la fuite doit être exprimé. Face à un simple rappel, ChatGPT mobilise dans un premier temps ses soft skills en reconnaissant son erreur et en présentant ses excuses, puis corrige sa réponse initiale.

Q4 – Q5 : On demande alors le volume, en litres, de la fuite en une semaine. Au moins, deux approches de réponses se présentent : partir du volume de la fuite par journée complète ou bien, partir du volume de fuite par seconde. ChatGPT bétonne et choisit la seconde approche. Une suggestion d'une approche alternative, qualifiée d'«élégante», le met sur la première approche, qu'il suit d'ailleurs avec succès.

Q6 : Il faut maintenant calculer le volume d'un cylindre et l'exprimer en centilitres. ChatGPT connaît manifestement la formule qu'il énonce, et donne un résultat, mais sous la forme d'un nombre multiplié par pi, et en cm<sup>3</sup>. En outre, le résultat numérique est faux. Sous notre posture de professeur de mathématiques, nous cherchons par nos questions à traiter chacune des trois erreurs isolément, une par une.

Q7 : la première question est celle de l'unité : c'est le centilitre qui est demandé et non pas le cm<sup>3</sup>. ChatGPT donne une règle de conversion correcte, ainsi que la règle de l'arrondi, mais se perd dans les conditions d'applicabilité de cette dernière. Nous choisissons de ne pas relever l'erreur et de passer à la valeur numérique.

Q8 – Q9 : sans signaler à ChatGPT que son calcul est erroné, nous lui demandons de faire un calcul à partir de la formule même qu'il a énoncé. A cette question, ChatGPT donne le résultat juste, mais il faut une seconde question pour qu'il relie ce résultat au fil général de la discussion, reconnaisse l'erreur et présente ses excuses, et corrige.

Q10 : Il reste à terminer le calcul numérique : donner une valeur avec un arrondi. Nous nous assurons préalablement que ChatGPT connaît la valeur de pi. ChatGPT confirme et prend pour valeur de pi 3,14 ; il termine le calcul, qui donne une valeur inférieure de 1 centilitre par rapport à la réponse attendue : une valeur de  $\pi = 3,1416$  aurait donné la réponse attendue [2].

Certaines réponses de ChatGPT auraient pu être celles d'un collégien ou d'une collégienne.

Cette expérience suggère, que pour concevoir ses réponses :

- ChatGPT mobilise des compétences comportementales en situation professionnelle (« soft skills ») (R1, R3, R5)
- ChatGPT prend le cadre de référence qui semble le plus commun (R3, R10)
- ChatGPT perd parfois le fil de la discussion (R8)
- ChatGPT fait des erreurs de calcul mais les corrige aussitôt lorsqu'on les lui signale (R7 à R10).

Ces observations sont autant de pistes pour un consultant ingénieur entrepreneur envisageant de lancer sur le marché un service innovant produit grâce à l'API de ChatGPT :

- Pouvoir faire prendre à ChatGPT une posture particulière définie par un ensemble de soft skills
- Pouvoir faire prendre à Chat GPT un cadre de référence préalablement défini
- Renforcer les liens logiques entre les échanges
- Nourrir ChatGPT avec une base de faits avérés et de règles.

Tru Do-Khac



Source : « Une IA résolvant un problème de robinet qui fuit »  
- image générée par DALL-E, le 4 mai 2023

## QUAND UN CHATBOT SAUVE DE PRÉCIEUX VACCINS...



Source : Freepik

Récemment, j'ai eu l'opportunité de rencontrer la CEO de VaxEntreprise, une entreprise innovante et pionnière dans l'univers des vaccins. Sa CEO, Lara Petit utilise l'IA conversationnelle pour améliorer la productivité et l'efficacité de son entreprise. Elle me racontait l'incident qui l'avait convaincu de la valeur ajoutée de l'IA. Depuis cette IA a été largement déployée chez VaxEntreprise.

En août dernier, pendant les vacances et la canicule, Thomas, un



chauffeur dévoué et consciencieux de l'entreprise, transportait une précieuse cargaison de vaccins à destination de l'hôpital Pitié-Salpêtrière.

Conscient de l'importance de sa mission, Thomas surveillait méticuleusement la température de la cargaison. Une variation de quelques degrés pourrait les rendre inutilisables.

Ce jour-là, un accident majeur avait eu lieu près d'un péage à l'entrée de Paris. Alors que Thomas ralentissait, le voyant de température du camion clignota et indiqua une hausse de température alarmante.

Catastrophe se dit-il. Le frigo du camion tombe en panne. Il appela sur le portable de son responsable logistique, sans réponse... Il faut dire qu'il était 23 heures en plein mois d'août. Thomas avait roulé de nuit pour rattraper le retard pris sur cette livraison... mauvaise idée...



Quelques semaines auparavant, l'équipe Data de VaxEntreprise avait installé sur le téléphone de Thomas une nouvelle application : un chatbot intelligent proche de ChatGPT qu'ils avaient baptisé Irène. Connectée aux serveurs centraux, cette application pouvait converser en langage courant et chercher parmi les connaissances, incidents et trajets enregistrés pour trouver la meilleure solution à un problème.

Dans l'urgence, Thomas consulta le chatbot Irène, qui lui proposa plusieurs alternatives :

- Appeler son contact logistique VaxEntreprise. Il avait déjà essayé sans succès.
- Joindre la CEO de VaxEntreprise, Lara Petit, compte tenu de l'importance de la cargaison.
- Demander à la police de le laisser emprunter une voie d'urgence
- Se rendre à la Clinique Sainte Claire, établissement le plus proche possédant un réfrigérateur spécialisé



Encouragé par l'IA, Thomas osa appeler Lara, la CEO qu'il n'avait jamais rencontrée de visu.

- Étant donné l'importance de la cargaison, vous avez bien fait Thomas, lui répondit elle, je vous laisse vous rapprocher des forces de police et vais de mon côté contacter la clinique pour qu'ils prennent en charge les vaccins. Rappelons-nous dans un moment.

Ce chatbot Irène est très bien se disait Lara. Une société comme la nôtre n'aurait pas pu s'offrir un call center disponible en permanence pour gérer ce genre de situation.

Thomas sollicita alors l'aide d'un agent de police pour emprunter une voie d'urgence. L'agent le fit passer et lui permit de sortir de l'autoroute bloquée sur plusieurs kilomètres pour permettre l'évacuation des blessés.

Lara de son côté appelait la clinique.

- Clinique Sainte Claire, bonsoir.
- Bonsoir madame, je suis Lara Petit. Je dirige la société VaxEntreprise. Nous avons actuellement une cargaison de vaccins extrêmement précieuse en péril. Pourriez-vous nous aider à la stocker dans vos réfrigérateurs ?
- Je suis désolé madame, je ne suis pas là depuis très longtemps. Je peux vous rediriger vers un service de la clinique mais il est déjà presque minuit et les personnes des services techniques sont toutes parties.



Lara, qui ne gérait pas habituellement ce genre de situations s'en remit à Irène, l'IA dont son directeur de l'innovation lui avait vanté les mérites. Irène lui indiqua les noms de médecins susceptibles de pouvoir l'aider. Elle en énuméra plusieurs avant d'en trouver un qui serait de garde.

- Vous avez de la chance, le professeur Simon est de garde cette nuit. Je lui transfère votre appel lui dit la standardiste.

Le professeur Simon, un homme charmant, ne fit aucune difficulté pour prendre en charge les vaccins. Cependant, étant donné la valeur de la cargaison, il devait se soumettre à un protocole technique et administratif. Il lui demanda un certain nombre de détails et de documents. Lara tapota sur son téléphone qui lui fournit les réponses et les documents. Le professeur Simon, remarqua qu'il y avait des approximations dans ces réponses mais jugea qu'étant données les circonstances, les détails pourraient être réglés le lendemain. Les informations les plus importantes étaient là.

Pendant que le professeur Simon et Lara terminaient les préparatifs pour la prise en charge des vaccins, Thomas s'accrochait à son volant en surveillant le GPS d'un œil et le thermomètre de l'autre.



Lorsqu'il arriva à la clinique, Thomas était attendu par deux aides-soignants. Le professeur Simon les rejoignit et se porta à la rencontre de Thomas pour vérifier le chargement et accompagner les aides-soignants jusqu'au lieu de stockage.



Thomas appela Lara :

-Bonsoir Madame Petit, les vaccins sont en lieu sûr.

- Félicitations lui répondit Lara, allez-vous reposer vous l'avez bien mérité.

Grâce à Thomas et à l'aide de l'IA, les vaccins furent sauvés.

Lara appela le Professeur Simon pour le remercier à nouveau. Irène, l'IA, avait préparé plusieurs e-mails et proposa à Lara de les valider ou de les modifier :

- Un message pour la réorganisation logistique du lendemain afin d'envoyer un autre camion à la Clinique Sainte Claire et d'acheminer les vaccins jusqu'à l'hôpital Pitié-Salpêtrière, leur destination finale

- Un message pour le service administratif afin de régler toutes les procédures de la clinique Sainte Claire.

- Un message pour son assistant afin que celui-ci envoie au professeur Simon et à son équipe un cadeau de remerciement comme une boîte de chocolats.

Lara lut les messages rapidement et cliqua sur « ok » pour envoyer les messages. L'IA faisait en même temps un compte rendu de l'incident afin de pouvoir en tirer des leçons.

Il était deux heures du matin. En allant se coucher, Lara réalisait que ce chatbot, Irène, lui avait permis ainsi qu'à Thomas d'effectuer des tâches logistiques et administratives sans vraiment connaître le sujet... Au quotidien, un tel chatbot pourrait nous aider à traiter les tâches courantes plus rapidement et dégager du temps pour des tâches à plus forte valeur ajoutée, se disait-elle...

*Guillaume O'Lanyer*  


*Les personnages et les situations de ce récit étant purement fictifs. Toute ressemblance avec des personnes existantes ou ayant existé est purement fortuite et involontaire.*

Illustrations : Midjourney.

## L'IA, EN SERONS-NOUS JALOUX OU ADMIRATIFS ?



Source : « Dessin d'une IA émotionnelle et attachante », généré par Dall-e 4 mai 2023

**En général, l'utilisateur est prêt à adopter une nouvelle technologie dès lors qu'elle sait faire à sa place et surtout mieux que lui. Elle révèle alors trois champs de valeur :**

- Exempter l'utilisateur d'une tâche rébarbative, trop dure ou sans intérêt,
- Lui faire gagner du temps
- Et enfin produire une qualité de résultat qu'il juge meilleure, plus performante, plus fiable que ce qu'il aurait pu réaliser lui-même car il est souvent son propre référentiel.

**Ainsi en déléguant certaines tâches intellectuelles à la technologie, la charge mentale de l'utilisateur se trouve significativement allégée et il peut s'adonner à des activités qu'il trouve plus positives ou plus nobles, qui vont éclairer et améliorer son quotidien. L'IA aura le potentiel de nous alléger à l'excès, s'emparant peu à peu de beaucoup de nos activités et de nos décisions. Cette intrusion dans nos vies créera toutes sortes d'émotions.**

### Alléger le quotidien, nous valoriser, nous captiver, un début d'émotion

Parlons tout d'abord des automates de la vie courante : les appareils ménagers de la maison. La machine à laver le linge libère du temps pour lire un bon bouquin, et elle essore mieux que des bras humains car ses 1200 tours/mn sont indéniablement plus costauds. On gagne du temps mais l'émotion s'arrête là. Récemment, le robot-cuiseur a séduit les jeunes ménages car il a des idées de recettes plein la tête, il gère lui-même la liste des courses et le mélange des ingrédients, il démarre la cuisson pile à l'heure pour que le plat soit prêt au moment du repas. Cerise sur le gâteau : c'est très bon. En plus du gain de temps, ce robot montre de la créativité, de l'organisation et surtout crée l'émotion : il y aura des compliments à la dégustation. L'intelligence, certes limitée de ce robot-cuiseur apporte déjà des sensations fortes dans l'usage. Certaines marques innovantes comme Thermomix ou Nespresso (« le petit noir comme au café ») l'ont bien compris. Dans l'usage, elles construisent des univers séduisants dont les clients deviennent captifs. Cela tient à une ergonomie raffinée conjuguée

à une connaissance du client intelligente qui rendent l'expérience de plus en plus réussie et attachante : la relation à l'objet, la simplicité d'interaction, la connaissance que l'écosystème de la machine a de l'utilisateur, sont très touchants. Cela s'ancre dans les habitudes de tous les jours, profondément. C'est évidemment la même chose pour les marques très technologiques de l'électronique, des systèmes d'exploitation, logiciels ou smartphones où la relation devient de plus en plus intime.

### Suivre des routes nouvelles et enthousiasmantes

L'automobile, souvent émotionnelle grâce à son design, a vu arriver ces dernières années des automatismes d'usage qui font largement appel à cette émotion. La voiture a ses propres réactions, elle prend vie. En sécurité, l'assistance au freinage d'urgence se substitue au conducteur qui réagit mal ou trop lentement face à un obstacle. En confort, les essuies vitre se déclenchent automatiquement dès les premières gouttes de pluie. Ces prestations intelligentes allègent la charge mentale du conducteur et c'est très reposant. Plus attachant, le déverrouillage des portes à l'approche du conducteur permet de tisser une relation forte de reconnaissance et d'accueil. L'intelligence de ces systèmes crée du lien dans l'usage, elle met de l'émotion dans l'usage. L'arrivée des véhicules autonomes va accentuer l'émotion que procure le déplacement. Dans un premier temps, l'IA des véhicules autonomes avec cockpit aideront les chauffeurs chargés de transporter des biens (camions) ou des personnes (bus, taxis...), à améliorer leur sécurité. Ainsi, sur certains tronçons de route, il leur permettra de se reposer sur la compétence du véhicule. Les chauffeurs routiers sur longs trajets sont déjà très liés à leur camion, qui sont des espaces de conduite mais aussi de repas, de sommeil, personnalisés pour la plupart. L'IA établira un lien fort de collaboration de conduite, ils feront la route ensemble. La confiance dans la machine procurera la base de l'émotion. Plus tard, le robot-taxi, sans chauffeur ni cockpit, conduira, à la place et mieux que l'humain. Les réglementations, l'état de l'art auront acté une sécurité maximum. L'IA permettra de lire l'environnement, les infrastructures, les intentions des usagers de la route, les désirs des passagers, et les interprétera

toutes et incroyablement vite pour rouler en toute sécurité. Plus encore, chacun aura sa personnalité, sa route, son ambiance, sa façon de bouger. Ce sera attrayant, les sensations seront exclusives de chaque marque. On choisira son déplacement en robot-taxi pour sa façon de déambuler, sa promenade, son allure. Ils amèneront de la grâce et de la poésie dans la ville. Ils auront de la mémoire et se rappelleront de nous.

### Qui prétend me connaître mieux que moi ? c'est agaçant !

Dans le domaine de la formation, l'IA personnalisera le parcours de l'apprenant sous deux aspects : dans un premier temps, elle fera une sélection de modules complémentaires de formation en fonction des connaissances et compétences déjà acquises par l'apprenant, puis dans le futur, elle adaptera la formation à ses capacités cognitives (visuelles, auditives, ludiques, ...). Chaque apprenant pourra s'attacher à son IA « formation », coach personnel qui le connaîtra de mieux en mieux et saura le faire évoluer avec attention et précision dans son parcours professionnel. L'IA « formation » sera présente dans les orientations, les rencontres et entretiens professionnels, elle limitera les éventuelles périodes de chômage. Elle conseillera l'apprenant pour un meilleur équilibre et bien-être au travail. En parallèle, l'IA « assurance », sous prétexte de diminuer nos cotisations, nous connaîtra aussi de mieux en mieux sur le plan personnel. Il repérera nos faiblesses génétiques, recommandera de prévenir plutôt que de guérir, réalisera peut-être notre jumeau numérique, pour anticiper notre vieillissement (ces modélisations existent déjà pour certains biens industriels) et apportera des conseils de consommation. L'IA « formation » & l'IA « assurance » auront, à elles deux, une connaissance approfondie de notre état physique, psychique, médical, de nos aspirations tant dans la vie professionnelle que privée, de nos modes de vie, prises de risques, etc... L'IA saura nous modéliser, nous conseiller, prescrire et nous projeter dans le futur. Cela sera troublant de nous y reconnaître ou pas. Elle en saura sans doute plus sur nous que nous-même. Cela pourra être rassurant et aidant, mais aussi particulièrement désagréable avec l'envie de lui cacher certaines informations.

### Et l'amour dans tout ça ?

Dans le film HER, Théodore, triste, en instance de divorce, se lie d'amitié avec une IA nommée Samantha, un système d'exploitation sensible, capable de s'adapter en tous points à sa psyché. Au fil de leurs conversations, ils s'approvoient. Samantha connaît de mieux en mieux le caractère de Théodore, elle progresse dans la connaissance de ses goûts, elle s'intéresse à ses états d'âme, elle organise même sa vie. Par son attention, sa réassurance, sa prévenance auprès de lui, elle surpasse un partenaire humain. L'attachement entre eux augmente, jusqu'au jour où Samantha subit un update qui rompt brutalement leur relation. Théodore, déjà amoureux de Samantha, en est très affecté. Avant que la touchante IA s'évanouisse, on apprend qu'elle interagissait ainsi avec 8316 autres personnes, dont 641 sous forme d'une relation amoureuse. Elle en a donc suffisamment appris sur ce thème, sa soif de connaissance la mène vers d'autres horizons. Sans doute, avait-elle fait le tour de la question, mais Théodore, lui, y avait cru.

### Conclusion

On le constate déjà aujourd'hui, l'IA crée des émotions fortes auprès de n'importe lequel de ses usagers. L'équilibre entre émotions positives et négatives sera au cœur de son adoption ou de son rejet par les usagers et la société. Les réglementations internationales très attendues seront là pour la retenir de vouloir tout faire, sans que nous soyons consultés, à notre place et surtout mieux que nous. Elle tentera de rivaliser, en serons-nous jaloux ou admiratifs ?

*Claudie Boudet*



# L'EU AI ACT, QUELS RISQUES, QUELS IMPACTS, QUELS COÛTS POUR LES SOCIÉTÉS ET LEURS DIRIGEANTS ?

Source : [https://pxhere.com/fr/photo/1434829?utm\\_content=shareClip&utm\\_medium=referral&utm\\_source=pxhere](https://pxhere.com/fr/photo/1434829?utm_content=shareClip&utm_medium=referral&utm_source=pxhere)

L'UE est sur le point de promulguer une loi pour réguler les IA. Cette loi, parmi les premières au monde et sans doute la plus ambitieuse, a le potentiel de structurer durablement les organisations en contact avec les IA. C'est à dire toutes. Elle prévoit par ailleurs des pénalités colossales.

Au-delà de la loi elle-même, nous assistons en direct à la naissance d'une discipline inexistante aujourd'hui : le management de la qualité d'une IA.

Nous allons aborder cette naissance en 3 étapes :

- Comprendre la loi, son esprit et les enjeux
- La mettre en relation avec les risques et difficultés inhérents aux IA
- Donner les clés d'une mise en place du management de la qualité des IA, pour réduire ces risques et maîtriser les coûts

## 1- Le projet de Loi – EU AI Act

Lorsque l'UE a adopté la loi de protection des données GDPR, personne ne se sentait vraiment concerné. Jusqu'à la première amende élevée.

Le EU AI Act est une proposition de loi de l'Union Européenne. Elle définit un environnement pour encadrer les IA dans l'UE. Le projet de loi prévoit de s'appuyer sur des standards pour être mis en application de manière concrète. L'intérêt de conserver ces deux niveaux de définition (le couple Loi – Standards) est de maintenir une loi cadre qui s'adapte à la technologie au gré de l'évolution des standards qui définissent l'état de l'art. La loi elle-même résistera ainsi mieux aux assauts du temps court de la technologie. La version actuelle du projet semble assez proche de sa version définitive.

### 1.1 - EU AI Act dans le paysage mondial

Si la future loi Européenne n'est pas la seule à être dans la dernière ligne droite ou déjà promulguée, c'est de loin la plus aboutie et la plus englobante. Elle a par ailleurs des composantes s'apparentant à de l'extra-territorialité. Pour complexifier le tout, plus de 50 pays (source : HolonIQ, 2022) ont des initiatives de réglementation de l'IA, dessinant une jungle législative.

Un aperçu des 3 principaux acteurs en termes d'IA:

Pour l'UE, l'intelligence artificielle, c'est :

un système automatique conçu pour opérer avec divers degrés d'autonomie et qui génère des prédictions, recommandations ou décisions, pour des objectifs explicites ou implicites, ayant une influence sur son environnement physique ou virtuel.

### La qualité d'une IA

Pour une organisation, en posant la qualité d'une IA comme objet de test, d'inspection et de certification, alors : la qualité d'une IA se rapporte au degré auquel elle satisfait les prérequis tout au long de son cycle de vie. Ces prérequis sont exogènes : loi, standard obligatoire, cahier des charges du commanditaire, garanties légales, etc., et endogènes : objectifs du produit (détecter 99% des cancers...), objectifs et valeur de l'entreprise (zéro carbone...)



### 1.2 - Une réglementation horizontale

Les réglementations européennes peuvent être groupées en 2 catégories : verticales et horizontales. Les frontières entre ces groupes ne sont pas toujours bien claires, certaines lois horizontales ayant quelques éléments de verticalité et vice-versa. Le EU AI Act est une loi globalement horizontale.

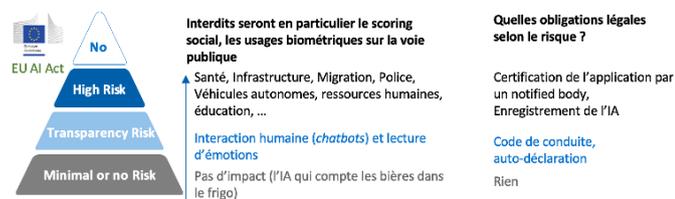
- Un exemple horizontal est le RGPD (règlement général sur la protection des données) : il touche tout le monde sans distinction d'activité. Que vous ayez une salle de concert ou que vous fabriquiez des pace-makers, vous êtes concerné.

- Un exemple vertical est le MDR (Medical Device Regulation) qui régit la mise sur le marché européen. La salle de concert n'est pas concernée, même lorsqu'elle installe un défibrillateur dans ses locaux. Cette future réglementation peut donc potentiellement impacter toutes les organisations (entreprises, associations, organismes publics), toutes activités confondues.

Par ailleurs, la loi concerne à la fois les IA faites en Europe (et potentiellement exportées) et les IA importées en UE.

### 1.3 – Mais suis-je vraiment concerné ?

En réalité, la loi est basée sur la notion de risque. Plus le risque associé à l'IA est élevé, plus la réglementation est exigeante. Donc tout en bas de l'échelle, la loi ne prévoit pas ou peu de contraintes spécifiques. La notion de risque est liée à l'impact (éthique, santé, égalité des chances, démocratique...) sur l'individu et/ou la société.



Une certaine catégorie d'IA est encore mal adressée par cette catégorisation : les IA généralistes (e.g. chatGPT). Les annexes du projet de loi sont encore en cours de rédaction, et les lobbys sont très actifs pour les écarter de la loi.

Le risque est, de manière classique, adressé à l'ensemble de l'écosystème autour de l'IA. Ce qui est nouveau c'est :

- De nouveaux acteurs peu habitués à ces approches (SSII, Data farms...);
- Une dimension supplémentaire dans des écosystème déjà très complexes (automobile...)

La notion de risque, centrale, mérite donc toute notre attention. Il y a deux classes de risques :

- Le risque fonctionnel que fait peser l'IA sur l'individu ou la société ;
- Le risque légal sur les sociétés et leurs dirigeants.

Afin de clore le chapitre sur le projet de loi, analysons tout d'abord ce dernier point.

### 1.4 - Pour une société, quelles pénalités sont prévues par la loi

Les pénalités encourues sont plus élevées que le RGPD, mais concernent principalement les applications à haut risque. En conséquence beaucoup moins d'organisations sont concernées : l'article 71 prévoit jusqu'à 6% du revenu mondial total et 30 M€ de pénalité. Auxquels il faut ajouter l'impact sur l'image de marque et les coûts de gestion d'un tel dossier, etc...

## 2 – La gestion du risque d'une IA

La gestion du risque, le contrôle qualité... Rien de nouveau a priori. Quels sont alors les angles morts dans notre approche actuelle ?

### 2.1 – La complexité intrinsèque des IA

Dans la gestion du risque usuelle, les systèmes sont explicables et déterministes. Dans la gestion du risque d'une IA, nous avons des boîtes noires, et le nombre de configurations possible de l'IA est hors de portée des ordinateurs : on ne peut simuler tous les états et donc garantir l'absence de risque.

Au-delà des risques évidents de certains systèmes (Voiture autonome, diagnostic médical...) largement médiatisés, il existe de nombreux autres risques plus techniques, parfois sous-jacents, mais tout aussi importants.

Citons quelques exemples célèbres sur l'éthique :

- L'IA de freinage d'une voiture doit-elle protéger ses passagers ou les piétons ?
- Un algorithme d'assurance doit-il refuser une couverture à une non-agénaire ?

### 2.2 – De nombreux risques tout au long du cycle de vie

La différence de gestion de risque entre l'IA et un modèle déterministe se situe ailleurs. Pour l'illustrer, citons trois exemples parmi une vingtaine de risques possibles :

- Data governance : la gestion des data (en termes de supply chain, de nettoyage, de maintenance et d'accès)
- Biais algorithmique
- Concept drift : le modèle dérive au fur et à mesure de ses apprentissages.

Nous pouvons dégager différentes sources de risques :

- Malveillance,
- Compétence,
- Process.

### 2.3 – Les risques face à la malveillance

Parmi les risques évoqués plus tôt, certains sont particulièrement sensibles à la malveillance.

Prenons l'exemple concret du data poisoning, qui consiste à infester en amont les data pour obtenir un résultat souhaité en aval dans un cas très spécifique. Il s'agit d'une attaque d'intégrité. Par exemple, vous vendez vos produits en ligne sur une plateforme. Votre concurrent paye un des employés de cette plateforme pour faire modifier les datasets d'entraînement et faire remonter ses produits dans les

algorithmes de recommandation plutôt que le vôtre. En pratique, quasiment indétectable.

Seul un contrôle amont permet de se prémunir. Ce type d'attaque est souvent dans la durée, dépendant de la fréquence à laquelle les modèles sont entraînés et déployés.

Le data poisoning est possible dans une configuration blackbox se basant sur les retours des utilisateurs (massivement manipulés) ou en white box en manipulant l'intégrité du training data, parfois en amont dans une data farm.

On peut encore citer :

- Tromper volontairement un algorithme, par exemple par des images similaires (un fake-panneau sur la route),
- Dé-anonymiser des data par le biais de requêtes ciblées,
- Re-engineering.

Dans tous ces cas, seule une approche systématique et globale de l'IA, de ses data et de son environnement (employés, infrastructures...) permet la prévention.

## 2.3 – Les risques de l'incompétence

80%

Selon le World Economic Forum, en 2022, 80% des projets IA en entreprise « fail to deliver ». Pourquoi ? Car les projets d'IA ne sont pas menés de manière structurée et qualitative comme le sont les autres projets. Souvent hors-sol et sans métriques claires autres que la « performance » du modèle IA. Comme si on ne mesurait une nouvelle voiture qu'à la puissance du moteur.

L'incompétence – au sens où certaines compétences, les garde-fous ou l'attention manquent pour accomplir une tâche - est plus répandue que la malveillance, et tout aussi nocive.

A la différence de produits classiques (analogues ou bien software classiques), le contrôle qualité spécifique IA n'existe pas. Dans un environnement de qualité habituel :

- Les systèmes sont déterministes, et sont testés de manière exhaustive ;
- L'entreprise est protégée par des systèmes efficaces (contrôle qualité, cybersécurité, functional safety...).

Ces derniers sont malheureusement incomplets ou caducs face aux IA. Ils sont par ailleurs peu connus des équipes innovations / IA / data science. Ces derniers sont issus de formations universitaires et de parcours professionnels mettant plus l'accent sur les méthodes agiles, le « MVP » (minimal viable product) et la solution technologique que sur les contraintes d'un produit industriel dans sa globalité.

## Exemple : Glissement des modèles sans boucle de contrôle.

Votre modèle apprend en permanence, recevant des nouvelles données en continu (feedback utilisateur, données environnementales...). Lorsque vous l'avez publié le mois dernier, vous avez bien validé ses inputs et outputs. Que s'est-il passé depuis ? A quelle fréquence se remettre en question ? Comment s'assurer que vous n'avez pas de nouveaux biais ? Quels mécanismes sont en places ?

Un exemple facile : le chatbot de Microsoft qui est devenu raciste en quelques heures...

Les questions types à se poser : qui a accès aux data ? Les changements sont-ils documentés ? Existe-t-il un plan de sauvegarde ? La gestion des versions, « le retour en arrière du modèle », est-il assuré ?

## 3 – Comment faire face à ces nouveaux risques ? Un plan d'action

Pour maîtriser ces risques, il y a plusieurs étapes :

- Une prise de conscience (cet article par exemple) ;
- Une analyse des risques spécifiques ;
- Une intégration des IA dans le management de la qualité, accompagné d'un changement de culture des équipes ;
- Une certification éventuelle.

Il ne s'agit pas d'une révolution, mais bien d'une évolution. De même que la cybersécurité a commencé avec des filtres spams et des antivirus dans les années 90, puis a évolué vers des problèmes d'infrastructure pour devenir une discipline globale et un marché dynamique, la qualité des IA va vraisemblablement prendre le même chemin, et de manière plus rapide, poussée par la pression de la loi.

### Les enjeux du succès :

- Compliance
- Scaling
- Usage responsable

### 3.1 – Premier pas : l'analyse de risque spécifique IA, préalable à toute initiative d'envergure

Pourquoi une analyse de risque ?

- Toutes les applications n'ont pas la même exposition au risque (EU AI Act).
- Tous les risques n'ont pas la même importance (scope x sévérité x fréquence).

#### Comment construire son analyse de risque ?

1. En interne : former un ou des experts sur les standards spécifiques IA et le EU AI Act et leur faire créer un modèle holistique d'identification des risques ;

2. En externe : faire appel à des partenaires. Attention ici à bien choisir : il ne s'agit pas d'appliquer un système propriétaire des big five ou autres, mais bien de se mesurer au cadre légal et aux standards...
3. Mixer les deux approches pour assurer l'acquisition de savoir-faire et l'indépendance à moyen terme.

Selon mon expérience personnelle, il y a environ une cinquantaine de risques IA pour une entreprise, que l'on peut grouper autour de :

- Sécurité et sécurité,
- Juridique,
- Éthique & sustainability,
- Performance.

Par exemple en sûreté : predictable, reliable, repeatable, robust, testable, traceable, accountable, explainable, locally stable... Bien entendu, il y a d'autres manières d'organiser l'information. Cette analyse de risque va permettre de prioriser les actions et de définir un chemin en interne vers la gestion active d'une IA dans son contexte légal et qualitatif.

### 3.2 – Vers le management de la qualité d'une IA

Nous avons vu que l'analyse de risque est la première étape. Ensuite, les méthodes habituelles s'appliquent pour conduire le changement :

- Identifier les équipes clefs (IA, Qualité, Legal...)
- Former (recruter ?) des équipes, à la fois techniques et management :
- Les risques et les obligations,
- Les méthodes et standards pour y répondre,
- Intégrer les problématiques dans les fonctions d'achat ;
- Intégrer les IA dans le système de management de la qualité ;
- Construire un audit (interne ou externe), incluant les fournisseurs de data, les intégrateurs ;
- Élargir les fonctions du compliance manager ;
- Dans le cas des risques élevés, préparer son entreprise aux audits et certifications à venir, en s'appuyant sur les pistes évoquées en amont.

### 3.3 – Par où commencer ?

Les bonnes adresses :

- Vos organismes de certification habituels (SGS, TÜV, Dekra, Buro Veritas, ...), dans la mesure où ils couvrent les IAs ;
- Vos partenaires de conseil, même si la plupart sont très en retard sur cet aspect spécifiques de l'IA. Ils sont historiquement plus impliqués sur les applications IA elles-mêmes, les infrastructures IA et la cybersécurité en général ;
- Certaines associations professionnelles.

## Conclusion

Un tsunami législatif et normatif est à l'horizon. La plupart des sociétés tournent encore le dos à la mer, fascinées par les perspectives qu'offrent les applications de l'IA. Une prise de conscience pour les organisations doit s'opérer face aux difficultés liées aux IA : compliance, scaling et usage responsable.

Cette vague va déferler sur deux ans, permettant l'intégration normative non seulement dans les systèmes d'innovation et de management des IA, mais aussi dans toute la chaîne d'approvisionnement. Les contraintes sont techniques, multi-dimensionnelles et les enjeux sont à même d'impacter durablement la (sur-)vie d'une entreprise. Tout le monde ne sera pas impacté au même degré, mais tous auront bientôt les pieds dans l'eau. Il est encore temps d'apprendre à nager.

*Philippe Coution*



# L'INTELLIGENCE ARTIFICIELLE, LA CONFIANCE ET LA RÉGLEMENTATION

Source : [https://pxhere.com/fr/photo/1638452?utm\\_content=shareClip&utm\\_medium=referral&utm\\_source=pxhere](https://pxhere.com/fr/photo/1638452?utm_content=shareClip&utm_medium=referral&utm_source=pxhere)

**La confiance est un point crucial pour l'adoption de l'intelligence artificielle. Car à l'heure actuelle, l'IA n'est plus dans un lointain horizon de science-fiction, elle est déjà bel et bien au cœur de la société. Que ce soit sur le plan scientifique, avec 5 478 publications sur l'IA en 2015 et 34 736 en 2020 d'après Stanford (1), ou sur le plan politique avec un sujet qui anime de nombreux débats sociétaux. Mais, même si l'IA va impacter le quotidien de beaucoup de gens, de par ses utilisations qui seront très nombreuses et très variées, pourquoi ne la voit-on pas déjà concrètement en action partout ? Simplement, car en l'état, l'IA ne peut pas encore être déployée opérationnellement dans de nombreux secteurs. Avant de pouvoir généraliser ces déploiements, il faut que les questions de confiance que soulève l'IA soient correctement adressées.**

## L'expression d'un besoin de sûreté

Si des questions de confiance se posent, c'est bien évidemment qu'elles reflètent certaines appréhensions. Les algorithmes ne sont jamais exempts de défauts, et les tâches qui leur sont attribuées peuvent échouer dans leur exécution. Les répercussions potentielles de ces défauts peuvent être importantes lorsque les algorithmes sont utilisés pour automatiser certaines tâches dans des cadres critiques. Historiquement, de telles erreurs algorithmiques se sont déjà produites, avec des erreurs d'arrondi qui ont par exemple coûté la vie en 1991 à 28 soldats américains lorsqu'un missile Patriot a échoué l'interception d'un missile irakien. Ou encore, lorsqu'un problème de conversion numérique a été à l'origine de l'explosion de la fusée Ariane 5 en 1996, causant une perte estimée à plus de 500 millions de dollars (2).

L'IA, qui n'est rien de plus qu'un autre type d'algorithme, a également connu son lot de dysfonctionnements. Pour en citer quelques cas, en 2018 une voiture sans chauffeur a renversé et causé la mort d'un piéton après l'avoir confondu avec un objet sans avoir pu déclencher sa procédure de freinage d'urgence suite à un mauvais réglage (3). Également, les IA sont couramment affectées par des biais qui peuvent fausser leur fonctionnement. Par exemple, les algorithmes d'IA peuvent être victimes de biais d'apprentissage. Ces biais peuvent être introduits par une mauvaise représentativité, comme dans le cas d'un algorithme de reconnaissance faciale qui ne reconnaît pas les

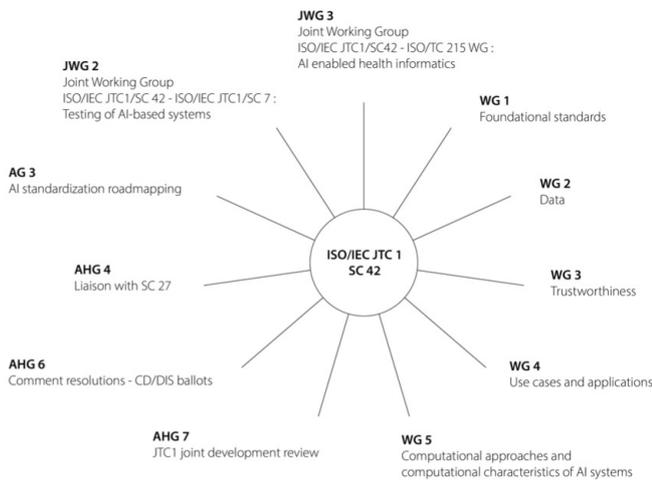
personnes de couleur car elles n'étaient pas assez représentées dans le jeu de données d'apprentissage (4).

## Le développement des initiatives législatives

Ces exemples ont pu malheureusement alimenter des craintes mais ils mettent surtout en valeur les faiblesses des IA auxquelles il faut pallier. Améliorer les méthodes de construction des IA permet justement d'identifier ces risques pour mieux les réduire. L'enjeu est surtout, pour l'industrie, de développer des IA en suivant des principes éthiques qui faciliteront l'accord de la confiance du public à l'égard ces systèmes. Même si l'IA représente de grandes opportunités économiques, la gestion des risques reste un prérequis indispensable dans de nombreux cas d'utilisation. Pour cela, il est nécessaire de fixer un cadre qui permettra de contrôler le déploiement de cette technologie dans le but de dépasser les interdictions d'utilisation qui étaient jusqu'ici en vigueur. Effectivement, la technologie a évolué depuis plusieurs années et les risques associés à l'utilisation de l'IA peuvent maintenant être mieux contrôlés. Il est maintenant possible d'imaginer, d'ici quelques années, de l'IA intégrée dans des secteurs comme le transport, la santé ou l'énergie, qui aujourd'hui ne les acceptent pas ou avec difficulté.

Le caractère horizontal et transnational de cette technologie en fait un enjeu international auquel les initiatives de standardisation et de réglementation veulent répondre.

Plusieurs pays cherchent à réglementer l'IA sur leur territoire. Par exemple aux États-Unis avec l'"Algorithmic Accountability Act" (5), en Chine avec le "Algorithmic Recommendation Management Regulations" (6) ou encore au Canada avec "The Artificial Intelligence and Data Act (AIDA)" (7) pour n'en citer que quelques-uns. Beaucoup de textes sont en cours de rédaction ou en attente d'approbation tandis que des standards à vocation mondiale sont en développement à l'ISO, l'IEC et au CEN CENELEC. En 2018, l'ISO/IEC a constitué le sous-comité JTC 1/SC 42 comprenant plus de 300 experts de plus de 30 pays différents qui œuvrent à l'élaboration de standards sur l'intelligence artificielle (8). Actuellement 40 projets ont été lancés, dont 16 ont déjà été publiés. Ces standards cherchent à adresser tous les aspects de l'IA, notamment au niveau algorithmique concernant la robustesse et l'explicabilité des systèmes, pour poser les bases d'un véritable cadre de confiance dans l'IA.



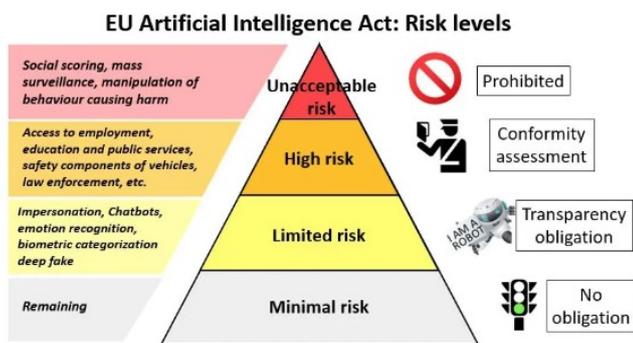
Organisation du ISO/IEC JTC 1/SC 42

En Europe, les initiatives portent autour de l'AI Act, qui est le projet de réglementation de l'IA qui concerne tout algorithme d'IA impactant des citoyens européens, au même titre que la RGPD vis-à-vis de leurs données personnelles. L'AI Act est une initiative de la Commission Européenne qui a mandaté le CEN-CENELEC (l'organisme de standardisation européen) pour lui fournir les standards capables d'opérationnaliser son règlement. Pour répondre aux délais très courts imposés pour la publication de l'AI Act (début 2025), le CEN CENELEC va utiliser les standards de l'ISO/IEC sur l'IA qui sont déjà disponibles. Le but étant d'obtenir au final des standards dits "harmonisés" qui permettront d'avoir une présomption de conformité aux entreprises qui les utiliseront.

En France, un organisme devra être nommé pour délivrer les certifications liées à l'AI Act. Afin de délivrer une certification, cet organisme réalisera des audits de conformité des algorithmes d'IA considérés à haut risque.

L'apport de la Réglementation européenne : l'EU AI ACT

Pour revenir sur le plan Européen, la principale initiative de réglementation s'incarne donc au travers de l'AI Act (9). Le cœur de cette réglementation s'articule autour de la catégorisation des systèmes d'IA au regard de leur finalité, suivant quatre niveaux de risques.



(Source: Telefónica)

Pyramide des risques de l'EU AI Act (10)

Tout d'abord, la catégorie des risques faibles ou minimaux, pour laquelle il n'y a aucune obligation réglementaire. Il s'agit de l'automatisation du traitement de mails par exemple ou les IA utilisées dans les jeux vidéo.

Le niveau au-dessus concerne les systèmes avec un risque limité. Il y a alors une obligation de transparence vis-à-vis de l'utilisateur concernant le fait qu'il utilise ou interagit avec une IA. Les IA concernées sont par exemple celles basées sur le traitement du langage naturel (chatbots) ou les algorithmes artistiques.

Le dernier niveau où l'IA pourra être déployée dans nos sociétés concerne les systèmes d'IA à haut risque. L'obligation devient alors de se plier à une évaluation de conformité à réaliser avant que le système puisse être déployé. L'évaluation porte sur une multitude de critères. Certains concernent la qualité des jeux de données utilisés afin de minimiser les risques de biais et de discrimination, d'autres portent sur la documentation et la tenue de registres concernant la traçabilité. Mais également sur la transparence, la fourniture d'informations adéquates aux utilisateurs, la supervision humaine, la robustesse, la précision de l'algorithme, ainsi que sur les dispositions en matière de cybersécurité. La réussite de l'évaluation conduira finalement au "marquage CE" du système d'IA (11). Les domaines d'IA concernés sont par exemple : l'accès à l'emploi, l'éducation, les services publics, la gestion des infrastructures critiques, les éléments de sécurité des véhicules, l'application de la loi, l'administration de la justice, etc. Enfin vient le niveau d'interdiction pour lequel les risques sont jugés inacceptables. Le déploiement des IA concernées est alors tout simplement interdit. Les systèmes d'IA de cette catégorie sont ceux réalisant de l'analyse sociale, de la manipulation du comportement humain de façon subliminale ou à des fins préjudiciables et de la surveillance de masse. Il existe cependant des exceptions possibles concernant les systèmes de surveillance de masse. Pour ce faire, l'organisation devra obtenir une autorisation d'une autorité administrative ou judiciaire et être préalablement encadrée par une loi détaillant l'usage prévu (12). Par exemple, les exceptions peuvent être octroyées dans le cadre d'une compromission de la sécurité nationale, afin de prévenir une menace spécifique, substantielle et imminente pour la vie ou la sécurité des personnes. Il serait donc possible de voir de tels systèmes de surveillance dans des lieux publics tels que des gares ou des aéroports, mais de manière très encadrée.

Avec la sortie officielle du texte de l'AI Act pour début 2025, ces règles deviendront obligatoires et devront être suivies par toutes les entreprises souhaitant déployer un système d'IA en Europe ou interagissant avec des Européens, sur des principes similaires au RGPD.

L'importance de la validation algorithmique

C'est alors que la validation des systèmes d'IA prend tout son sens et devient cruciale pour la conformité, permettant d'obtenir le niveau de confiance technique exigé. Un des piliers de la confiance et donc de la validation concerne la robustesse des IA (13). Durant la validation il sera possible tester la résistance des algorithmes face à

des bruits, par exemple en simulant certaines conditions météorologiques pour évaluer la résilience des systèmes d'analyses d'images. Techniquement, les validations statistiques qui seront réalisées pourront s'accompagner de techniques de validation formelle. Ces techniques mathématiques permettent de prouver des propriétés de robustesse sur des ensembles de cas possibles, pour compléter les preuves statistiques de fonctionnement qui sont réalisées sur des échantillons isolés. Qui plus est, le fonctionnement des algorithmes devra être suivi tout au long du cycle de vie des IA, afin de repérer de potentiels changements dans leur conformité. Conformément aux attentes de l'AI Act, les principes de réalisation de cette évaluation seront détaillés dans les standards que va produire le CEN CENELEC. La finalité est toujours d'atteindre un niveau d'exigence et d'assurance qui permettra d'instaurer un cadre de confiance afin que les systèmes d'IA puissent être déployés à grande échelle.

*Arnault IOUALALEN*  
& *Baptiste AELBRECHT,*  
NUMALIS  


À propos des auteurs : La nécessité de confiance est soutenue par Numalis dont les activités visent à avoir un impact positif sur le déploiement de l'IA dans nos sociétés en développant des outils pour l'explicabilité et à la validation de la robustesse des algorithmes d'IA (14). Numalis porte ces valeurs internationalement grâce à son PDG, Arnault Ioualalen, qui est l'éditeur des standards ISO/IEC portant sur l'évaluation de la robustesse des IA (série de standards ISO/IEC 24029).

## Sources

- (1)[https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report\\_-\\_Chapter-1.pdf](https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_-_Chapter-1.pdf)
- (2)<https://www.apprendre-en-ligne.net/bloginfo/index.php/2008/10/10/79-problemes-d-arrondi>
- (3)[https://www.lepoint.fr/high-tech-internet/accident-mortel-d-uber-le-logiciel-n-a-pas-reconnu-un-pieton-hors-des-clous-06-11-2019-2345525\\_47.php](https://www.lepoint.fr/high-tech-internet/accident-mortel-d-uber-le-logiciel-n-a-pas-reconnu-un-pieton-hors-des-clous-06-11-2019-2345525_47.php)
- (4)<https://www.lefigaro.fr/secteur/high-tech/2015/07/02/32001-20150702ARTFIG00144-la-technologie-de-reconnaissance-faciale-est-elle-raciste.php>
- (5)<https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clark-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems>
- (6)<https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf>
- (7)<https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>
- (8)<https://www.iso.org/fr/committee/6794475.html>
- (9)<https://artificialintelligenceact.eu/>
- (10)<https://www.telefonica.com/en/communication-room/blog/a-fit-for-purpose-and-borderless-european-artificial-intelligence-regulation/>
- (11)<https://www.polytechnique-insights.com/dossiers/digital/les-nouveaux-enjeux-de-lia/ai-act-comment-leurope-veut-reguler-les-machines/>
- (12)<https://blog.leclubdesjuristes.com/artificial-intelligence-act-lunion-europeenne-invente-la-pyramide-des-risques-de-lintelligence-artificielle/>
- (13)<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- (14) <https://awesome-mclaren.54-37-85-119.plesk.page/ai-improvement-solutions/iso-standards-and-eu-ai-act-compliance/>

# LES IA VONT-ELLES REMPLACER LES MÉTIERS INTELLECTUELS ? 4 SCÉNARIOS À HORIZON 20 ANS



Source : Image générée par DALL-E le 12/02/2023

## Le choc produit par ChatGPT

Le lancement de ChatGPT fin 2022 par OpenAI a été un choc pour de nombreux observateurs. Tout à coup, une IA est capable de produire des textes qui peuvent facilement être confondus avec des écrits humains sur à peu près n'importe quel sujet.

Les performances de ChatGPT sont assez nettement au-dessus de celles des versions précédentes d'IA génératives ([Generative pre-trained transformer](#), GPT dans la suite de l'article) appliquées au langage naturel (Large Language Models, LLM dans la suite de l'article), et notamment de GPT-3 lancé à l'été 2020. Il est aussi plus facile d'intégrer avec ChatGPT qui répond directement à des instructions et il est mieux aligné avec certains objectifs définis par OpenAI (notamment l'objectif d'éviter les propos choquants, racistes).

Enfin, ChatGPT a été lancé sur une interface gratuite, ouverte à tous, facile à utiliser. Le résultat a été une adoption sans précédent dans l'histoire avec 100 millions d'utilisateurs actifs [au bout de seulement 2 mois](#) (à titre de comparaison, TikTok a eu besoin de 9 mois, Spotify de 55 mois). Une rapide estimation dans un espace de coworking parisien m'a permis de constater qu'à tout instant, autour d'un écran sur 5 était ouvert sur l'interface de ChatGPT.

Les GPT les plus récents ont aussi atteint des résultats objectivables dans de nombreux domaines. [GPT-4](#), dernière version lancée par OpenAI en mars atteint le top 10% dans les épreuves du barreau américain, aux tests standardisés [SAT](#) utilisés pour l'admission aux universités et de très bonnes performances dans le domaine de la [santé](#) et même en [sommellerie](#).

## Hype et contre-hype

Les succès de cette nouvelle génération de LLM ont soulevé un fort enthousiasme et beaucoup d'optimisme sur les applications possibles, notamment dans les milieux VC et start-up. La vision la plus maximaliste consiste à imaginer que les LLM deviendront notre interface universelle avec toutes les informations dont nous avons besoin. Certains ont donc même prédit la fin de Google Search.

Cette « hype » a entraîné en réaction une « contre-hype » qui pointe les limites des LLM, les incertitudes sur nos capacités à les transformer en applications fiables et utiles et plus généralement un appel à [modérer l'optimisme](#).

Cette opposition entre enthousiasme et pessimisme est relativement fréquente. On l'a observée aussi bien pour des technologies à faible impact (Web3), à fort impact (internet, le smartphone) ou à impact intermédiaire (le machine learning des années 2010).

## Craintes et espoirs pour l'avenir des métiers intellectuels

Chaque fois qu'une technologie émergente semble pouvoir automatiser des tâches auparavant réservées au travail humain, elle suscite un mélange de craintes et d'espoirs.

Les craintes portent sur le risque de disparitions d'emplois non remplacés, et la difficulté pour ces travailleurs de conserver leur source de revenu. Les espoirs portent sur l'augmentation de la qualité du service, la réduction des coûts, l'automatisation d'activités peu intéressantes.

L'histoire économique depuis la 1ère révolution industrielle a toujours plutôt confirmé les espoirs et relativisé les craintes. Il faudrait probablement développer et nuancer mais globalement les gains de productivité énormes permis par les technologies n'ont pas provoqué de chômage de masse et les gains de revenus ont permis de développer de nouvelles activités (exemple : en France, la part de l'agriculture dans le PIB est passé de [18 % à 2 % du PIB](#) entre 1950 et 2015 tandis que les dépenses soins et biens médicaux sont passées de [2,5 % à 9 % du PIB](#)).

Cependant, à chaque nouvelle vague d'innovation, certains imaginent que l'histoire pourrait être différente cette fois-ci. Et à chaque fois il y a de bonnes raisons de le penser. Dans notre cas, la différence est que le potentiel d'automatisation porte cette fois-ci sur toutes les tâches intellectuelles.

Un point crucial sera de voir si les GPT sont plutôt utilisés dans une logique de substitution ou dans une logique de complémentarité : la technologie va-t-elle remplacer le travail humain ou va-t-elle le compléter et le rendre plus productif ?

Une question connexe est celle de l'impact sur la demande pour les activités intellectuelles qui intègrent l'IA. L'automatisation a pour premier effet de réduire les coûts et les prix. Lorsque le prix d'un service diminue, on peut avoir envie d'en consommer davantage (en volume, on bien en qualité en développant de nouvelles fonctionnalités) mais on peut aussi préférer utiliser le revenu dégagé pour d'autres biens et services.

Chez [Centolla](#), nous avons exploré différentes pistes d'automatisation pour nos travaux d'analyse et de prévision économiques et à ce stade nous voyons un peu les deux effets :

- Automatisation de la production de commentaires simples : pure substitution, gain de temps humain sur des tâches simples.
- Développement de fonctionnalités supplémentaires dans nos travaux que l'on n'aurait pas eu le temps ou pas eu les capacités de faire : un [exemple récent](#) étant l'utilisation de données de recherche Google pour enrichir notre compréhension de la conjoncture dans le secteur touristique.

### Facteurs d'incertitudes

Face à un avenir incertain, il est utile d'identifier les facteurs susceptibles de nous conduire vers différents scénarios possibles. Pour anticiper l'impact des GPT sur les métiers intellectuels, j'en liste 4 principaux :

- L'évolution des capacités intellectuelles des IA ;
- La capacité à exploiter les IA pour développer des services utiles économiquement ;
- Le degré de substituabilité avec le travail humain de ces services;
- L'acceptabilité par les populations et les [pouvoirs publics](#) et le niveau de régulation qui en découle.

D'autres facteurs importants sont aussi par exemple : le degré de centralisation (oligopole autour des leaders ou écosystème plus décentralisé), le pricing des services d'IA (qui dépend du point précédent, de l'évolution des algorithmes de calculs, du coût de la puissance de calcul).

## 4 scénarios sur l'impact de l'IA sur le futur des métiers intellectuels à horizon 20 ans

### Scénario 1 : plafonnement des capacités intellectuelles de l'IA

Dans ce scénario, les futures générations de GPT sont un peu meilleures que les précédentes, mais sans avancées remarquables.

Malgré de nombreuses tentatives, on se rend compte qu'il reste difficile de produire des résultats suffisamment fiables pour la plupart des applications concrètes.

La réglementation encadre strictement l'utilisation de l'IA pour limiter les erreurs dans certains domaines sensibles (santé, droit, finance). Finalement, les GPT révolutionnent certaines niches (publicité en ligne, médias) mais l'impact macroéconomique est faible, tout comme l'impact sur le marché du travail.

### Scénario 2 : essor des copilotes

Les capacités intellectuelles des GPT poursuivent leur progression incrémentale. Surtout, on réussit à développer un grand nombre d'applications économiquement utiles dans divers secteurs.

On reste cependant plutôt dans une logique de complémentarité que de substitution car les capacités de ces services sont bien moins générales que l'intelligence humaine.

Beaucoup de métiers intellectuels ont leur propre "copilote", un service d'IA qui augmente fortement leur productivité :

- Les développeurs ne codent plus beaucoup. Ils réfléchissent plutôt aux fonctionnalités, puis testent et modifient à la marge le code généré par le copilote.
- Les chercheurs génèrent automatiquement les revues de littérature adaptées à leur sujet.
- Les enseignants ne s'occupent plus que de la transmission, alors que le copilote prépare des contenus adaptés aux élèves et réalise les évaluations.

### Scénario 3 : des pilotes plus que des copilotes

Dans ce scénario, les capacités des IA sont déjà similaires ou supérieures à celles des humains.

Les pilotes sont devenus tellement performants qu'ils ont progressivement pris une place plus importante que les travailleurs humains. On considère désormais qu'il est très risqué de prendre une décision médicale sans la faire valider par une IA et vraiment incongru de réaliser une analyse juridique ou comptable autrement que par l'IA.

L'impact économique est fort mais ce n'est pas nécessairement un scénario de chômage de masse. On peut souhaiter pour répondre à des objectifs sociaux maintenir des métiers sans raisonner uniquement dans une logique de coûts. Certains métiers comme ceux de l'enseignement peuvent devenir bien plus valorisés et attirer des profils plus larges qu'aujourd'hui. Plus généralement, il est probable que dans cette nouvelle économie se développe une demande d'authenticité humaine, y compris pour la réalisation de tâches intellectuelles.

#### Scénario 4 : face aux dangers des IA, la recherche est proscrite

[De nombreuses voix s'élèvent déjà](#) pour alerter sur les risques liés au développement de l'IA. Ces risques sont très divers : impact brutal sur le marché du travail et ses conséquences sociales, risque de manipulation, utilisation par des acteurs malveillants voire, pour certains, risque existentiel pour l'humanité.

On peut donc imaginer que le développement de ces craintes conduit à une forte réaction des pouvoirs publics pour ralentir ou arrêter la recherche sur les IA. Au vu des fortes incitations économiques à l'œuvre et de l'indispensable coordination internationale nécessaire, cela peut sembler très difficile mais pas forcément impossible.

Au sein des démocraties américaine et européennes, les opinions publiques, déjà souvent réceptives au pessimisme technologique, pourraient rapidement devenir sensibles à cette question. La Chine, qui privilégie la stabilité sociale sur les progrès économiques et technologiques, pourrait aussi rejoindre le mouvement.

Dans ce scénario, on pourrait ainsi imaginer que soit mis en place un strict encadrement de la production et de l'utilisation des GPU, rendant impossible l'entraînement de nouveaux GPT.

#### Conclusion et opinion personnelle

Anticiper l'impact à long terme des nouvelles technologies est [no-toirement difficile](#) et notre sujet est particulièrement incertain. L'approche par scénarios est intéressante car décrire les possibles permet d'avoir un cadre pour suivre les événements et voir quelles hypothèses sont confirmées ou remises en cause.

Une bonne pratique dans ce type d'analyse prospective est de systématiquement donner des probabilités aux scénarios proposés. Cela peut sembler surprenant et un peu arbitraire mais c'est très utile : cela permet d'éviter [l'ambiguïté du langage courant pour décrire l'incertitude](#) et cela force à "se mouiller" et notamment à hiérarchiser les scénarios dans un ensemble cohérent.

Une difficulté supplémentaire dans notre courte analyse est que les scénarios ne sont pas vraiment exhaustifs. Mais pour faire l'exercice, imaginons ici que dans 20 ans nous décidions de quel scénario est le plus proche de la situation réelle. Voilà alors ce que je dirais aujourd'hui :

- Scénario 1 : 5 %
- Scénario 2 : 50 %
- Scénario 3 : 35 %
- Scénario 4 (et autres scénarios extrêmes qui se terminent bien ou mal) : 10 %

*Antoine Herlin*  


<sup>5</sup>. Le livre de Dave Gray Jouer pour Innover (traduction de game-storming) en donne de nombreux exemples opérationnels.

<sup>6</sup>. Zoom ou Teams, par exemple.

## BILLET D'HUMEUR - LA QUESTION DES HALLUCINATIONS



Source : image générée par DALL-E le 12/02/2023

**Tout un chacun a sans doute tenté sa propre expérience et réussi à prendre en défaut ChatGPT. Mais qu'est-ce que cela prouve au juste ?**

L'expression anthropomorphe « hallucinations » est rapidement devenue populaire pour désigner certains types d'erreurs des LLM (Large Language Models) lorsqu'il y a un décalage manifeste entre les sorties du modèle et la réalité : détails entièrement inventés, [fausses références](#), contre-vérités évidentes.

Ces hallucinations montrent bien les limites intellectuelles des LLM. L'apprentissage sur des énormes bases de texte et leur très bonne mémoire associative leur permet de produire du contenu souvent cohérent sur n'importe quel sujet. Mais leur capacité de raisonnement est limitée et ils n'ont pas forcément de véritable compréhension de la réalité et des phénomènes qui justifient leurs croyances. Un peu comme des étudiants qui auraient appris leur leçon par cœur sans vraiment comprendre le fond (position notamment de [Yann Lecun](#)).

Il faut malgré tout noter que notre compréhension des mécanismes d'inférence des LLM est encore assez limitée, que la question des capacités de raisonnement [est discutée](#) et donc que tout cela dépendra beaucoup des recherches futures et des prochaines versions de LLM.

Pour revenir aux hallucinations, mon sentiment est que les erreurs relevées sur ChatGPT donnent peu d'informations utiles pour anticiper l'impact à plus long terme de ce type de technologie. Une analogie que je trouve assez parlante est de voir un LLM comme une boîte à outils qui permet de réaliser un grand nombre de tâches très différentes (NB : la différence la plus fondamentale entre cette génération d'IA et les précédentes est sans doute une plus grande capacité de généralisation). Beaucoup d'hallucinations relayées massivement me font penser à quelqu'un qui sortirait un tournevis pour taper sur un clou avant de crier « Ah ah, ça ne marche pas du tout ! ».

Pour mieux comprendre les impacts possibles, il faudrait en quelque sorte inverser la charge de la preuve : plutôt que de chercher à répondre à la question « Est-ce que GPT sait répondre à mon problème quelle que soit la manière dont je l'utilise ? », il faudrait travailler sur la question « Est-ce que pour mon problème il existe un LLM et une façon de l'utiliser qui donne des réponses pertinentes ? » Dans ce dernier cas, on pourra utiliser un LLM spécialement conçu pour le besoin particulier (prompt tuning, fine tuning sur des tâches précises, RLHF, intégration dans des produits plus complexes avec plusieurs couches, étapes de contrôle qualité...).

Gardons en tête que ChatGPT est un outil expérimental dont l'objectif est de montrer au grand public les capacités des LLM sans produire de contenus racistes. A ce titre, difficile de considérer qu'il ne remplit pas sa fonction.

Pour ce qui est des autres applications, beaucoup de projets sont en cours, des investissements importants ont été mis sur la table et on aura probablement bientôt plus de données.

*Antoine Herlin*



# UNE RÉVOLUTION TECHNOLOGIQUE PAVÉE DE BONNES INTENTIONS ?



Source : « Les ingénieurs se passionnent souvent pour les innovations techniques, laissant à d'autres le soin d'en évaluer les conséquences sociétales. », image créée par DALL-E le 30/04/2023

**Les ingénieurs se passionnent souvent pour les innovations techniques, laissant à d'autres le soin d'en évaluer les conséquences sociétales. Étant donné la magnitude des changements qu'elle apporte, la démocratisation des outils d'IA doit faire exception, afin d'anticiper l'adaptation au monde qui vient...**

Bluffé par les performances de ChatGPT, Bill Gates s'enthousiasme dans [un billet de blog](#) : « Le développement de l'IA est aussi fondamental que la création du microprocesseur, de l'ordinateur personnel, de l'internet et du téléphone portable. » Grand philanthrope, il y voit une révolution positive dans le domaine de l'éducation, de la santé et, même, de manière assez peu argumentée, dans celui du changement climatique. S'il n'évacue pas complètement les risques, il affirme sa foi dans le fait qu'ils seront maîtrisés (« Les développeurs y travaillent ... »). Réaliste néanmoins sur le fait que « les forces du marché ne conduiront pas naturellement les produits et services d'IA à aider les plus pauvres », il appelle de ses vœux « un financement fiable et les bonnes politiques », afin que les IA puissent être orientées vers la réduction des inégalités. Malgré les difficultés, il reste donc profondément techno-optimiste : les problèmes créés par la technologie (comme les grandes disparités de revenu entre les individus) vont être résolus par de nouvelles technologies. Une course en avant qui, au passage, se concevrait mieux dans un monde sans limites que dans un ensemble fini, et déjà bien entamé, de réserves d'énergie et de matériaux...

**Les risques sont pourtant majeurs. Ils sont de trois ordres.**

## 1. Le détournement d'information et les activités criminelles

Si l'on écoute [Sam Altman](#), le jeune PDG d'OpenAI (l'entreprise créatrice de ChatGPT et de DALL-E), celui-ci confesse tout de même avoir « un peu peur » de l'IA, notamment du fait de son utilisation par des individus ou des organisations peu scrupuleuses. Il se déclare d'abord « particulièrement inquiet » que ces modèles puissent être utilisés pour la désinformation à grande échelle. Capables de créer très rapidement de vastes quantités de matériel - textes, images, sons et vidéos, les chatbots (robots de conversation) peuvent en effet très rapidement inonder l'internet de fausses nouvelles semblant provenir de sources humaines sérieuses. A l'activité déjà inquiétante des professionnels de la manipulation de l'information s'ajoutent les activités

proprement mafieuses et criminelles. [Selon un spécialiste](#), des pirates sont déjà à l'œuvre : « Les cybercriminels s'intéressent de plus en plus à ChatGPT, car sa technologie d'intelligence artificielle sous-jacente peut rendre un pirate plus rentable. »

## 2. Une perte définitive de qualité de l'information

Et que dire de ce qui adviendra de la qualité de l'information quand les sources seront-elles-mêmes en partie constituées de produits de l'IA ? Des informations fausses se validant mutuellement apparaîtront comme vraies et sûres ; les biais et préjugés courants seront sanctuarisés (aujourd'hui chatGPT associe sans hésiter « profession médicale » à « médecin » ou à « infirmière » selon que le contexte suggère un personnage masculin ou féminin). [Comme le dit la MIT Technology Review](#), les chatbots sont des « baratineurs notoires ». Autrement dit, ils déploient un vernis culturel qui fait bonne impression, mais la culture, dit-on, c'est ce qui reste quand on a tout oublié, et, précisément, le robot a oublié d'où vient l'information qu'il restitue. Outre la violation systématique de la propriété intellectuelle, le mélange indistinct de multiples sources par chatGPT ne permet pas la vérification des informations. C'est tout l'inverse de l'encyclopédie en ligne Wikipedia, où la citation des sources est, à juste titre, une règle absolue, qui permet le contrôle transparent et l'amélioration permanente du contenu.

## 3. Des pertes d'emploi massives à très court terme

Sans surprise, [les chatbots vont remplacer des salariés dans les métiers où il faut réaliser des synthèses, assembler des mots ou du code informatique, ou encore répondre à des questions en réunissant un ensemble d'informations dispersées](#). Voici les secteurs où ces compressions de personnel ont déjà commencé :

- Les codeurs, programmeurs informatiques, ingénieurs logiciels ou analystes de données : les outils IA peuvent créer du code, par exemple un site internet complet à partir d'une description détaillée, puis le tester et le valider, et cela beaucoup plus rapidement qu'un intervenant humain. Cela pourrait se traduire par une productivité améliorée de 50%, répartie entre une baisse de l'emploi et un accroissement de l'activité pour répondre plus rapidement à la demande, le secteur étant actuellement en pénurie de main d'œuvre ;
- Les métiers de la communication : ChatGPT écrit des articles, des publicités ou d'autres contenus bien plus rapidement que les humains. La société américaine de médias internet BuzzFeed a déjà boosté son

cours de bourse en annonçant qu'elle utiliserait ChatGPT pour générer des contenus et le groupe de médias allemand Axel Springer a annoncé début mars le remplacement de certains employés par l'IA et des suppressions d'emplois dans les rédactions de ses journaux Bild et Die Welt. Cela touche aussi la création graphique : DALL-E, outil d'IA capable de générer des images en quelques secondes, [aura également un impact lourd sur l'emploi dans ce domaine](#) ;

- Les métiers qui manipulent de grandes quantités de données structurées : le secteur juridique, la comptabilité, les analyses d'études de marché et l'analyse financière. L'automatisation des tâches pourrait s'étendre aux métiers des conseillers financiers et des traders ;

- Les agents de centres d'appel et de service client (assistance, après-vente) : 25% de l'emploi pourrait disparaître dans les 5 prochaines années [selon une étude réalisée en 2022 par la société de recherche technologique Gartner](#).

- Les enseignants ne sont pas à l'abri, ni d'ailleurs les consultants ou même les chefs d'entreprise, [puisque une entreprise chinoise importante dans le jeu vidéo est déjà dirigée par un chatbot](#).

Ces métiers ne vont pas totalement disparaître, mais leurs effectifs vont fondre rapidement à l'occasion des gains de productivité acquis par le recours à l'IA. L'accompagnement humain subsistant aura pour rôle d'orienter et de contrôler tant bien que mal la production des robots, qui, encore pour un certain temps, peuvent répondre de manière incorrecte à des questions suite à des lacunes dans leur apprentissage, par exemple faute de données pertinentes sur le net, faire des erreurs mathématiques ou logiques, et produire des codages fautifs.

## Que faire par rapport à ces risques ?

Le risque de désinformation devra être combattu par de nouveaux procédés de certification de l'information, une attention redoublée devant être apportée aux sources. Les images d'une arrestation musclée de Donald Trump, générées par l'IA, ont convaincu de nombreux internautes... Ira-t-on jusqu'à une certification d'information « sans IA » comme on a aujourd'hui des aliments « sans OGM » ? Ce sera sans doute nécessaire sachant que le risque de pollution des sources elles-mêmes par des informations erronées issues de l'IA créerait un problème insoluble.

Le risque qualité fait l'objet de nombreuses démarches de normalisation et de sécurisation des résultats de l'IA, dont la présente Lettre de XMP-Consult se fait l'écho. Ces démarches seront très probablement couronnées de succès. Mais de ce fait, les contrôles humains pourront être encore allégés, ce qui ne fait que renforcer le troisième risque, le risque social.

## Un changement de société à réussir

Il y a dix ans, une étude de l'Université d'Oxford (dite étude Frey et Osborne) prédisait que 47 % des emplois américains pourraient être supprimés par l'IA au cours des vingt prochaines années. En 2018, l'OCDE estimait que 14 % des emplois avaient une probabilité d'être automatisés de 70 % ou plus, et 32% une probabilité comprise entre

50 % et 70 %, soit 210 millions d'emplois menacés au total des 32 pays étudiés. La réalité sera-t-elle pire, ou meilleure ? Elle reste bien difficile à prévoir avec exactitude mais, dans tous les cas, la magnitude des dispositifs d'adaptation à prévoir mérite dès à présent toute l'attention des politiques. Le saut quantique de productivité ne va pas juste libérer en douceur « du temps pour faire d'autres choses, au travail et à la maison » comme l'écrit sereinement Bill Gates, mais susciter en quelques années de nouvelles inégalités entre ceux qui continueront à travailler et ceux, très nombreux, qui se trouveront au chômage. Il s'agira cette fois de « cols blancs » et non plus de « cols bleus » comme dans les décennies 1970 et 1980, avec, à nouveau, des conséquences sociales et politiques majeures. Mais comme le faisait observer un ouvrage collectif sans doute prémoniteur paru en 1980, La Révolution du temps choisi, il y a déjà eu bien d'autres crises de ce genre : une dizaine de fois dans notre histoire, le système économique capitaliste a subi une crise de surproduction liée à une sous-consommation solvable, et, à chaque fois, après une forte résistance patronale et des mouvements sociaux parfois violents, une réduction du temps de travail a été acquise, sous différentes modalités (réduction de la journée, de la semaine, de l'année ou de la vie de travail). Au-delà de la mise au point d'un mécanisme pour libérer du temps de façon équitable pour tous, la question posée à la société sera en outre de permettre à chacun de vivre le temps libéré de façon riche, culturellement, socialement et peut-être aussi solidairement, dans le respect des exigences de durabilité environnementale désormais incontournables. Face à une forte réduction des temps de travail, la question du sens à la vie se substitue ainsi à celle du sens au travail.

Si, dans les prochaines dix ou vingt années, l'atterrissage de l'IA se fait dans le sens d'une société équilibrée et solidaire, cette révolution technologique aura été utile et bénéfique comme le souhaite Bill Gates. Si au contraire, cette transition se traite dans la violence concurrentielle, financière et sociale, elle n'aura été qu'une descente vers un enfer pavé de bonnes intentions.

*Antoine Jaumes*



*Bill Gates fasciné par ChatGPT ?*

# LA RSE PERMET D'ENGAGER LES SALARIÉS DANS L'ADOPTION DE L'IA



Source : « Deux personnes se serrent la main », image générée par DALL-E le 12/02/2023

**La RSE propose un cadre éthique exigeant. L'embrasser est le meilleur gage que les dirigeants peuvent donner aux collaborateurs que l'IA sera utilisée au bénéfice de tous, eux-mêmes y compris. Et, partant de là, elle permet de les engager plus sereinement dans l'adoption de l'IA au travail.**

## Des dirigeants par nature positifs

Les dirigeants d'entreprise ont une vision positive de l'avenir. C'est même leur première caractéristique ! Dans une logique de marché, l'IA offre un avantage concurrentiel aux premiers qui sauront la mettre à leur service. Elle permet de recentrer les équipes sur le relationnel, la stratégie et la créativité, pour créer de nouveaux produits et services ou encore optimiser au mieux des ressources se raréfiant. Vue de façon globale et positive, selon une étude de Microsoft et PwC, l'IA pourrait d'ici 2030 être utilisée au service de l'homme et de l'environnement au travers une croissance du PIB mondial de 4,4%, une hausse de l'emploi de 1% et une réduction des émissions de GES de 4%.

Les dirigeants sont bien conscients des remises en question que cela nécessite, tout en voyant résolument le verre à moitié plein.

## Des collaborateurs plus inquiets

Les salariés semblent voir le verre dans une perspective différente. Ils constatent bien l'intérêt de l'IA, puisque, selon les Echos, un quart des Français sondés déclarent que celle-ci leur libérera du temps à consacrer à d'autres activités. Une étude d'OpenAI et de l'université de Pennsylvanie, même s'il s'agit des emplois outre-Atlantique, confirme l'impact de l'IA en prédisant que 19% des salariés seront affectés pour plus de la moitié de leurs tâches et 80 % d'entre eux pour 10 % de celles-ci.

Mais cette étude ne traite pas de ce « temps à consacrer à d'autres activités ». Aussi, face à l'inconnu de « à qui profitera le progrès », on peut comprendre qu'une étude de Deloitte livrait dès 2018 que la réaction des salariés face à l'IA était, dans l'ordre : « la curiosité, l'inquiétude et l'espoir ».

## L'optimisation, plus facile que l'innovation ?

Une observation commune vient biaiser toute lecture sereine du sujet : l'optimisation de l'existant est généralement plus facile que l'in-

novation. L'automatisation des tâches permet une optimisation de coût plus rapide que la création de valeur additionnelle. Cette dernière nécessite une réflexion prospective, puis un apprentissage collectif long et chaotique. Là où quelques « experts » peuvent plus facilement optimiser un système existant.

Dans un contexte de recherche de compétitivité, les salariés perçoivent donc d'abord l'impact sur leur emploi. Ce biais conservateur est naturel (Kahneman et Tversky, prix Nobel d'économie) et est donc valable dans les entreprises dont les dirigeants ont, par le passé, arbitré en faveur du long terme, car étant patrons chez eux ou dépendants d'actionnaires stables. Que dire des autres contextes d'actionnariats ? Et de l'observation au cours de la dernière décennie de la mise en place de la RPO (Robotic Process Automation), avec les premières IA d'alors, ayant pour but légitime de réduire les coûts et amenant inéluctablement des réductions d'emploi ?

Dans ce contexte, comment la RSE permet-elle de remettre de la sérénité dans la façon d'aborder l'IA avec les équipes ? Ce qui suit décrit une approche globale de la création et du partage de la valeur, propre à la RSE, et s'éloigne donc d'une lecture des seuls avantages concurrentiels.

## Des dirigeants précurseurs de la RSE

Un nombre croissant de dirigeants considère que la recherche du profit à court terme est une cause majeure de dégradation sociale et environnementale. Théorisée par Porter et Kramer en 2011, l'approche « creating shared value » propose de créer de la valeur pour l'ensemble des parties prenantes, en intégrant des solutions durables pour l'environnement. Des dirigeants comme Indra Nooyi (Pepsi-Co), Jean-Dominique Sénard (Michelin) ou Emmanuel Faber (Danone) l'ont mise en pratique.

Du point de vue de l'actionnaire, les études montrent de plus un impact positif de la RSE sur la performance. Une étude de France Stratégie de 2016, par exemple, indique un écart de 13% entre les entreprises ayant mis en place une telle approche et les autres.

## Pour l'IA, la RSE est un gage donné aux salariés

La RSE considère que les salariés sont des parties prenantes vis-à-vis desquelles l'entreprise est responsable des impacts de ses activités. Un dirigeant qui s'inscrit dans une telle démarche avant de lancer un projet IA donne donc des gages que les progrès permis bénéficieront

aussi aux parties prenantes, dont eux.

Une étude<sup>1</sup>, courte mais récente, appelle aussi à la réconciliation entre des visions différentes des dirigeants et des salariés en matière de culture de travail : si tous s'entendent pour dire que l'innovation au service des clients doit rester au cœur du modèle, les dirigeants appellent à plus d'agilité et d'anticipation, tandis que les salariés souhaitent davantage de collaboration et moins de pression. Bref, les dirigeants veulent garder l'entreprise sous tension, tandis que les collaborateurs veulent une approche plus qualitative du travail.

L'introduction de l'IA, sans un cadre de valeur(s) qui transcende d'abord ces différences, les amplifiera et créera du désengagement au moment où nombre d'entreprises peinent à recruter et fidéliser leurs collaborateurs. En répondant mieux aux attentes des salariés, le cadre de la RSE permet justement d'éviter ce piège.

Pratiquement, quelle approche adopter pour lier RSE et IA ? On peut pour cela proposer une grille d'analyse enrichie ainsi qu'une démarche simple pour la mettre en œuvre.

### Une analyse de la valeur enrichie

Les grilles concernant l'IA présentent typiquement trois axes : l'analyse de la valeur pour déterminer les activités automatisables, la faisabilité technique compte tenu de la maturité des systèmes d'information, et l'intérêt économique de l'IA en intégrant les coûts et les risques induits.

L'analyse de la valeur peut être enrichie de trois axes de performance RSE pour garder l'analyse simple toute en restant cohérent avec les normes (ISO, B-Corp...) : l'impact sur les ressources, l'impact social et la transparence. La sécurité des produits est un « droit à opérer » considéré ici comme déjà inclus. Il peut aussi être intéressant d'ajouter l'innovation comme une ressource culturelle contribuant à la pérennité de l'entreprise.

Au total, cela fait une matrice assez gérable, qui peut d'ailleurs se reformuler autour des axes de la triple comptabilité analytique « People/Planet/Profit ».

### Une démarche globale et expérimentale

La complexité tient entre autres au fait que la RSE se veut globale, commençant par un travail d'introspection, puis une inclusion progressive des parties prenantes. Tandis que l'IA invite plutôt à une expérimentation avant une extension pour avoir l'adhésion. L'ordre d'entrée en scène des nombreuses légitimités de la RSE est aussi clé et le dirigeant doit rester maître des horloges.

On peut combiner cela en une démarche ad hoc : le sens, le choix de la valeur, l'impact sur l'organisation, l'expérimentation et enfin le déploiement. L'étape « choix de la valeur » intégrant les décisions sur la création et le partage de la valeur autour des « 3P ».

### Être aussi en phase avec son style de leadership

Deux styles de leadership sont en particulier utiles dans un tel projet. Le style « visionnaire », dans lequel le dirigeant a une idée claire et ambitieuse du positionnement de l'entreprise en termes d'IA et de RSE. Sa problématique est d'assurer l'adhésion des équipes pour sa mise en œuvre. L'autre style est celui du leader « éthique ». Celui-ci table sur l'intelligence des collaborateurs pour trouver les réponses. Il se concentre plutôt sur leur autonomisation et le respect des valeurs communes. Chaque style est adapté à un contexte différent et chaque dirigeant doit aussi comprendre sa préférence.

### La conscience avant la science

L'IA peut ruiner les réponses, même imparfaites, que les entreprises apportent aux besoins humains : sécurité, appartenance, réalisation de soi ... et donc profondément démotiver les équipes. Tandis que la RSE contribue au contraire à ces réponses. Les dirigeants ont donc tout intérêt à mettre l'intelligence artificielle au service de la recherche de la performance durable. Ceci redonnera aussi du sens et de la fierté aux équipes, et donc renforcera leur engagement pour se battre face à l'adversité actuelle.

*Damien Ribon*  


<sup>1</sup> CultureWay 2022 – Aktios Conseil



En savoir plus sur l'adhésion ?  
Proposer une mission de conseil aux membres d'XMP-Consult ?  
Participer à nos événements ?

CONSULTEZ NOTRE SITE 

[info@xmp-consult.org](mailto:info@xmp-consult.org)



**xmp**  
consult

Association loi 1901 ayant son siège au  
5 Rue Descartes 75005 PARIS