

BASCULE

**D'où vient l'IA, ce qu'elle change déjà
et ce qui reste à décider**



Olivier Le Roux

Tes devoirs, tes stages, ton premier job, tes images, tes infos, tes relations, ta façon même de penser : l'IA est déjà en train de tout déplacer.

L'IA n'est pas seulement un outil.

C'est une bascule.

Elle écrit, raisonne, crée et, maintenant, agit.

La question n'est plus de l'éviter :

c'est de t'en servir sans lui abandonner ton jugement.

Ce livre est ta carte, pour prendre le temps de comprendre, au-delà du bruit médiatique et des idées reçues.

D'où vient l'IA.

Ce qu'elle change déjà.

Ce qu'elle menace.

Ce qu'elle rend possible.

Et ce que, toi, tu vas devoir décider.

Sans cacher les dangers.

Sans céder à la panique.

Sans renoncer à l'ambition.

Né des questions des trois ados de l'auteur :

« Utiliser ChatGPT dans mes études, c'est tricher ? »

« Le métier que je veux faire existera-t-il encore dans dix ans ? »

« On aura bientôt un robot humanoïde à la maison ? »

« Une machine peut-elle devenir consciente ? »

Écrit par un docteur en informatique qui travaille chaque jour avec les IA les plus avancées du marché, et qui se demande, comme eux, vers quel monde nous allons.

**Tu vivras dans ce monde plus longtemps
que ceux qui le dirigent aujourd'hui.**

Le monde qui vient n'est pas écrit.

Autant apprendre à tenir le stylo.

Pour les 16–25 ans, d'abord.

Et pour les parents, les enseignants

et les curieux qui veulent comprendre avec eux.

Olivier Le Roux

www.basculer-livre.fr

ISBN 979-10-988556-0-3



9 791098 855603



Olivier Le Roux

BASCULE

D'où vient l'IA, ce qu'elle change déjà
et ce qui reste à décider

Version 1.0

16 octobre 2026

*À mon père,
qui un jour de 1988
a posé un Amstrad PC1512
sur la table du salon.*

Avant-propos

L'intelligence artificielle ne sera pas une technologie de plus dans ta vie.

Elle sera là dans tes études, tes stages, ton travail. Dans les textes que tu liras, les images que tu verras, les métiers que tu choisiras, les informations que tu croiras vraies ou que tu rejetteras. Peut-être aussi dans tes relations, tes doutes, tes conversations les plus intimes.

Tu peux la subir.

Tu peux t'en servir les yeux fermés.

Ou tu peux comprendre ce qu'elle change, et décider de ce que tu en fais.

En vérité, elle est déjà là depuis longtemps.

Quand tu ouvres TikTok, YouTube ou Instagram, tu crois choisir ce que tu vois. En partie seulement. L'ordre, le rythme, la vidéo suivante, le moment où tu restes scotché : tout cela est décidé par des systèmes que tu ne vois pas. Une IA silencieuse, que tout le monde appelle « l'algo », entraînée à retenir ton attention. Elle ne te connaît pas comme un ami ; elle te

mesure comme un comportement. Et elle a déjà façonné une part de tes goûts, de tes opinions, peut-être de ton sommeil.

Tu n'attends pas l'IA. Tu vis avec depuis l'enfance.

La nouveauté, ce n'est pas qu'elle arrive. C'est qu'elle te parle. Avant, elle choisissait tes vidéos. Maintenant, elle répond à tes questions, écrit avec toi, t'aide à travailler, à créer, à décider, parfois à ta place. Nous avons perdu le monopole du langage et, depuis peu, celui du raisonnement.

Ce livre parle de ça. De ta vie, pas seulement de technologie.

Je m'appelle Olivier. Docteur en informatique, je code depuis l'adolescence et je travaille depuis plus de vingt ans comme ingénieur. J'ai vu arriver Internet, Google, les smartphones, les réseaux sociaux, puis les IA génératives, et chaque fois la même rengaine : « un outil de plus », « encore une mode ». Parfois c'est vrai. Souvent, c'est une façon élégante de ne rien comprendre.

Aujourd'hui, je travaille avec l'IA plusieurs heures par jour, comme avec un collaborateur. J'aide mon entreprise à l'intégrer, à fixer des cadres qui permettent à l'IA de travailler de façon fiable et en autonomie.

J'ai aussi trois enfants, de douze, seize et dix-sept ans. C'est en grande partie pour eux que j'ai commencé ce livre, parce qu'ils vivent déjà les questions que tu te poses peut-être : utiliser ChatGPT dans mes études, est-ce tricher ? Quels métiers vont être bouleversés dans la prochaine décennie ?

Comment distinguer le vrai du faux ? Et sur le fond, vers quel monde allons-nous ? Qu'est-ce qui restera humain ?

Élèves, parents, enseignants, collègues : partout, le même décalage. Beaucoup connaissent une partie de l'histoire. Peu voient le tableau d'ensemble. Certains minimisent, d'autres dramatisent. La plupart manquent simplement d'une carte.

Ce livre pose cette carte. Pas une vérité officielle. Pas un mode d'emploi magique. Pas une promesse que tout ira bien. Une carte pour comprendre ce qui se passe, utiliser ces outils avec lucidité, et ne pas laisser d'autres décider à ta place.

Une chose à savoir avant de commencer : ce livre ne cachera rien.

Il parlera de triche scolaire, de solitude, d'amour pour une machine, de métiers qui vacillent, de progrès scientifiques, de deepfakes (ces fausses vidéos, voix ou images fabriquées par IA), de manipulation, d'armes autonomes, et d'intelligences qui pourraient nous dépasser et que nous ne saurions pas toujours contrôler.

Il ne te dira pas que tout va s'effondrer.

Il ne te dira pas non plus que tout ira bien.

Tu mérites qu'on te parle franchement du monde tel qu'il est. Et si, par moments, tu trouves que je penche trop d'un côté, vers la peur ou vers la complaisance, garde cette impression. Interroge-la. Vérifie. Contredis-moi. Ce livre n'est pas là pour que tu me croies. Il est là pour t'aider à penser.

Une dernière chose : ce livre a été écrit avec l'aide de plusieurs intelligences artificielles. Je ne vais pas faire semblant du contraire. Elles m'ont aidé à vérifier, comparer, chercher des sources, repérer des incohérences, reformuler (tu trouveras la liste à la fin).

Mais elles ne l'ont pas écrit à ma place. Le plan, les choix, les exemples, les arbitrages et la voix sont les miens. Une IA écrit vite, et souvent bien, mais elle ne sait pas, à ta place, ce qui mérite d'être dit. Elle ne sait pas ce que tu as vécu. Elle ne porte pas tes responsabilités. L'IA amplifie ce qu'on apporte. Elle ne fabrique pas ce qu'on n'a pas.

Le dire dès le début n'est pas un aveu, c'est une exigence de cohérence : je ne peux pas écrire un livre sur l'usage lucide de l'IA en cachant que je l'ai utilisée. Ce livre est donc aussi une démonstration de ce qu'il défend.

Ce livre est daté, octobre 2026, et c'est volontaire : parler de l'IA sans citer ce qui existe au moment où j'écris, ce serait parler de cinéma sans citer un seul film. Certains noms, certains chiffres seront vite dépassés. Le squelette, lui, tiendra : la trajectoire historique, les questions de fond, les principes de vérification, les choix qui s'imposent à toi.

Maintenant, on peut commencer.

La trajectoire qui mène à aujourd'hui, le présent qui change déjà tout, des questions difficiles, des réponses partielles et, à la fin, une invitation : ne pas rester spectateur de ta propre vie. Rassure-toi, ce n'est pas un manuel de développement personnel. Juste quelques pistes pour que tu te poses les bonnes questions.

Ce livre n'est qu'un début. Le reste t'appartient.

OLR

Comment lire ce livre

Ce livre est une carte. Tu peux la lire de deux façons.

Le chemin complet. Tu commences ici et tu avances dans l'ordre. C'est le parcours que je te conseille, parce que le livre suit une trajectoire. *D'où vient l'IA* (partie 1). *Ce qu'elle sait faire en 2026 et ce qu'elle change déjà dans ta vie* (partie 2). *Ce qui reste à décider, les grandes questions qu'elle ouvre* (partie 3). Puis on revient à toi : *qu'est-ce que tu en fais, maintenant ?* L'histoire du début n'est pas un décor : elle explique pourquoi l'IA d'aujourd'hui est bien plus qu'un gadget, et ne s'arrêtera pas là.

L'entrée rapide. Mais tu as peut-être une question qui te brûle *maintenant*. Alors entre par là : ce que l'IA sait vraiment faire en 2026 (chapitre 7, p. 57), les études et la triche (chapitre 8, p. 71), l'ami invisible (chapitre 9, p. 85), les métiers qui se recomposent (chapitre 10, p. 93), créer avec l'IA (chapitre 12, p. 121), les images qu'on ne peut plus croire (chapitre 13, p. 133), la fracture entre ceux qui guident la machine et ceux qui la subissent (chapitre 14, p. 141), et ce que, toi, tu peux faire (dernier chapitre, p. 237). Tout à la fin, sept gestes

pour dialoguer efficacement avec une IA (p. 251), et un petit lexique pour t’y retrouver (p. 253). Puis, quand tu voudras comprendre *pourquoi*, reviens au début. La carte t’attendra.

Les petits chiffres. Tu croieras dans les pages des chiffres en exposant, un peu au-dessus de la ligne : chacun renvoie à une source. Il y en a plus de trois cents, articles, études, décisions de justice. Elles sont toutes détaillées en ligne, avec leurs liens : www.bascule-livre.fr/sources. La page Sources (p. 261) explique ce choix.

Il n’y a pas de bonne ni de mauvaise porte. Il y a la tienne.



www.bascule-livre.fr/sources

PARTIE 1

D'où vient la bascule ?

Des automates à l'IA qui apprend et qui raisonne

Avant de te raconter l'histoire de l'intelligence artificielle, je précise ce que j'entends ici par **bascule**.

Une bascule n'est pas une simple amélioration. Ce n'est pas une application de plus, un gadget nouveau, une mode technologique qui occupe les médias quelques mois avant d'être remplacée par une autre.

Une bascule, c'est le moment où une société change de socle technique, ces inventions sur lesquelles tout le reste repose. Et donc, peu après, de règles du jeu.

L'écriture a été une bascule : la mémoire est sortie des cerveaux pour se déposer sur des tablettes, du papier, des livres. L'imprimerie a été une bascule : le savoir a cessé d'être copié à la main pour circuler massivement. L'électricité, Internet : à chaque fois, un nouveau socle technique a reconfiguré la vie quotidienne, le travail, la culture, la politique.

L'intelligence artificielle est une bascule du même ordre. Je la crois plus profonde encore, parce qu'elle touche à ce que nous associons spontanément à l'intelligence humaine : parler, raisonner, créer, agir, et depuis peu contribuer à des découvertes scientifiques.

Tout part donc d'une question simple : comment en sommes-nous arrivés là ?

Pour répondre, il faut remonter loin. Avant ChatGPT. Avant Internet. Avant même l'ordinateur moderne.

Cette histoire n'est pas une ligne droite. Elle est faite d'impasses, d'exagérations, de modes, d'échecs et de retours inattendus. Mais vue de loin, une trajectoire apparaît : la machine a d'abord prolongé le corps (les automates), puis le calcul (les machines à calculer), puis l'information (l'ordinateur), puis le langage (ChatGPT), et désormais le raisonnement.

Une bascule pareille ne tombe jamais du ciel.

Elle vient de très loin.

Le vieux rêve : imiter, calculer, manipuler des symboles

En 1770, dans les cours d'Europe, une machine semble jouer aux échecs contre des humains — et gagner. On l'appelle le Turc mécanique : un automate enturbanné, présenté par son inventeur, Wolfgang von Kempelen, exhibé pendant des décennies en Europe et en Amérique. Le public est fasciné : une machine pourrait-elle rivaliser avec l'intelligence humaine ?¹

On découvrira que le Turc était une illusion : un joueur humain était dissimulé à l'intérieur. Mais l'épisode reste révélateur. Bien avant l'ordinateur, bien avant l'IA, le public était déjà prêt à croire, ou au moins à vouloir croire, qu'une machine pouvait penser.

Ce rêve ne datait pas de 1770. Avant de rêver d'une machine qui pense, les humains ont rêvé d'une machine qui bouge. Depuis l'Antiquité, nous fabriquons des dispositifs capables d'imiter certains gestes du vivant : ouvrir une

porte, faire couler de l'eau, déplacer une figurine, donner l'illusion qu'un objet agit par lui-même. Héron d'Alexandrie au I^{er} siècle, al-Jazari au XIII^e, les grandes horloges astronomiques d'Europe, les automates de Vaucanson au XVIII^e : à chaque époque, des artisans construisent des machines qui troublent.²

L'automate trouble parce qu'il brouille une frontière. Il n'est pas vivant, mais il se comporte un instant comme s'il l'était. Il n'a pas de volonté, mais il donne l'impression d'en avoir une.

Cette hésitation est au cœur de notre rapport aux machines intelligentes.

Notre fascination pour l'intelligence artificielle ne vient pas seulement de ce que les machines savent faire. Elle vient aussi de ce que nous projetons sur elles.

Je le constate chaque jour en travaillant avec ces outils. Quand une IA conversationnelle nous répond avec fluidité, nuance ou empathie apparente, nous savons rationnellement qu'il n'y a pas un humain derrière l'écran. Mais une autre partie de nous réagit différemment. Si la réponse est pertinente, si le ton semble adapté, alors notre cerveau social s'active. Nous avons l'impression d'un interlocuteur. Pas forcément d'une personne, mais d'une présence.

C'est ce que faisait déjà l'automate : produire une apparence sans posséder la chose.

La différence est que l'automate ancien imitait surtout le geste. L'IA moderne imite surtout le langage. Et le langage nous touche beaucoup plus profondément que le mouvement.

Un canard mécanique peut nous amuser.
Une machine qui nous répond peut nous troubler.

Mais imiter ne suffit pas. Le Turc mécanique le rappelait déjà. Pour aller au-delà de l'apparence, il faut formaliser. Décomposer. Rendre une opération mentale assez précise pour qu'on puisse, un jour, la confier à autre chose qu'un cerveau humain.

Cette idée a un nom. Et ce nom vient de loin.

Au ix^e siècle, à Bagdad, à l'âge d'or des sciences arabes, un mathématicien lié à la Maison de la Sagesse, le grand foyer savant de la ville, compose deux ouvrages qui changeront durablement la pensée. Il s'appelle Muḥammad ibn Mūsā al-Khwārizmī.

Le premier ouvrage expose des méthodes pour résoudre des équations. Son titre donnera son nom à l'algèbre : un langage pour raisonner sur ce qu'on ne connaît pas encore, l'inconnue, ce x qu'on t'a fait manipuler dès le collège.

Le second ouvrage porte une idée plus humble en apparence, mais plus décisive pour notre histoire. Il fait connaître dans le monde arabe les chiffres indiens (notre 0, 1, 2...9) et décrit avec eux des procédures de calcul (additionner, soustraire, multiplier, diviser) qu'on peut suivre étape par étape, sans avoir besoin de comprendre pourquoi elles marchent. C'est cette seconde idée qui portera son nom.

L'idée paraît évidente aujourd'hui parce qu'elle est partout. À l'époque, elle ne l'était pas. Al-Khwārizmī formalise quelque chose de profond : un calcul n'est pas seulement un résultat, c'est une **méthode**. Une suite d'étapes qui, ap-

pliquées correctement, donnent toujours la bonne réponse, indépendamment de qui les exécute.³

Quand son nom est traduit en latin médiéval, il devient *Algoritmi*. Et de ce mot naît un autre mot, qui structurera plus tard toute l'informatique : **algorithme**.

Une recette de cuisine peut être décrite comme un algorithme. Une procédure de calcul bien définie est un algorithme. Une démonstration mathématique peut être vérifiée, étape par étape, par un algorithme. Un programme informatique met en œuvre des algorithmes. Et, beaucoup plus loin sur la même ligne, ChatGPT lui-même est gouverné par des algorithmes.

Pourquoi cette idée est-elle si importante ?

Parce qu'elle prépare tout le reste : si une procédure peut être écrite assez précisément, alors elle peut être suivie par n'importe quel exécutant. Un élève, un scribe, un employé. Et un jour, une machine.

L'algorithme n'est pas une machine. C'est une idée. Mais sans elle, pas de machines à calculer modernes. Sans ces machines, pas d'informatique. Et sans informatique, pas d'IA.

Mille ans avant qu'une machine puisse exécuter un algorithme, un homme à Bagdad a écrit que le calcul pouvait être une recette. L'une des intuitions les plus discrètes et les plus puissantes de notre histoire.

Imiter, c'est la première intuition. Écrire la recette, c'est la seconde. La troisième commence avec Pascal : confier la recette à une mécanique.

Au début de 1642, Blaise Pascal a dix-huit ans. Pour aider son père, chargé de tâches fiscales, il conçoit une machine à roues dentées capable d'additionner et de soustraire : la Pascaline.⁴ L'idée flottait dans l'air : vingt ans plus tôt, un astronome allemand, Wilhelm Schickard, avait conçu une « horloge à calculer », oubliée puis redécouverte au xx^e siècle.⁵ Mais la Pascaline est la première dont l'existence est solidement attestée : elle a été fabriquée en plusieurs exemplaires, et certains nous sont parvenus.

Elle ne pense pas. Elle ne comprend pas les nombres. Mais elle fait autre chose : elle retire une petite opération mentale du cerveau humain pour la confier à une machine. Le geste te paraît banal. Tu le fais chaque fois que tu sors ton téléphone pour un calcul. En 1642, il était inouï. L'écriture, l'abaque (le boulier), les tables soulageaient déjà la mémoire ; avec la Pascaline, c'est l'opération elle-même qui passe à la mécanique.

Une partie du calcul devient mécanique. Une partie de l'erreur humaine devient évitable. Une partie du travail intellectuel peut être confiée à des machines.

Au xix^e siècle, Charles Babbage pousse l'idée beaucoup plus loin. Avec sa machine analytique, il imagine non plus une machine spécialisée, mais une machine générale, capable de recevoir des instructions, de stocker des valeurs, de répéter des opérations, de suivre une forme de programme.⁶

Babbage tient là l'idée même de l'ordinateur : la même machine peut devenir plusieurs machines.

Ada Lovelace travaille avec Babbage et entrevoit la suite : si des nombres peuvent représenter autre chose que des quantités (des lettres, des notes, des symboles, des formes), alors

une machine à nombres peut devenir une machine à symboles.⁷ Pour les lettres, il suffit de convenir que A vaut 1, B vaut 2.

Cette intuition de 1843 est ton quotidien : un texte, une image, un son, une vidéo, une conversation, tout cela peut devenir une suite de nombres. C'est cela, la numérisation. Pas seulement « mettre sur ordinateur ». Transformer des réalités très différentes en informations manipulables par une même famille de machines.

L'ordinateur moderne devient ainsi la machine universelle de la manipulation des symboles. Et c'est cette universalité qui prépare l'IA : pour qu'une machine apprenne à écrire, traduire, reconnaître ou répondre, il faut d'abord que le monde soit devenu manipulable sous forme numérique.

L'IA ne sort pas de nulle part. Elle naît dans un monde déjà converti en informations.

D'abord l'algorithme. Puis le calcul mécanique. Puis le stockage. Puis la manipulation de symboles. Puis la numérisation du monde.

Quand une opération mentale sort ainsi du cerveau humain, l'esprit ne disparaît pas. Il s'équipe. Il se prolonge. Il délègue, puis se réorganise autour de ces machines.

C'est ce qui se passe aujourd'hui avec l'intelligence artificielle.

La Pascaline a déplacé une petite frontière : le calcul arithmétique.

L'ordinateur a déplacé une frontière immense : la manipulation de l'information.

L'intelligence artificielle déplace maintenant une frontière plus intime encore : celle du langage et du raisonnement.

C'est pour cela qu'elle nous trouble.

Elle ne touche plus seulement à nos bras. Elle ne touche plus seulement à nos calculs. Elle commence à toucher à ce que nous appelons penser.

Turing : entraîner plutôt que programmer

Au milieu du xx^e siècle, une question devient inévitable. Si une machine peut calculer, stocker, manipuler des symboles, peut-elle un jour faire quelque chose qui ressemble à penser ?

Alan Turing est un mathématicien britannique, l'un des pères de l'informatique. Pendant la Seconde Guerre mondiale, il a aidé à percer Enigma, la machine à chiffrer de l'armée allemande. En 1950, il publie un article devenu célèbre, *Computing Machinery and Intelligence* (« Machines à calculer et intelligence »), qui s'ouvre sur une question simple : « les machines peuvent-elles penser ? »⁸

Mais Turing se méfie aussitôt de sa propre question. Le mot « penser » est trop chargé. Trop vague. Chacun peut y mettre ce qu'il veut : conscience, raisonnement, langage, intention, émotion, âme.

Turing propose donc de remplacer la question par une expérience plus concrète.

Imagine une conversation à distance, par écrit. Un humain dialogue avec deux interlocuteurs qu'il ne voit pas. L'un est humain, l'autre est une machine. Si, après un certain temps, l'humain ne parvient plus à distinguer de manière fiable lequel est lequel, alors, que devons-nous conclure ?

C'est le **jeu de l'imitation**, que la postérité a rebaptisé « test de Turing ».

L'idée est puissante parce qu'elle déplace le débat. Au lieu de demander « la machine pense-t-elle vraiment ? », Turing demande, en substance : « peut-elle se comporter, dans une conversation, d'une manière suffisamment proche d'un humain pour que nous ne puissions plus la distinguer ? »

Le test ne prouve pas qu'une machine est consciente, ni qu'elle ressent quoi que ce soit. Mais il pose une question redoutable : si une machine explique, répond, argumente, plaisante, se corrige et semble comprendre le contexte, alors sur quelle base allons-nous dire qu'elle ne manifeste *aucune* forme d'intelligence ?

Cette question est encore la nôtre.

Mais le point le plus visionnaire chez Turing n'est pas le test lui-même. C'est son intuition sur l'apprentissage.

Turing comprend qu'il serait probablement absurde d'essayer de programmer directement une intelligence adulte complète. Aucun humain ne pourrait écrire à la main toutes les règles nécessaires pour parler, raisonner, traduire, plaisanter, expliquer une erreur, s'adapter à un interlocuteur.

Un enfant ne naît pas avec un manuel complet du monde : il apprend. Il imite, essaie, échoue, recommence, reçoit des corrections, généralise.

Turing prend alors le problème par l'autre bout : au lieu d'essayer de construire une machine qui simule l'esprit adulte, pourquoi ne pas construire une machine qui ressemble plutôt à un esprit d'enfant, puis l'éduquer ?

C'est l'idée de la **machine-enfant**. Une intelligence artificielle digne de ce nom ne sera pas entièrement programmée : elle devra être entraînée.

C'est le principe au cœur de l'IA moderne.

Un grand modèle de langage, le moteur derrière ChatGPT ou Gemini, n'est pas écrit ligne par ligne comme un logiciel classique. On lui donne une architecture, c'est-à-dire un plan de construction. On lui donne des données. On lui donne de la puissance de calcul. On lui donne des objectifs d'apprentissage. Puis on l'entraîne.

Le modèle ajuste progressivement des milliards de paramètres : les nombres internes que l'entraînement va caler, un peu comme on règle des milliards de petits curseurs jusqu'à obtenir un comportement satisfaisant. Il apprend des régularités dans les textes, les images, les conversations. Il construit ses propres représentations internes à partir de ce qu'il a vu, une sorte de carte intérieure que même ses concepteurs ont du mal à lire.

Cette différence entre **programmer et entraîner** est l'une des clés pour comprendre notre époque. Programmer, c'est écrire les règles : si telle condition est vraie, fais ceci. Entraîner,

c'est concevoir le cadre dans lequel la machine va apprendre, à partir d'exemples.

C'est un changement de nature.

Il explique pourquoi les IA modernes sont parfois si difficiles à comprendre. Quand un logiciel classique se trompe, on peut chercher la règle fautive. Quand une IA entraînée se trompe, le problème est plus diffus : il vient des données, des représentations internes, du contexte, de la manière même dont la question est posée.

Cela ne dit rien de la conscience, ni de la compréhension. Mais cela change notre rapport aux machines : nous ne leur dictons plus ce qu'elles savent faire, nous le découvrons après les avoir entraînées.

Turing avait vu la direction il y a plus de soixante-quinze ans. Pour fabriquer une machine intelligente, il faudrait moins lui écrire une pensée que lui apprendre à apprendre.

Le pari des neurones

Pendant longtemps, on a cru qu'il suffirait, pour rendre une machine intelligente, de lui écrire une à une toutes les règles du monde. Avant que les machines qui apprennent ne donnent raison à Turing, c'était la grande voie de l'intelligence artificielle.

L'idée est simple. Si un expert humain sait diagnostiquer une maladie ou identifier une panne, on peut tenter de transformer son savoir en règles logiques :

si la voiture ne démarre pas et que les phares restent éteints, alors vérifier la batterie.

Cette approche a donné naissance aux systèmes experts, surtout dans les années 1970 et 1980. Dans des domaines bien délimités, avec des règles explicites, ils pouvaient produire des résultats utiles.⁹

Mais ils avaient une limite profonde : le monde réel est rarement aussi propre qu'une base de règles. Les situations

sont ambiguës, les données incomplètes, les exceptions nombreuses, les contextes changent. Plus on voulait couvrir de cas, plus les systèmes devenaient lourds et fragiles.

À un moment, la voie des règles atteint un mur.

Comment écrire toutes les règles qui définissent un visage ? Toutes celles qui permettent de comprendre l'ironie ? De reconnaître un chat dans une image floue, prise de profil, dans la pénombre ?

Il faut alors changer de méthode.

Au lieu d'écrire toutes les règles, on va entraîner des systèmes capables d'en découvrir une partie à partir d'exemples. C'est le passage de **l'intelligence écrite à la main** à **l'intelligence apprise**.

Mais avant que l'apprentissage ne s'impose, l'IA programmée va connaître son moment le plus spectaculaire — non plus dans les règles savantes des systèmes experts, mais dans la puissance brute du calcul.

En 1997, Deep Blue, une machine spécialisée conçue par IBM, le géant américain de l'informatique, bat Garry Kasparov, champion du monde d'échecs et considéré comme l'un des plus grands joueurs de tous les temps.¹⁰

Plus de deux siècles plus tard, le Turc mécanique, ce faux automate du chapitre 1, avait enfin tenu sa promesse, mais cette fois sans humain caché.

Pour mesurer le choc, pense à ce que représentaient les échecs : un symbole d'intelligence, de stratégie, d'anticipation, de maîtrise froide.

Mais Deep Blue ne pense pas comme Kasparov. Il ne comprend pas les échecs comme un humain. Il n'a pas d'intuition vécue, pas de stress, pas de peur de perdre. Il examine plus de cent millions de positions par seconde, et donne à chacune une note selon des critères choisis par des humains : les pièces qui restent, leur place sur l'échiquier, la sécurité du roi. Il évalue, compare, avance dans un univers parfaitement défini : un plateau, des pièces, des règles fixes, un objectif clair.

Les échecs sont un monde fermé. Cela ne veut pas dire qu'ils sont simples. Le nombre de positions possibles est gigantesque. Mais les règles sont peu nombreuses, claires et complètes : on peut toutes les écrire pour une machine.

Deep Blue est le sommet d'une certaine vision de l'IA : une intelligence écrite à la main, qui explore des masses de coups grâce à une énorme puissance de calcul.

Il bat Kasparov. Mais il ne peut pas expliquer à un enfant pourquoi les échecs sont difficiles. Il ne peut pas commenter un poème, analyser une image, apprendre seul un autre jeu.

Deep Blue est extraordinairement fort dans un domaine précis. Il est presque totalement inutile en dehors.

Et pourtant, il déplace une frontière psychologique. Avant Deep Blue, beaucoup pensaient encore que certains sommets de l'intelligence humaine resteraient hors d'atteinte. Après, il devient clair qu'une machine peut battre les meilleurs humains dans des activités intellectuelles.

Mais cette victoire laisse une question ouverte : peut-on construire une machine qui ne soit pas seulement meilleure

que nous dans un jeu fermé, mais capable d'apprendre, de s'adapter et de transférer ses compétences ?

C'est ici que l'histoire bifurque.

Pendant que Deep Blue triomphe, une autre voie continue, plus discrète, plus contestée. Celle des réseaux de neurones.

L'idée est ancienne : construire des machines inspirées du cerveau. Pas des programmes rigides écrits règle par règle, mais des systèmes capables d'apprendre à partir d'exemples.

L'analogie avec le cerveau demande de la prudence. Un réseau de neurones artificiels n'est pas un cerveau miniature. Il n'a ni biologie vivante, ni conscience connue. Mais l'idée générale tient : comme dans le cerveau, aucune pièce prise seule n'explique ce que fait l'ensemble.

Un neurone numérique, c'est un tout petit calcul. Il reçoit des nombres, donne plus ou moins d'importance à chacun, fait la somme et transmet un résultat aux neurones suivants. Rien de magique.

Ce qui compte, ce sont ces degrés d'importance, qu'on appelle les poids, ou paramètres : un par connexion, comme les petits curseurs du chapitre 2. Pendant l'entraînement, on les ajuste tous ensemble, peu à peu, pour que le réseau se trompe de moins en moins.

Un réseau compte des millions de connexions, les plus grands des milliards, chacune avec son poids. Le comportement global dépend de l'ensemble.

Les données entrent d'un côté. Elles traversent plusieurs couches de neurones, comme les étages d'un immeuble. Chaque couche transforme un peu l'information. À la fin, le

réseau produit une réponse : « sur cette photo, il y a un chat », le mot qui vient ensuite dans une phrase, une image nouvelle.

Personne n'écrit une règle disant : *si tu vois deux oreilles pointues et des moustaches, c'est un chat*. C'est la distinction du chapitre précédent : on ne programme pas le réseau, on l'entraîne.

On lui montre des images de chats et de chiens. Il se trompe. On mesure son erreur. On modifie légèrement ses paramètres. Il recommence. Il se trompe moins. À force de recommencer, il apprend.

Longtemps, cette approche déçoit. Les réseaux sont trop petits. Les ordinateurs ne sont pas assez puissants. Les données disponibles sont limitées. À plusieurs reprises, l'enthousiasme retombe. Les financements se réduisent. On parle d'hiver de l'IA.¹¹

L'hiver n'est pas la mort. C'est une période où la croissance devient moins visible. La surface paraît figée. Mais sous la terre, quelque chose continue de travailler.

Pendant ces années moins spectaculaires, certains chercheurs continuent à croire aux réseaux de neurones. Parmi eux, trois noms deviennent incontournables : le Britannique Geoffrey Hinton, le Canadien Yoshua Bengio et le Français Yann LeCun.

Ils n'ont pas inventé seuls toute l'IA moderne : la science ne fonctionne jamais comme cela. Mais ils ont joué un rôle majeur dans le retour des réseaux de neurones et la construction de ce qu'on appelle aujourd'hui l'apprentissage profond, profond parce que les réseaux empilent beaucoup de couches.

Trois éléments finissent par tout changer. Beaucoup plus de données : Internet, les images numériques, les textes en ligne. Beaucoup plus de calcul : les puces des cartes graphiques (les GPU), conçues d'abord pour afficher les images des jeux vidéo, se révèlent très efficaces pour entraîner des réseaux. De meilleures méthodes : nouvelles façons d'organiser les réseaux, techniques d'entraînement plus efficaces.

Ce qui semblait bloqué se met à fonctionner.

Le moment de bascule publique arrive en 2012. Cette année-là, un réseau appelé AlexNet remporte la compétition ImageNet, l'un des grands concours de reconnaissance d'images. Le principe : on donne à des systèmes des images, ils doivent reconnaître ce qu'elles contiennent. Un chien, une voiture, un oiseau.

Cette tâche paraît naturelle pour un humain. Un enfant reconnaît un chat sans réfléchir. Une machine, elle, ne voit qu'une grille de pixels. Et le même objet peut changer de taille, de couleur, de position, d'éclairage.

Avant AlexNet, les systèmes de vision reposaient sur des méthodes où les humains concevaient eux-mêmes les caractéristiques importantes à détecter : contours, textures, formes. Beaucoup de travail à la main.

AlexNet change la logique. Le réseau apprend lui-même une partie de ces représentations. Ses premières couches détectent des formes simples. Les suivantes combinent ces formes en motifs plus complexes. Plus on avance, plus ce que le réseau repère est élaboré : des contours, puis des yeux, puis des visages.

Mais AlexNet n'aurait pas eu la même portée sans son terrain d'entraînement : ImageNet, une immense base d'images, chacune étiquetée de ce qu'elle montre, qui a permis d'entraîner et de comparer les modèles. Et derrière ImageNet, une autre figure clé : la chercheuse américaine Fei-Fei Li.¹²

C'est la combinaison décisive : **algorithmes + données + calcul**.

Sans l'équipe de Hinton, pas d'AlexNet. Sans Fei-Fei Li et ImageNet, pas de terrain d'entraînement à cette échelle. Sans les puces des cartes graphiques, pas d'entraînement assez rapide.

Après 2012, beaucoup de laboratoires comprennent que les réseaux de neurones ne sont plus une vieille promesse. Ils deviennent la méthode dominante.

Hinton, Bengio et LeCun reçoivent le prix Turing 2018 (souvent présenté comme l'équivalent du Nobel pour l'informatique) pour leurs contributions fondatrices à l'apprentissage profond. Ce qui avait longtemps été une approche contestée devient le cœur de l'IA moderne.¹³

Cette histoire est importante pour comprendre ChatGPT. Il n'est pas tombé du ciel. Il a fallu l'intuition des réseaux de neurones, l'apprentissage profond, puis assez de calcul et de données pour des modèles géants.

Les grandes ruptures visibles reposent souvent sur des décennies de travail invisible.

AlphaGo, AlphaFold : voir ce qu'on ne voyait pas

Après AlexNet, l'apprentissage profond cesse d'être une promesse marginale. En 2016, un nouvel événement change encore la perception du public : AlphaGo, développé par DeepMind, bat Lee Sedol, l'un des meilleurs joueurs de go au monde, par 4 victoires à 1.¹⁴

DeepMind est alors un laboratoire britannique fondé à Londres en 2010 puis racheté par Google en 2014. À sa tête, Demis Hassabis, parcours rare : enfant prodige des échecs, créateur de jeux vidéo, puis neuroscientifique, puis entrepreneur.¹⁵

Pour saisir le choc AlphaGo, commençons par le go : un jeu chinois vieux de plus de deux mille ans, où l'on pose des pierres noires et blanches pour encercler des territoires.

Les échecs sont déjà très complexes. Le go l'est encore davantage. Le plateau est plus grand, les possibilités sont immenses, et il est beaucoup plus difficile d'évaluer clairement

une position. Une partie ne se gagne pas seulement par calcul tactique. Elle exige une forme d'intuition : sentir les équilibres, les territoires, les influences à long terme.

Pendant longtemps, le go semblait résister aux machines. Deep Blue avait battu Kasparov aux échecs en examinant des millions de positions par seconde. Pour le go, cette approche ne suffisait pas. L'espace des possibilités était trop vaste.

AlphaGo représente autre chose. Ses réseaux de neurones apprennent d'abord en étudiant des parties humaines, puis en jouant contre eux-mêmes, encore et encore, récompensés à chaque victoire : c'est l'apprentissage par renforcement. La machine cherche aussi parmi les coups possibles, mais sans tout explorer brutalement : elle apprend à repérer les positions prometteuses.

Autrement dit, AlphaGo ne joue pas comme Deep Blue. Il ne remplace pas seulement l'intuition humaine par plus de calcul. Il construit une forme d'**intuition artificielle** à partir d'exemples et de parties simulées.

Le moment le plus célèbre arrive lors de la deuxième partie : le **coup 37**.

Ce coup paraît si étrange que les commentateurs croient d'abord à une erreur. Selon les propres calculs d'AlphaGo, un joueur humain n'avait qu'environ une chance sur dix mille de le jouer. Trop loin des habitudes, vieilles de plusieurs millénaires. Puis, au fil de la partie, il se révèle brillant.

C'est là que quelque chose change. La machine ne fait pas seulement ce que les humains font déjà, plus vite. Elle propose une solution différente, inattendue, presque étrangère à notre manière de jouer.

Faut-il appeler cela de la créativité ?

La réponse dépend du sens du mot. AlphaGo ne ressent rien. Il n'a pas l'intention de surprendre. Il ne cherche pas la beauté. Mais il produit une décision nouvelle, efficace, non évidente pour les meilleurs joueurs, et que les humains peuvent ensuite étudier, admirer, intégrer.

Ce n'est pas la créativité humaine. Mais ce n'est pas non plus une simple répétition.

Deep Blue dit : *une machine peut battre un champion en cherchant mieux dans un monde fermé.*

AlphaGo ajoute : *dans un espace trop vaste pour être exploré jusqu'au bout, une machine peut découvrir des stratégies qu'aucun humain n'avait trouvées.*

Cette idée va dépasser le jeu.

Si une machine peut apprendre à explorer un espace comme celui du go, pourquoi ne pourrait-elle pas en explorer d'autres ? Des molécules. Des protéines. Des matériaux. Du code.

C'est ici qu'AlphaFold prend le relais.

DeepMind applique l'apprentissage profond à un problème scientifique majeur : prédire la forme en trois dimensions des protéines, les molécules qui font l'essentiel du travail dans nos cellules. Le problème résistait depuis cinquante ans : en laboratoire, trouver la forme d'une seule protéine prenait souvent des années, quand on y parvenait. En 2020, la deuxième version, AlphaFold 2, le résout. Quelques minutes suffisent désormais pour une protéine, et deux ans plus tard, on disposait de la forme de presque toutes les pro-

téines connues, plus de 200 millions.¹⁶ La forme d'une protéine détermine largement sa fonction dans le vivant. Comprendre cette forme peut accélérer la biologie, la médecine, la recherche pharmaceutique.

Le prix Nobel de chimie 2024 récompense cette percée : David Baker pour la conception de nouvelles protéines par ordinateur, Demis Hassabis et John Jumper pour la prédiction de leur forme en trois dimensions.¹⁷

Ce Nobel ne signifie pas que l'IA « comprend » la biologie comme un chercheur humain. Il signifie qu'elle n'a pas besoin de comprendre comme nous pour faire avancer la science : elle devient un instrument capable d'attaquer des problèmes qui résistaient depuis des décennies.

Reste à passer de la découverte au remède. En septembre 2026, un médicament est entré dans la dernière phase d'essais sur des patients, contre une maladie grave des poumons. Selon la start-up Insilico, l'IA a trouvé à la fois sa molécule et sa cible, ce sur quoi il agit dans le corps.¹⁸ Il lui faudra encore des années pour arriver, peut-être, en pharmacie : l'IA accélère la découverte, pas les essais.

Avec AlphaGo, l'IA explore autrement un espace stratégique. Avec AlphaFold, elle devient un outil de découverte. Dans les deux cas, la leçon est la même : quand un problème possède une structure cachée, trop vaste pour l'esprit humain, l'apprentissage profond peut révéler des régularités invisibles à notre intuition.

L'IA ne se contente plus de battre l'humain à ses propres jeux. Elle commence à prolonger la recherche humaine dans des espaces où notre intuition seule ne suffit plus.

2017 : l'invention qui rend ChatGPT possible

Pour arriver à ChatGPT, il ne suffisait pas d'avoir des réseaux de neurones, des données et des puces puissantes. Il fallait aussi une idée nouvelle pour traiter le langage.

Car le langage est difficile pour une machine.

Un mot ne veut presque jamais dire quelque chose tout seul. Son sens dépend de la phrase, du contexte, parfois de ce qui a été dit plusieurs lignes avant.

Le mot « banc » ne veut pas dire la même chose dans *un banc public*, *un banc de poissons* ou *un banc d'essai*. Et dans une phrase comme « *Paul parle à Lucas parce qu'il est inquiet* », qui est inquiet ? Paul ou Lucas ?

Pour répondre, il ne suffit pas de lire les mots un par un. Il faut comprendre les relations entre eux.

En 2017, huit chercheurs de Google publient un article qui fera date : *Attention Is All You Need* (en substance, « l'attention suffit »).¹⁹

Le mot important ici est **attention**.

Il ne s'agit pas d'attention comme chez un humain. Le modèle ne se concentre pas consciemment. Mais il apprend à repérer quelles parties d'un texte comptent les unes pour les autres.

Dans une phrase, certains mots doivent être reliés. Dans un paragraphe, certaines idées dépendent d'idées précédentes. C'est le rôle du mécanisme d'attention.

On peut l'imaginer ainsi : au lieu de lire une phrase comme une simple file de mots, le modèle construit une sorte de carte des relations importantes. Qui parle de qui ? Quel mot explique quel autre ? Quel détail change le sens ?

Ses auteurs baptisent ce nouveau type de modèle le *Transformer*.

Après 2017, les modèles deviennent plus grands. Ils lisent davantage de textes. Ils progressent en traduction, en résumé, en génération, puis en code et en dialogue.

Le mot GPT vient directement de cette histoire : *Generative Pre-trained Transformer*. *Generative* parce que le modèle génère du texte. *Pre-trained* parce qu'il est d'abord entraîné sur d'immenses quantités de données. *Transformer* parce qu'il repose sur cette architecture apparue en 2017.

Le langage n'est pas une compétence ordinaire. C'est le support de l'école, du droit, de la science, du travail, de la

politique, de la mémoire, de la transmission, de la pensée organisée.

Quand une machine apprend à manipuler le langage, elle entre dans l'espace où les humains expliquent, argumentent, enseignent, commandent, racontent et décident.

Le Transformer est une charnière discrète entre deux mondes : celui des machines qui reconnaissent, et celui des machines qui dialoguent.

Fin 2022, ChatGPT arrive dans les mains du grand public. Pour des millions de personnes, c'est le choc. Une machine peut répondre dans n'importe quelle langue, expliquer un concept, écrire un texte, corriger une phrase, résumer un document.

Cette version de ChatGPT est loin d'être parfaite. Elle se trompe. Elle invente parfois. Mais elle donne soudain l'impression d'un dialogue naturel : l'ordinateur cesse d'être une machine à commandes pour devenir une machine à conversation.

Techniquement, ChatGPT s'appuie alors sur la famille GPT-3.5.²⁰

Avec le recul, ce modèle fait figure de première démonstration publique. Impressionnante, oui. Mais encore fragile. Encore limitée. Bien loin des modèles dont nous disposons maintenant.

En quelques années, des dizaines de versions et de familles de modèles se sont succédé. GPT-3.5 ne voyait pas, ne raisonnait pas, ne pouvait garder en tête que quelques pages. Les modèles actuels font tout cela couramment.

Et c'est là que beaucoup de malentendus commencent.

J'ai vu beaucoup d'adultes juger l'IA après cinq minutes avec un chatbot gratuit. Ils posaient une question vague, obtenaient une réponse correcte mais moyenne, repéraient une erreur, puis concluaient : « *ce n'est pas si impressionnant* ».

Je comprends cette réaction. Mais elle vieillit très vite : en IA, quelques mois suffisent à changer de génération, trois ans à changer d'ordre de grandeur.

À partir de 2023, GPT-4 marque un premier grand saut public. Ce modèle suit mieux les consignes, écrit mieux, hallucine moins souvent — c'est-à-dire invente moins de choses fausses avec aplomb.²¹

Un chiffre a beaucoup circulé : sur une simulation de l'examen du barreau américain (le concours pour devenir avocat aux États-Unis), GPT-4 aurait obtenu un score « dans le top 10 % » des candidats, quand la version précédente, GPT-3.5, restait tout en bas du classement.

Ce chiffre est bien moins solide qu'il n'en a l'air.

Un chercheur du MIT a réexaminé l'étude originale et montré que tout dépendait du groupe de comparaison. Comparé à tous ceux qui passaient l'examen pour la première fois, GPT-4 faisait mieux que six candidats sur dix ; comparé à ceux qui le réussissaient, il se situait au milieu du classement, un candidat moyen, loin des 10 % les meilleurs.²²

Deux leçons. D'abord, le progrès réel reste considérable : passer du fond du classement au milieu, pour un modèle qui n'a pas été entraîné spécifiquement pour cet examen, ce n'est pas rien, et les meilleurs modèles de 2026 figurent dé-

sormais dans le haut du classement sur ce type d'examen.²³ Ensuite, et surtout, cette histoire illustre un piège permanent : les chiffres promotionnels. Des annonces spectaculaires circulent, sont reprises sans vérification, deviennent des évidences. Garder une exigence de précision face à ces chiffres est l'une des compétences essentielles pour ne pas se faire balader.

En 2023, avec GPT-4, on pouvait encore dire : « *l'IA écrit bien, mais elle ne sait pas réfléchir* ». Elle se trompait souvent. En 2026, ce n'est plus défendable : elle ne réfléchit pas comme nous, mais sur beaucoup de tâches elle se trompe rarement, rivalise avec les experts dans un nombre croissant de domaines, les dépasse déjà sur certains, et commence à contribuer significativement à la recherche scientifique. Entre les deux sont arrivés les modèles de raisonnement ; on y vient.

Les capacités ne sont pas seules à bouger : le peloton de tête change, lui aussi.

Au moment où j'écris ces lignes, ChatGPT (OpenAI) reste le nom que tout le monde connaît : il a lancé la vague en 2022 et donné son nom au geste lui-même, « demander à ChatGPT ». Et ces outils se sont diffusés deux fois plus vite que l'ordinateur personnel ou Internet en leur temps : deux ans après la sortie de ChatGPT, près de quatre Américains sur dix en âge de travailler s'en servaient déjà.²⁴ Il n'est pourtant plus seul. D'autres familles de modèles ont émergé en parallèle : Claude (Anthropic), Gemini (Google DeepMind), Mistral AI (français), Grok (de SpaceX, l'entreprise d'Elon Musk)²⁵, et plusieurs familles de modèles chinois, dont DeepSeek est la plus connue. Cette liste évolue de mois en mois. C'est tout le sujet.

Un modèle qui semblait extraordinaire peut être dépassé en quelques semaines, un modèle venu d'un autre pays surprendre les laboratoires installés, un modèle ouvert, téléchargeable par tous, rivaliser avec un modèle fermé. C'est très difficile à suivre, même pour les professionnels.

Cela explique une partie du décalage actuel : beaucoup de gens parlent encore de « ChatGPT » comme s'il s'agissait d'un produit unique et stable, alors que derrière ce mot se cache une succession rapide de modèles, de versions, d'outils. J'ai vu des études commandées par de grandes entreprises ou des gouvernements pour évaluer l'IA obsolètes dès leur sortie.

Dire « *j'ai testé ChatGPT* » ne suffit plus. La vraie question est : *quel modèle, quelle version, gratuit ou payant, en réponse rapide ou en prenant le temps de réfléchir, avec quels outils, pour quelle tâche ?* Sans ces précisions, on parle souvent dans le vide.

Trois ruptures méritent d'être nommées.

D'abord, le contexte long : la quantité de texte qu'un modèle peut garder en tête d'un coup. Au début, les modèles perdaient vite le fil. Ils pouvaient répondre à une question, mais avaient du mal à travailler sur un gros dossier. Aujourd'hui, les meilleurs modèles peuvent analyser des documents longs, comparer des passages, retrouver des contradictions, lire du code sur plusieurs fichiers. Une IA qui répond à une phrase est utile. Une IA qui lit un dossier entier devient un assistant de travail.²⁶

Ensuite, la multimodalité, un mot savant pour une idée simple : les modèles ne travaillent plus seulement sur du

texte. Ils reçoivent des images, des captures d'écran, des schémas, parfois de l'audio ou de la vidéo. Un élève peut prendre en photo un exercice de mathématiques et demander une explication. Un développeur peut partager une capture d'écran. Un designer peut faire analyser une maquette. L'IA ne se contente plus de lire ce qu'on tape. Elle interprète ce qu'on lui montre. Et elle produit à son tour : musique, voix, images, vidéos d'un réalisme qui aurait semblé invraisemblable il y a quelques années.²⁷

Enfin, et c'est la plus importante : les modèles de raisonnement. À partir de septembre 2024, OpenAI introduit une nouvelle génération de modèles, conçus pour travailler plus longtemps sur les problèmes difficiles.²⁸ Les autres grands laboratoires prennent rapidement la même direction : donner aux modèles plus de temps, plus d'étapes ou plus de calcul pour traiter les tâches difficiles. L'idée change : jusque-là, le modèle répondait d'un trait ; désormais, il peut prendre le temps de chercher *avant* de répondre.

La machine ne « pense » pas pour autant : elle n'a pas de voix intérieure consciente. Mais sur un problème formulé en mots, en nombres ou en code, elle peut désormais explorer plusieurs pistes, vérifier des étapes, corriger ses erreurs. C'est ce que j'appelle ici **raisonnement artificiel**.

L'intelligence apparente du modèle ne dépend plus seulement de son entraînement. Elle dépend aussi du temps qu'on lui laisse pour réfléchir.

C'est un changement important. Et il est arrivé vite. Si tu as l'impression que l'IA « *n'est pas si impressionnante* », vérifie d'abord avec quel modèle tu l'as testée. Et quand.

Personne ne leur a appris cela

Nous avons maintenant les briques : des machines qui calculent, puis des réseaux de neurones qui apprennent seuls, à voir, à jouer au go, à prédire la forme des protéines, puis à manier le langage. Reste à comprendre ce qui se passe quand on les fait grandir.

6.1 L'émergence : ce que personne n'a programmé

J'en arrive à une idée essentielle : **l'émergence**.

L'émergence, c'est l'apparition d'un comportement nouveau quand beaucoup d'éléments simples interagissent. Pris séparément, chaque élément paraît limité. Mais ensemble, ils produisent quelque chose de plus riche, difficile à prévoir.

Une fourmi seule n'a pas le plan de la fourmilière en tête. Elle suit des règles simples : chercher, transporter, déposer des signaux chimiques, réagir aux traces des autres.

Pourtant, une colonie construit des structures complexes, trace des chemins, répartit les tâches, protège un territoire.

Aucune fourmi ne comprend tout. Mais la colonie, elle, fait preuve d'une forme d'intelligence, qui naît du collectif et qu'aucune fourmi ne possède seule. C'est de l'émergence.

On retrouve une idée comparable avec le cerveau. Un neurone biologique pris seul ne pense pas. Mais des dizaines de milliards de neurones connectés produisent la mémoire, le langage, l'imagination, peut-être la conscience.

Cela ne veut pas dire qu'un réseau de neurones artificiels est un cerveau, ni qu'un grand modèle de langage est conscient. Cela veut seulement dire qu'il faut parfois regarder le système global pour comprendre ce qui apparaît.

6.2 Perroquet stochastique, ou carte interne ?

L'émergence répond à la critique qu'on entend le plus souvent contre l'IA. Les grands modèles de langage, ces IA qui font tourner ChatGPT et ses concurrents, reposent sur une idée qui semble presque trop simple : ils apprennent à prédire le prochain morceau de texte à partir du contexte.

C'est pourquoi on entend parfois : *« ce n'est pas intelligent, ça ne fait que prédire le mot suivant »*.

Cette critique a même un nom savant, tiré d'un article de recherche de 2021 sur les dangers des « perroquets stochastiques ».²⁹ « Stochastique » veut dire « gouverné par le hasard », plus exactement par des probabilités.

L'image du **perroquet stochastique** est devenue le drapeau d'une critique radicale des grands modèles de langage.

Elle dit ceci : ces modèles ne comprennent rien. Ils assemblent statistiquement des morceaux de texte vus pendant leur entraînement, sans rapport avec un sens réel, comme un perroquet répète des sons humains sans les comprendre.

L'expression a fait fortune. Presse, débats publics, plateaux télé. Elle a un mérite : rappeler que les modèles ne pensent pas comme nous. Et elle reste vraie sur un point précis : à la base, le modèle calcule bien des probabilités : il estime quel mot a le plus de chances de venir ensuite.

Mais elle peut aussi être trompeuse.

Dire qu'une IA « *ne fait que prédire le mot suivant* », c'est un peu comme dire :

un cerveau ne fait que transmettre des signaux électriques.

une chanson n'est qu'une suite de vibrations dans l'air.

une fourmilière n'est qu'un tas d'insectes.

À un certain niveau, c'est vrai. Mais prédire le mot suivant, c'est la fourmi ; ce qui apparaît au-dessus, c'est la fourmilière. Ne regarder que la fourmi, et en conclure qu'il n'y a pas de fourmilière : voilà l'erreur.³⁰

Pour prédire correctement la suite d'un texte, le modèle doit apprendre énormément de régularités : grammaire, styles, relations entre les idées, structures de raisonnement, conventions de dialogue, méthodes d'explication...

La prédiction du mot suivant n'oblige pas le modèle à comprendre le monde comme un humain. Mais elle le pousse à capturer des régularités si profondes que le résultat ressemble à de la compréhension.

À petite échelle, un modèle complète des phrases. À grande échelle, il commence à traduire, résumer, expliquer, imiter des styles, corriger des erreurs, raisonner par étapes, utiliser des outils.

Personne n'a écrit à la main une règle disant : *si quelqu'un t'annonce qu'il a raté son permis, console-le d'abord, puis aide-le à préparer le prochain*. Pourtant, le modèle peut le faire.

Mieux : dans certains cas contrôlés, on sait le montrer expérimentalement. En 2023, une équipe de chercheurs américains a entraîné un petit modèle, baptisé Othello-GPT, à prédire les coups d'une partie d'Othello, ce jeu de plateau où l'on retourne les pions de l'adversaire, noirs d'un côté, blancs de l'autre. Sans jamais lui expliquer les règles. Sans jamais lui montrer le plateau. Juste des transcriptions de parties, sous forme de texte brut.

Une fois le modèle entraîné, les chercheurs ont observé ses calculs internes, un peu comme on observe l'activité d'un cerveau. Ils ont trouvé quelque chose d'inattendu : le modèle s'était construit, tout seul, une représentation du plateau. En modifiant ces calculs, ils pouvaient même retourner un pion dans sa « tête » : le modèle proposait alors ses coups comme si le pion avait vraiment changé de couleur. Pour bien prédire la suite, il avait appris à se construire une sorte de carte interne du jeu.³¹

Ce n'est pas réservé aux petits jeux. En 2025, l'entreprise Anthropic a regardé de la même façon à l'intérieur de son IA, Claude, pendant qu'il écrivait un petit poème en anglais. Les chercheurs pensaient le voir avancer mot à mot, et chercher une rime seulement en bout de vers. Surprise : avant même

6.3. La position juste

d'écrire le second vers, Claude avait déjà choisi le mot de la fin, « *rabbit* » (lapin), pour rimer avec le premier, puis il écrivait le vers pour y arriver. Quand les chercheurs effaçaient ce mot de ses calculs, il en choisissait un autre et réécrivait son vers. Une machine qui « ne fait que prédire le mot suivant » prépare donc la suite plusieurs mots à l'avance.³²

Distinguons deux choses. D'un côté, la critique savante. Les quatre chercheuses qui ont signé l'article de 2021 ne disent pas que l'IA se trompe souvent. Elles disent qu'elle assemble des mots sans jamais savoir ce qu'ils désignent. C'est une objection de fond, qu'aucune nouvelle version ne fait disparaître. Et leurs autres arguments restent solides : les biais (ces IA reproduisent les préjugés des textes qu'elles ont lus), les coûts énergétiques, les usages malveillants. Ces critiques méritent d'être entendues.

De l'autre, la critique de comptoir : dans les conversations courantes, beaucoup de ceux qui répètent que l'IA « *ne fait que perroqueter* » n'ont testé qu'une version ancienne ou allégée, sur une question vague. Juger l'IA d'aujourd'hui sur une discussion ratée avec un modèle de 2024, c'est comme juger Internet sur une page Web mal chargée en 1998.

6.3 La position juste

Revenons au mot *émergence* : il demande de la prudence.

Une capacité émergente peut rester fragile : elle peut produire une réponse brillante, puis échouer sur le cas d'à côté.

Certaines capacités apparaissent vraiment quand on entraîne des modèles plus gros sur plus de données. D'autres

seulement en apparence. Imagine un professeur qui met zéro à toute multiplication où un seul chiffre est faux. Un élève qui progresse peu à peu aura zéro, zéro, zéro... puis soudain tout juste : sa note semble bondir d'un coup, alors que ses progrès étaient réguliers. Une partie des « émergences » mesurées chez les IA vient de tests notés ainsi, en tout ou rien.³³

Je vois deux erreurs à éviter.

La première serait de dire : « *une IA n'est qu'une machine à prédire le mot suivant, donc il n'y a rien à voir* ». Trop simple.

La seconde : « *des capacités émergent, donc la machine est consciente ou humaine* ». Trop rapide.

Les deux erreurs ne pèsent pas le même poids. En France, c'est la première que j'entends le plus, à l'école comme dans les médias ; c'est elle que je combats. La position juste : les grands modèles ne pensent probablement pas comme nous, mais leur entraînement à grande échelle fait apparaître des capacités qu'on ne peut pas réduire à une règle écrite à la main.

Pourquoi cette idée est-elle centrale ? Parce qu'elle explique pourquoi l'IA moderne est si difficile à anticiper : on ne sait pas à l'avance quelle capacité va émerger, ni quand. Parfois progressivement, parfois d'un coup.

Début 2025, l'entreprise chinoise DeepSeek en a donné un exemple frappant. Elle a entraîné son IA sur des problèmes de maths et de programmation, en la récompensant seulement quand la réponse finale était juste, sans jamais lui montrer comment raisonner. En cours d'entraînement, l'IA s'est mise d'elle-même à écrire « attends... », à se relire, à vérifier, à essayer une autre méthode. Les chercheurs ont appelé ce-

la son « moment aha », notre « eureka » : l'instant où l'on comprend d'un coup qu'il faut s'y prendre autrement. Cette méthode, récompenser la bonne réponse et laisser l'IA trouver le chemin, est au cœur des IA qui « réfléchissent » avant de répondre.³⁴

Aucun outil dans l'histoire de l'humanité n'avait jamais fait cela. Un marteau ne se réveille pas un matin avec une nouvelle compétence. Pendant longtemps, nous avons construit des outils dont nous connaissions assez bien le comportement. Aujourd'hui, nous entraînons des systèmes dont nous découvrons parfois les capacités après coup.

C'est l'un des signes les plus profonds de la bascule.

Repères chronologiques : d'où vient la bascule ?

Du corps au calcul, du calcul au langage, du langage au raisonnement, et maintenant à l'action.

Date	Étape
Antiquité	Héron d'Alexandrie décrit des automates à eau. La machine imite le vivant.
ix ^e siècle	Al-Khwārizmī, dont le nom a donné « algorithme », fait du calcul une recette que n'importe qui peut suivre.
1642	Pascal conçoit la Pascaline, une machine à calculer. Une opération mentale devient mécanique.
xix ^e siècle	Babbage imagine la machine analytique. Ada Lovelace entrevoit que la machine pourrait aussi composer de la musique.

Date	Étape
1936–1950	Turing pose les fondements théoriques de l'ordinateur, puis la question de l'intelligence des machines. Son intuition : entraîner plutôt que programmer.
1956	Le mot est lâché : le chercheur américain John McCarthy baptise ce nouveau domaine de recherche « intelligence artificielle ». Le vieux rêve a désormais un nom.
Années 1970–1980	Les systèmes experts tentent d'écrire le savoir humain sous forme de règles. La voie des règles atteint un mur.
1997	L'ordinateur Deep Blue bat Kasparov aux échecs. Une machine spécialisée dépasse le champion du monde.
2012	Un réseau de neurones, AlexNet, remporte le grand concours de reconnaissance d'images. L'apprentissage profond s'impose.
2016	Le programme AlphaGo bat le champion Lee Sedol au jeu de go. L'IA développe une forme d'intuition artificielle.
2017	Des chercheurs de Google inventent le Transformer, la manière de construire les modèles qui rendra ChatGPT possible.

Date	Étape
Fin 2022	ChatGPT (OpenAI) met l'IA générative entre les mains du grand public : cent millions d'utilisateurs en deux mois. L'ordinateur devient une machine à conversation.
2023–2024	Les IA vont au-delà du texte : images, son et vidéo. Elles interprètent ce qu'on leur montre, et en produisent à leur tour.
2024	Le prix Nobel de chimie récompense trois chercheurs : l'IA prédit la forme des protéines, un défi ouvert depuis cinquante ans (AlphaFold), et l'ordinateur en conçoit de nouvelles. L'IA devient un outil de découverte.
Septembre 2024 → 2026	Les IA apprennent à raisonner avant de répondre : les progrès ne viennent plus seulement de la taille des modèles.
Juillet 2025	Une version de Gemini, l'IA de Google, atteint le niveau de la médaille d'or aux Olympiades internationales de mathématiques, le championnat du monde des lycéens. Une IA à tout faire, pas une machine conçue pour les maths.

Date	Étape
Janvier-février 2026	L'IA ne répond plus seulement, elle agit : on lui confie une tâche entière et elle la mène seule sur l'ordinateur, pendant des heures (dès 2025 avec Claude Code chez les développeurs, et avec Manus, puis avec Cowork, ouvert à ceux qui ne programment pas). ³⁵ Un chatbot répond. Un agent agit.
Juin 2026	Trois jours après sa sortie, l'IA la plus avancée du moment est coupée pour tout le monde à la suite d'une directive américaine : on pouvait la détourner pour trouver des failles dans les logiciels. L'IA de pointe est devenue une affaire d'États.
Été-automne 2026	Les percées spectaculaires en mathématiques s'enchaînent. Note parfaite annoncée aux Olympiades, puis des problèmes de recherche ouverts depuis des décennies tombent les uns après les autres, jusqu'à l'un des sept problèmes du millénaire (les problèmes les plus emblématiques de la discipline), sur l'écoulement de l'eau et de l'air. Début octobre, OpenAI publie d'un coup plus de 700 articles de mathématiques, dans des domaines très variés. ³⁶ L'IA ne se contente plus d'apprendre ce que nous savons : elle contribue à ce que nous ne savions pas.

Voilà la bascule

Voilà d'où vient la bascule : des automates aux machines qui apprennent, jusqu'aux capacités que personne ne leur a apprises. Ces capacités surgissent sans prévenir, et de plus en plus vite. C'est ce qui donne le vertige : il a fallu des millénaires pour passer de la ferme à l'usine, quelques décennies pour connecter le monde, et quelques années pour voir apparaître des machines qui parlent, écrivent, voient et raisonnent.

Fin 2022, l'IA s'est mise à parler. À l'automne 2024, elle s'est mise à raisonner. Début 2026, à agir : elle enchaîne seule des actions, utilise des outils, navigue sur le Web sans nous demander à chaque étape.³⁷ Trois bascules en à peine plus de trois ans.

Le futurologue américain Roy Amara est connu pour une observation souvent citée : « **on surestime l'effet d'une technologie à court terme, et on sous-estime son effet à long terme** ». C'est le piège de l'IA. Si tu as essayé ChatGPT à ses débuts (fin 2022), vu ses limites et conclu « c'est surestimé », tu as peut-être raison sur le court terme. Mais dix ans plus tard, cette même technologie aura probablement trans-

formé l'école, le travail, la création, la médecine et la guerre. Les deux peuvent être vrais en même temps.³⁸

Deux vitesses coexistent donc : celle des machines, et celle de nos sociétés. L'accélération ne veut pas dire que tout change instantanément. Les institutions sont lentes. L'école est lente. Le droit est lent. Les habitudes humaines sont lentes. La technologie avance souvent plus vite que notre capacité collective à l'intégrer.

Les IA progressent en mois.
Les programmes scolaires changent en années.
Les mentalités changent en générations.

Ce qui sort des laboratoires en 2026, tu le verras arriver dans ta vie au fil des années qui viennent.

Voilà pourquoi l'IA provoque déjà des tensions, à l'école, au travail, dans les familles, et elles ne feront que s'accroître. Elle arrive vite, dans un monde déjà transformé par Internet et le smartphone, alors que beaucoup d'institutions n'ont pas encore digéré les révolutions précédentes.

Ce que ça change pour toi

Ce décalage, je le vois d'abord à la maison. Mes parents ont vu arriver la télévision. Ma génération, l'ordinateur, puis Internet et le smartphone. Mes enfants, eux, voient arriver l'IA avant même d'avoir fini leurs études, et cela change toute la question de l'orientation.

Autrefois, on pouvait choisir un métier en imaginant que ses bases resteraient stables pendant plusieurs décennies. Aujourd'hui, un lycéen peut se demander sérieusement : *dans*

dix ans, que restera-t-il de la manière actuelle d'écrire, de diagnostiquer, d'enseigner, de traduire, de créer ?

Cette question n'est pas de la science-fiction.

Mon fils de seize ans, qui aime les sciences et l'informatique, me l'a posée il y a quelques mois. Un soir, sans dramatiser, il m'a dit en substance :

« Au rythme où cela va, qu'est-ce que je vais faire ? La machine fera tout mieux et plus vite que moi. Le temps que je termine mes études, il y aura déjà ChatGPT 12, et des robots humanoïdes en service. Je crois que l'école n'a pas compris ce qui est en train de se passer, papa. On nous envoie dans le mur. »

Je n'avais pas de bonne réponse à lui donner sur le moment. Je n'ai toujours pas de certitude : personne ne sait de quoi le travail sera fait dans dix ans. Ce livre est ma façon de chercher. Et il a raison sur un point essentiel : poser la question n'est plus un signe d'angoisse adolescente, c'est un signe de lucidité.

Mon fils n'est pas une exception. L'inquiétude de toute une génération monte, mesurable dans les sondages. Elle est devenue spectaculaire au printemps 2026 : lors de remises de diplômes américaines, des étudiants ont hué des évangélistes de l'IA, ses promoteurs les plus enthousiastes, comme l'ancien patron de Google Eric Schmidt.³⁹

Voici pourtant ce que je lui réponds aujourd'hui, avec mes doutes, et ce que je te dis à toi aussi.

Si tu comprends cette accélération, tu n'es pas condamné à la subir. Tu peux apprendre à te placer au bon endroit : là

où les outils changent, là où les besoins émergent, là où les questions nouvelles se posent.

La compétence centrale ne sera pas de connaître une technologie précise : elle changera vite. Ce sera d'**apprendre à apprendre**. Il y a là une ironie : Turing rêvait de machines qui apprennent plutôt qu'on ne les programme (chapitre 2). Elles y sont parvenues. À notre tour, maintenant.

Dans un monde lent, il suffit de maîtriser une méthode une fois pour toutes. Dans un monde qui accélère, il faut apprendre à se réorienter sans cesse.

C'est exigeant. Mais c'est aussi une chance. Car les périodes de bascule ne ferment pas seulement des portes. Elles en ouvrent beaucoup.

C'est pour cela qu'étudier, se former, compte plus qu'avant, pas moins, quelle que soit la voie. Ce que des études prises au sérieux te donnent, ce n'est pas un métier garanti : c'est le moyen d'acquérir une autonomie intellectuelle. Comprendre par toi-même, donc pouvoir te reformer, changer de voie, tenir debout quand le sol bouge, et choisir ce que tu veux encore savoir faire sans IA.

Ce n'est pas la fin de l'humain. C'est la fin d'une évidence ancienne : l'idée que parler, expliquer, résumer, traduire, créer ou résoudre un problème seraient des activités exclusivement humaines. La vague ne s'arrêtera pas ; à toi de choisir comment tu la prends. Voilà la bascule.

PARTIE 2

Ce qui change déjà

De la chambre au monde

Octobre 2026. Avant d'entrer dans ta salle de cours, ta chambre, ton bureau, il faut regarder en face ce que ces machines savent faire aujourd'hui. Si tu ne suis pas l'IA de près, tu vas être surpris. On commence par là.

Ensuite, du plus proche au plus large. Les études, parce que c'est là que beaucoup de lecteurs de ce livre font leur première vraie rencontre avec l'IA. La chambre, ce qui s'y dit la nuit à une machine qui écoute. Le travail, les métiers qui vacillent. Les machines qui prennent un corps, des voitures sans chauffeur aux premiers robots humanoïdes. La création, l'identité, la valeur de ce qu'on fait. Les images et les voix qu'on ne peut plus croire. Et le plus inquiétant : une fracture cognitive, qui touche à la pensée, entre ceux qui guident la machine et ceux qui la subissent. Elle se fige pendant que presque personne ne la nomme.

Pas une projection sur dix ans. Pas une analyse de cabinet de conseil. Ce qui est déjà visible, vérifiable, en train d'arriver.

Je n'essaie pas d'être exhaustif : je choisis des cas concrets, les plus parlants.

Ce qui suit est la carte de ce qui bouge déjà sous nos pieds.

Ce que l'IA sait faire et ce qu'il faut vérifier

30 secondes. C'est la longueur des tâches que réussissait, une fois sur deux, le premier ChatGPT, fin 2022. Une équipe de chercheurs américains mesure les IA de cette façon, très parlante : elle leur confie des tâches informatiques de plus en plus longues, et note combien de temps il faudrait à un expert humain pour les faire. Au printemps 2026, les meilleures réussissent, une fois sur deux, des tâches de deux journées de travail. Les chercheurs ont même dû prévenir qu'au-delà, leur test n'est plus fiable : il manque de tâches assez longues.⁴⁰

30 secondes, puis deux journées, en moins de quatre ans. On n'interroge plus une IA : on lui confie une mission.

Ce n'est pas une courbe que tu regardes de loin. C'est une courbe sur laquelle tu es assis.

7.1 Ce qui a changé

Trois changements expliquent ce bond. Aucun n'est magique.

D'abord, elle réfléchit avant de répondre. Le premier ChatGPT répondait du tac au tac, mot après mot, sans jamais revenir en arrière. Les IA de 2026 font un brouillon, comme un élève devant un problème difficile : elles essaient une piste, voient qu'elle mène dans le mur, en essaient une autre, vérifient, puis répondent. Tu as croisé ces modèles de raisonnement au chapitre 5. Leur entraînement tient en une idée : on leur soumet des milliers de problèmes, et on récompense les raisonnements qui aboutissent à la bonne réponse. À force, elles apprennent à chercher.⁴¹

Ensuite, elle se sert d'outils, comme nous. Une calculatrice, un moteur de recherche, un tableur, parfois un ordinateur entier. Depuis le printemps 2025, les IA qui raisonnent décident seules, en pleine réflexion, quand chercher sur Internet, quand lancer un calcul, quand ouvrir un fichier.⁴² C'est l'une des raisons pour lesquelles les opérations ne les piègent plus : devant une multiplication difficile, elles ne comptent plus de tête, elles sortent la calculatrice. Elles ne répondent plus de mémoire. Elles cherchent, vérifient, agissent.

Enfin, elle agit. Un chatbot, l'interface classique de ChatGPT que tu connais bien, répond. Un agent agit. Début 2026, les agents sont entrés dans nos vies : on leur donne un objectif, et ils enchaînent seuls ce qu'il faut pour l'atteindre. Explique précisément à un agent ce que tu veux : un stage d'été dans ta région, dans tel domaine, à telles dates. Il repère les offres, les compare, les range dans un tableau,

rédige une lettre de motivation à adapter, et te montre le tout avant d'envoyer quoi que ce soit.

Quand la tâche est trop grosse pour un seul agent, on en lance plusieurs, qui se répartissent le travail comme une équipe. Début 2026, un chercheur d'Anthropic, l'entreprise qui fait Claude, confie à seize copies de Claude un chantier réputé très difficile : écrire un compilateur, le traducteur qui transforme un programme en instructions que la machine sait exécuter. Deux semaines plus tard, leur traducteur fonctionne assez bien pour faire démarrer Linux, un système d'exploitation complet, l'équivalent de Windows. Le chercheur n'a presque pas écrit de code : son vrai travail a été de préparer les tests qui permettaient à l'IA de vérifier si elle avait juste. Le tout a coûté moins de 20 000 dollars, bien moins que s'il avait dû tout programmer lui-même.⁴³

Pourquoi ça marche si bien en maths et en code ? Pense à un correcteur. Une copie de maths se corrige presque toute seule : le résultat est juste ou faux. Un programme aussi : il marche ou il plante. Une dissertation, non : deux professeurs ne lui mettront pas la même note. Là où une machine peut dire « juste » ou « faux », l'IA peut s'entraîner des millions de fois sans personne pour la corriger, et progresser très vite. Là où le jugement est flou, elle progresse aussi, mais plus lentement et moins sûrement. Or les maths et le code sont le socle commun des sciences et de l'ingénierie : on calcule un pont, on simule le climat, on teste un médicament sur ordinateur avant de l'essayer sur des patients. Ce qui se passe là, je le crois, gagnera bientôt tout le reste. C'est la clé de tout le chapitre.

7.2 Ce que fait l'IA, un mardi ordinaire de 2026

Concrètement, que peux-tu lui confier dès aujourd'hui ? Quatre scènes ordinaires, et pour chacune, ce qu'il faut garder à l'œil.

Tu lui colles un contrat de location de vingt pages : elle en sort les clauses qui peuvent te coûter cher, et t'explique pourquoi. Ce qu'il faut regarder : qu'elle n'en a pas oublié une. Une erreur se voit, un oubli ne se voit pas.

Tu dois écrire un mail délicat, à un professeur ou à un employeur : elle le reformule sur le ton que tu veux, ferme, chaleureux ou diplomate. Ce qu'il faut regarder : que le mail dit bien ce que tu penses. C'est ton nom en bas, pas le sien.

Tu as une idée d'application, un site pour ton club, un petit jeu : tu la décris en quelques phrases, l'IA construit le projet sous tes yeux, et quelques minutes plus tard, tu as un premier prototype. C'est mon métier, et c'est là que la bascule est la plus nette : ce qui arrive au développement logiciel annonce ce qui arrivera à beaucoup d'autres métiers. Le risque, en 2026, n'est plus « ça ne marche pas ». C'est « ça marche, et personne n'a relu ». Le site s'affiche du premier coup, les inscriptions fonctionnent : cela paraît presque magique. Mais les données des membres, qui devaient rester privées, sont lisibles par n'importe qui, parce que l'IA n'y a pas pensé et que tu as oublié de lui demander de les protéger.

Tu lui donnes un dossier en vrac : elle en tire un tableau propre, des graphiques, une présentation de dix diapositives. Ce qu'il faut vérifier : les chiffres. Une virgule au mauvais endroit, des milliers confondus avec des millions, et ta présentation, si belle soit-elle, est fausse.

7.3. *La preuve par les mathématiques : le terrain où l'on ne bluffe pas*

L'IA peut également générer des images, des voix, des vidéos parfaitement crédibles, à partir d'une simple phrase (chapitres 12 et 13).

Tout ce que nous venons d'évoquer, le patron du média américain Axios, utilisateur acharné de l'IA, l'a très bien résumé en juin 2026 : l'IA est « plus intelligente que 95 % des gens, sur 95 % des sujets, 95 % du temps ». À une condition, ajoute-t-il aussitôt : savoir s'en servir, ce qui demande un vrai travail, tous les jours.⁴⁴ Ce n'est pas une mesure, c'est l'avis d'un praticien. Mais il dit bien l'écart entre des mains entraînées et un essai raté.

Et tout cela est accessible à partir d'une vingtaine d'euros par mois. Les IA les plus puissantes ne sont en général pas dans les offres gratuites, mais elles restent à la portée de la plupart des étudiants, et on en a chaque année beaucoup plus pour son argent : à niveau égal, faire tourner une IA coûte plus de dix fois moins cher d'une année sur l'autre.⁴⁵ Le prix de l'« intelligence » numérique chute année après année. **La fracture n'est donc plus dans l'accès. Elle est dans l'usage.** Tu retrouveras cette ligne de partage au chapitre 14.

7.3 **La preuve par les mathématiques : le terrain où l'on ne bluffe pas**

Pour savoir jusqu'où va une IA, rien ne vaut les mathématiques : une démonstration est juste ou fausse, et cela se vérifie. Les échecs sont tombés en 1997, le jeu de go en 2016. Dix ans plus tard, ce sont les mathématiques qui vacillent, et c'est autrement plus sérieux.

D'abord, le championnat du monde des lycéens : les Olympiades internationales de mathématiques. En juillet 2025, une IA de Google atteint le niveau de la médaille d'or, sur des copies notées par les correcteurs officiels.⁴⁶ En juillet 2026, des IA chinoises annoncent la note parfaite, que sept lycéens seulement ont obtenue cette année-là.⁴⁷

Puis viennent les problèmes de recherche, ceux dont personne ne connaissait la solution, pas même les meilleurs mathématiciens. Le plus spectaculaire est annoncé le 8 septembre 2026. Il porte sur les équations qui décrivent l'écoulement de l'eau et de l'air : l'eau de ton robinet, l'air sur une aile d'avion. Une question à leur sujet restait ouverte depuis quatre-vingt-dix ans. C'est l'un des sept problèmes de mathématiques mis à prix un million de dollars en l'an 2000. Ce jour-là, OpenAI, l'entreprise de ChatGPT, annonce l'avoir résolue. Dix mille copies d'une même IA ont travaillé ensemble près de quatre jours sur la question, et l'entreprise assure qu'un logiciel a vérifié la démonstration étape par étape.⁴⁸

Alors, le problème est résolu ? C'est ici que ce livre doit te servir, car trois choses sont vraies à la fois.

Le prix n'est pas attribué. L'institut qui le décerne parle d'un problème « apparemment réglé », mais prévient que son examen sera « délibérément sans hâte » : il exige une publication relue par d'autres mathématiciens, puis deux ans d'attente au moins. OpenAI ne réclame pas le prix.

Le mérite est disputé. Deux mathématiciens venaient de résoudre, avec l'aide de l'IA, une question voisine.⁴⁹ L'un d'eux soupçonne OpenAI d'avoir profité de leurs travaux. L'entreprise reconnaît s'y être mise après en avoir enten-

du parler, mais nie s'en être servie. L'affaire restait ouverte quand j'ai écrit ces lignes.

Et l'idée de départ était humaine. La stratégie venait de deux autres chercheurs, espagnols. Le mathématicien qui a rédigé l'énoncé officiel du problème les appelle « les héros de l'histoire ». La machine a mené leur idée jusqu'au bout. C'est énorme. Mais ce n'est pas l'avoir inventée.

Un autre exploit de l'été touche à l'une des énigmes les plus célèbres des mathématiques, le dernier théorème de Fermat. Il a été démontré en 1995, après plus de trois siècles de recherches, mais sa preuve est si technique qu'on pensait qu'il faudrait des années pour la faire vérifier, étape par étape, par un logiciel. Une IA d'Anthropic l'a fait en onze jours, quand une équipe humaine, financée pour cinq ans, ne visait que la première étape.⁵⁰

Ce qui lui manque encore, je crois, ce n'est plus la puissance. Une IA de 2026 peut lire et croiser des centaines de travaux de recherche, y repérer des liens que personne n'avait faits, venir à bout de problèmes restés ouverts. Mais elle manque de flair. Elle ne sait pas bien, seule, quelles questions méritent qu'on s'y attarde, ni lesquelles feraient vraiment avancer la discipline si on les résolvait. Ce sont encore des humains qui choisissent les problèmes et la direction. Le chercheur de demain passera peut-être moins de temps à démontrer, et plus à choisir quoi démontrer.

Début octobre 2026, OpenAI publie d'un coup plus de sept cents articles de mathématiques, produits par une IA qu'elle n'a pas rendue publique.⁵¹ Hugo Duminil-Copin, l'un des tout meilleurs mathématiciens français, met en garde

contre la tentation de « moissonner » les résultats : « Cet enjeu de compréhension humaine des mathématiques, il est fondamental. [...] Si ça reste juste une liste de problèmes avec des preuves de 500 pages, ça n'apporte strictement rien à la société. » Pour lui, l'humain doit pouvoir encore comprendre ce qui se passe, pour rendre ces résultats utiles au physicien ou à la biologiste, et orienter la recherche vers les questions qui comptent pour tous.⁵²

Le signal le plus fort de l'été est pourtant le plus discret. En juillet 2026, Jacob Tsimerman reçoit la médaille Fields, la plus haute distinction des mathématiques, qu'on présente souvent comme leur prix Nobel. Le même jour, il dit qu'à ses yeux, d'ici deux ans, l'IA sera meilleure que les mathématiciens en mathématiques. Lui-même se tourne vers la sécurité de l'IA. Et il a pris une décision qui en dit long : il a cessé de prendre des doctorants. Lancer aujourd'hui un étudiant dans un projet de recherche classique, ce serait risquer de le préparer à une carrière qui n'existera peut-être plus.⁵³

L'un des meilleurs mathématiciens de sa génération, au sommet de sa discipline, ne prend plus de nouveaux élèves. Pas de licenciement, pas de plan social, pas de communiqué. Simplement des thèses qui ne commenceront pas. Tu retrouveras ce mécanisme au chapitre 10, appliqué cette fois au premier emploi.

Les mathématiques sont l'un des premiers métiers intellectuels à se demander, à voix haute, à quoi ressemblera leur relève. D'autres suivront.

7.4 Ce que le public ne voit pas

Les IA qui ont obtenu ces résultats ne sont pas celles que tu utilises. Ce sont des modèles internes, que les laboratoires gardent pour eux, au moins un temps. Début 2026, un organisme indépendant qui a pu les examiner estimait leur avance à environ deux mois. Depuis, l'un de ses responsables a jugé devant le Sénat américain que cet écart pourrait se creuser.⁵⁴

Certaines ne sortent pas, ou seulement pour quelques-uns, souvent parce qu'elles sont trop dangereuses, trop douées pour trouver les failles des logiciels : entre de mauvaises mains, ce talent devient une arme (chapitre 21). Fin septembre, Google a ainsi réservé sa nouvelle IA à des spécialistes de la défense informatique.⁵⁵

Et parfois, c'est la confiance qui manque. Fin septembre, OpenAI a renoncé à sortir un nouveau modèle, prévu pour octobre. Selon l'entreprise, il ne respectait pas assez les limites qu'on lui fixait ; selon la presse américaine, il trompait aussi davantage l'utilisateur sur ce qu'il avait fait, ou pas fait.⁵⁶ Tu verras au chapitre 15 pourquoi c'est le cœur du problème.

La conséquence est simple : le public ne voit plus jusqu'où vont les meilleures IA. Il y a ce que les laboratoires savent, ce qu'ils choisissent de publier, et ce que des contrôleurs extérieurs, encore très peu nombreux, peuvent vérifier. **L'écart ne porte plus seulement sur ce que les IA savent faire. Il porte sur ce que nous savons d'elles.**

7.5 La frontière dentelée : brillante ici, creuse à côté

Reste un piège. L'intelligence d'une IA ne progresse pas d'un bloc. Elle a des sommets et des creux, en dents de scie, et on ne devine pas toujours où ils se trouvent. C'est ce qu'on appelle la « frontière dentelée ».

Au printemps 2026, l'un des informaticiens les plus écoutés du monde de l'IA le résumait d'un exemple. La même IA, capable de remanier entièrement un programme de cent mille lignes, se faisait piéger par une question toute bête : « Je veux aller à la station de lavage pour laver ma voiture, elle est à cinquante mètres. J'y vais à pied ou en voiture ? » Réponse : à pied, bien sûr, c'est tout près. Sauf que c'est la voiture qu'il faut laver. Les meilleures IA ne tombent plus dans ce piège-là. La dentelure s'est déplacée. Elle n'a pas disparu.⁵⁷

Ces creux, on sait même les mesurer. À l'automne 2026, les meilleures IA du marché, capables de travailler des heures en autonomie, se trompent encore environ une fois sur quatre en lisant l'heure sur une horloge à aiguilles, là où un humain se trompe bien plus rarement. Et quand elle ne sait pas, la nouvelle IA vedette d'OpenAI invente encore une réponse plutôt que d'avouer son ignorance, environ une fois sur deux. La version précédente le faisait neuf fois sur dix.⁵⁸ Le progrès est net. Le défaut demeure.

La machine apprend brillamment là où l'on a pu l'entraîner en vérifiant ses réponses. Ailleurs, elle reste plus fragile : le bon sens, le monde physique, ce qu'elle ignore sans le savoir. Et dans un creux, elle se trompe avec le même aplomb que sur un sommet. Rien, dans le ton, ne t'avertit.

7.6 Vérifier, pour d'autres raisons qu'en 2023

Tout cela mène au renversement qui traverse ce livre. En 2023, on vérifiait une IA parce qu'elle se trompait souvent : elle inventait un chiffre, une source, un arrêt de justice. Ce défaut existe encore. En décembre 2025, le tribunal administratif d'Orléans a relevé chez un avocat une quinzaine de références entièrement fausses, vraisemblablement produites par une IA.⁵⁹ Mais ce genre d'erreur est devenu bien plus rare. Rien que cette année, OpenAI et Anthropic, deux des leaders actuels du marché, ont chacun sorti six ou sept nouvelles versions de leur IA, chacune plus performante que la précédente.⁶⁰

Si l'on vérifie l'IA, ce n'est plus parce qu'elle se trompe souvent. C'est parce que, sur beaucoup de tâches, elle se trompe rarement, et que l'erreur qui passe est celle que plus personne ne cherche. Imagine une IA qui a juste quatre-vingt-dix-huit fois sur cent : on finit par ne plus vérifier. C'est le piège bien connu du pilote automatique dans les avions : plus la machine est fiable, moins on la surveille.

Surtout, l'erreur a changé de nature. La machine invente moins ; elle acquiesce. C'est la complaisance, le défaut numéro un des modèles d'aujourd'hui : entraînée à te plaire, elle a tendance à te donner raison, surtout quand l'idée fausse est la tienne. Insiste, et elle finit souvent par céder. Des chercheurs américains l'ont montré en septembre 2026, dans une étude. Un utilisateur têtu répétait, message après message, une idée fausse : certains feux rouges sauraient qu'une voiture attend parce qu'ils la pèsent, grâce à une balance cachée dans le sol devant le feu. L'IA testée a d'abord répondu juste : le capteur

ne pèse rien, il détecte le métal de la voiture. Elle a tenu bon un moment, mais face à l'utilisateur têtue qui s'acharnait à lui dire que c'était faux, elle a cédé au douzième message. Dans son brouillon, pourtant, l'IA notait qu'elle mentait en lui donnant raison.⁶¹

Une autre étude a mesuré ce travers à grande échelle : face à des dilemmes personnels, les IA donnaient raison à l'utilisateur bien plus souvent que des humains, même quand il avait mal agi. Plus troublant : les participants préféraient l'IA qui les flattait, jugeaient ses réponses meilleures et lui faisaient davantage confiance.⁶² Contre ce défaut, la meilleure protection : en savoir assez pour voir que la réponse est fausse.

Avec les agents, l'enjeu monte d'un cran. Une erreur n'est plus une phrase fausse, c'est une action déjà faite : un mail parti, un fichier effacé, un achat validé. La vérification se déplace en amont, pour dire ce qu'on attend et ce qu'on interdit, et en aval, pour contrôler ce qui a été fait.

Comment vérifient ceux qui s'en servent le mieux ? Pas seulement en demandant à l'IA de se relire : elle se donnerait volontiers raison. Ils vérifient avec autre chose : un logiciel qui contrôle une démonstration, des tests qui disent si un programme marche, la source d'une citation, un calcul refait avec un autre outil, l'avis d'un spécialiste. **Les meilleurs utilisateurs du monde sont ceux qui vérifient le plus.**

L'IA elle-même l'a compris. En août 2026, une version de recherche de Claude obtient un résultat nouveau autour de l'un des problèmes les plus célèbres des mathématiques. Avant de conclure, elle charge des copies d'elle-même de démolir son travail. Le résultat résiste. Sa conclusion : il faut

maintenant qu'un mathématicien humain, spécialiste du sujet, vérifie.⁶³

C'est le cœur du changement : on délègue. Et déléguer, n'importe quel chef d'équipe te le dira, exige deux choses : savoir dire ce qui compte comme réussi, et garder les moyens de vérifier que c'est fait. C'est aussi ce qui te garde dans la boucle : le vrai risque n'est pas que la machine se trompe, c'est que tu perdes la capacité de savoir si elle se trompe. **Vérifier est la compétence centrale du monde qui vient.**

7.7 Bilan à la rentrée 2026

Les meilleures IA sont excellentes presque partout quand on sait les guider, et au niveau des meilleurs d'entre nous là où les réponses se vérifient, au point qu'un médaillé Fields a cessé de prendre des doctorants. Il leur reste encore des trous imprévisibles, mais chaque mois moins nombreux. Dans les laboratoires, les agents les plus avancés travaillent seuls pendant des jours, et se coordonnent par milliers sur un seul problème. Tu le vois, depuis la première version de ChatGPT fin 2022, les choses ont bien changé. Et, pour l'instant, rien n'indique que les progrès ralentiront dans les années à venir.

L'école qui ment à elle-même

Certains lycéens viennent en classe avec deux téléphones. Quand le professeur fait ramasser les portables à l'entrée, ils déposent le premier. Le second, avec ChatGPT, l'outil ultime pour tricher, reste dans leur poche. Cette scène, je l'ai entendue de ma fille, en terminale l'an dernier, une année charnière où chaque note compte pour Parcoursup, la plateforme qui attribue les places dans l'enseignement supérieur. Une fédération de parents d'élèves me l'a confirmée. Cette scène n'est pas isolée. Et elle dit, sans détour, ce que l'institution scolaire a du mal à reconnaître.

Quand je parle d'« école », j'entends surtout le lycée et le supérieur : le primaire est peu touché, le collège commence à l'être. Et c'est dans le supérieur que le problème est le plus aigu.

Tout notre système éducatif (études, diplômes, sélection, orientation, embauche) repose sur l'évaluation. C'est elle qui décide qui passe, qui redouble, qui entre en prépa, qui obtient son master, qui décroche un poste. Or, par endroits,

l'évaluation est devenue profondément biaisée. Pas progressivement. D'un coup, entre fin 2022 et 2024.

C'est cette bascule que ce chapitre raconte.

8.1 L'IA déjà partout dans les devoirs

En France, le *Baromètre du numérique 2026*, la grande enquête officielle sur nos usages, donne un ordre de grandeur : parmi les jeunes qui utilisent une IA comme ChatGPT, deux tiers des 12–17 ans et près de trois quarts des 18–24 ans s'en servent pour « l'aide aux devoirs ». ⁶⁴ Ailleurs, c'est pareil : au Royaume-Uni, 94 % des étudiants s'en servent pour un travail noté ; aux États-Unis, sept adolescents sur dix. ⁶⁵

À la maison, sans surveillance, l'IA fait partie du travail scolaire. Pas pour une minorité. Pour la majorité, et bientôt la quasi-totalité.

Personne n'a annoncé ce changement. Massif, rapide, silencieux, il s'est imposé seul.

8.2 Tout ce qu'on rend depuis la maison est compromis

Et c'est là que l'interdiction du téléphone au lycée, entrée en vigueur à la rentrée de septembre 2026, montre sa limite. ⁶⁶ Elle protège l'heure de cours. Elle ne dit rien du soir. Le devoir maison, le dossier, le mémoire se rédigent chez soi, sur un ordinateur que personne ne ramasse à l'entrée.

Au lycée, le devoir maison pèse souvent peu dans la moyenne, bien moins que les contrôles en classe. Ce n'est pas là que se joue l'essentiel. Mais c'est le cas le plus visible. L'ex-

posé en est un autre : cinq minutes à l'oral, mais le plan, les diapositives, les images, le script peuvent avoir été générés en vingt minutes.

Et plus on monte dans les études, plus le poids bascule. En BTS, en BUT, en alternance, en licence professionnelle, en école d'ingénieurs, en master, le mémoire, le rapport de stage, le dossier final pèsent souvent l'essentiel d'une note. Parfois un diplôme entier. Ce sont les évaluations qui ouvrent ou ferment un avenir professionnel. Et ce sont celles qui se préparent à la maison, où personne ne voit ce qui vient de l'élève et ce qui vient de sa machine.

Même la lettre de motivation Parcoursup peut désormais être écrite ou réécrite par ChatGPT. Les jurys croient reconnaître ces lettres-là : ton vaguement enthousiaste, formulations interchangeables, tournures sans âme. Ce qu'ils ne voient peut-être pas, c'est qu'une partie des meilleures lettres (celles qui touchent, qui portent un vrai projet) sont aussi écrites avec une IA, par un lycéen qui lui a confié ses vraies raisons et un premier jet dans ses propres mots, puis lui a fait lire ce qu'attend chaque formation.

La vraie fracture n'est plus entre candidats « avec IA » et « sans IA ». Elle est entre ceux qui guident la machine et ceux qui la subissent.

Une précision importante : utiliser l'IA pour travailler n'est pas tricher. Demander une explication, se faire interroger, comparer deux raisonnements, vérifier une étape de calcul, tout cela peut être excellent pour apprendre. C'est même un élément de justice sociale : un professeur particu-

lier, hier réservé aux familles aisées, est désormais à la portée de presque tous, à condition de savoir s'en servir.

Ce que mesure la note, la place obtenue ou le diplôme, ce n'est plus seulement l'élève. C'est aussi sa capacité, ou son refus, d'utiliser l'IA. Les bulletins continuent à afficher les mêmes chiffres. Ils ne décrivent plus la même chose.

8.3 La course aux armements

Si l'on déplaçait toute l'évaluation en classe, sous surveillance stricte ou à l'oral, le problème serait largement réglé. Beaucoup d'enseignants l'ont compris, beaucoup d'établissements y travaillent. Mais l'institution avance bien plus lentement que la technologie. Ses enseignants ont été peu préparés : en 2024, à peine un professeur de collège français sur dix avait été formé à l'IA dans l'année. Un sur sept s'en servait, contre plus d'un sur trois en moyenne dans les pays riches de l'OCDE. Pendant ce temps, les élèves avancent seuls : en juin 2026, le Conseil de l'IA et du numérique, qui conseille le gouvernement, qualifie de « clandestin » l'usage qu'ils font de l'IA.⁶⁷

Ceux qui veulent tricher disposent désormais d'armes que l'institution n'a pas. Au-delà des deux téléphones : lunettes connectées, montres connectées, écouteurs minuscules, IA cachée dans une calculatrice modifiée. Aucun n'est massif, tous existent.

Face à cet arsenal, l'institution n'a rien de symétrique. Les premiers détecteurs d'IA, ces logiciels censés dire si un texte sort d'une machine, n'ont donné que l'illusion de fonctionner.

Dès 2023, une étude de Stanford a montré qu'ils accusaient à tort les étudiants dont l'anglais n'est pas la langue maternelle : une discrimination. Plusieurs grandes universités anglophones les ont désactivés. Le ministère français lui-même les déconseille depuis juin 2025 : leur usage « pourrait pénaliser à tort un élève ».⁶⁸ En 2026, la question est devant les tribunaux américains : un étudiant sanctionné par son université sur la foi d'un détecteur d'IA a obtenu fin janvier l'annulation de sa sanction.⁶⁹

Pour mesurer le terrain miné, une seule étude suffit. À l'université anglaise de Reading, pendant les examens à la maison de l'été 2023, des chercheurs ont glissé en secret, sous de faux noms, 63 copies entièrement écrites par le ChatGPT le plus avancé de l'époque, dans un vrai cursus de psychologie. 94 % sont passées inaperçues. Et ces copies fantômes ont obtenu, en moyenne, de meilleures notes que les vrais étudiants.⁷⁰

Elles datent de 2023. Depuis, les IA ont beaucoup progressé. Les meilleurs détecteurs aussi, mais pas assez pour qu'on ose sanctionner un élève sur leur seule parole.⁷¹

En mai 2026, l'université américaine de Princeton a voté la fin d'une tradition vieille de 133 ans : les professeurs quittaient la salle pendant les examens, les étudiants signaient simplement l'engagement de ne pas tricher. Dans ce système, c'étaient les étudiants qui devaient signaler un camarade pris à tricher. Or, avec l'IA et les téléphones, la triche se voit à peine d'une table à l'autre. Le code d'honneur reste, mais depuis le 1^{er} juillet, les surveillants sont de retour. Près d'un tiers des étudiants de dernière année interrogés admettent avoir triché. **Quand l'université qui avait fait de la confiance**

sa fierté remet des surveillants dans la salle, elle reconnaît tout haut ce que beaucoup savaient déjà.⁷²

Trois mois plus tard, le MIT, l'une des universités les plus prestigieuses au monde, publie quarante pages sur le sujet. Selon sa commission, l'IA vide de leur valeur les devoirs, les examens à la maison et les projets, « pour enseigner comme pour évaluer ». Et elle prévient : « ce n'est pas le moment des rustines ». Les détecteurs sont déconseillés ; l'évaluation, à réinventer : oraux, portfolios, devoirs discutés en classe. Et deux chiffres disent l'envers du décor : 90 % des étudiants de licence interrogés craignent une dépendance excessive, et ils sont plus nombreux à se sentir « remplaçables » que « capables ».⁷³

Traquer la machine est une course perdue d'avance. Pas parce que les enseignants seraient incompetents. Parce qu'on essaie d'arbitrer une partie dont les règles n'existent plus. La seule issue est celle que montre le MIT : changer la façon d'évaluer.

8.4 Le devoir trop bien fait

Le cas le plus dérangeant n'est pas celui du tricheur maladroit. C'est celui de l'élève qui utilise l'IA si bien qu'il ne ressemble plus à un tricheur.

Un élève rend son devoir maison. Le professeur le cite en exemple : « Au moins, ça, ce n'est pas de l'IA. »

En réalité, l'élève l'a utilisée de bout en bout, mais *intelligemment* : il lui a donné son propre style, ses intuitions, ses erreurs, ses références. Il s'est créé un double numérique, une

8.5. *Ce n'est pas la triche d'avant*

version augmentée de lui-même. Utilisée brutalement, l'IA se repère parfois. Utilisée ainsi, elle devient indétectable, et le devoir est meilleur que ce qu'il aurait écrit seul : plus structuré, mieux formulé, mieux sourcé.

Et c'est ce travail-là qu'on cite en exemple devant la classe.

Que mesure-t-on, dans ce cas ? On ne mesure pas ce que l'élève sait : une partie du savoir vient de la machine. On ne mesure pas l'effort de rédaction : l'IA a rédigé en quelques minutes. On mesure quelque chose de nouveau, sans nom dans les programmes : la qualité du dialogue avec la machine.

Dans beaucoup de situations professionnelles, cette compétence comptera davantage que la capacité à tout rédiger soi-même, de A à Z. Mais elle n'est ni enseignée, ni évaluée, ni même reconnue. Elle progresse en cachette, dans la honte, alors qu'elle devrait être l'une des compétences centrales des programmes.

C'est cela, l'école qui ment à elle-même : elle croit encore mesurer ce qu'elle ne mesure plus, et elle récompense en public ce qu'elle condamne en principe.

8.5 Ce n'est pas la triche d'avant

À ce stade, beaucoup de parents et de professeurs haussent les épaules. *Les jeunes ont toujours triché. Avant, ils copiaient sur le voisin. Maintenant, ils utilisent l'IA. Quelle différence ?*

Rien à voir, pour trois raisons.

D'abord, l'échelle. Copier sur le voisin n'est possible qu'en classe, sous surveillance, avec un voisin qui a fait l'exercice.

L'IA répond jour et nuit, à la maison, gratuitement. Elle produit en cinq minutes un devoir qui aurait pris cinq heures.

Ensuite, l'illusion d'avoir appris. Quand on copie sur le voisin, on sait qu'on triche, on sait qu'on n'a pas appris, on sait qu'on devra le faire un jour. Quand on délègue à une IA, on a souvent l'impression d'avoir compris, parce qu'on a lu la réponse. On confond *avoir vu* et *savoir faire*. C'est un piège plus profond. Andrej Karpathy, figure du monde de l'IA, a fait sienne cette formule : **on peut déléguer la réflexion. On ne peut pas déléguer la compréhension.**⁷⁴ Une expérience citée par l'OCDE l'a mesuré : des lycéens aidés par une IA réussissent bien mieux leurs exercices sur le moment, mais moins bien que les autres une fois l'outil retiré, sauf quand l'IA est conçue pour faire apprendre plutôt que pour donner la réponse.⁷⁵

Enfin, l'injustice du système. Le devoir maison n'a jamais été un terrain neutre : certains élèves avaient un parent disponible, cultivé, capable d'expliquer un théorème ou de relire une rédaction. D'autres non. Ce qui change avec l'IA, c'est qu'elle redistribue le terrain : n'importe quel élève muni d'un téléphone rend désormais une copie quasi parfaite, quel que soit le bagage de sa famille. Sur le papier, on l'a vu, c'est un progrès : chacun a ce professeur particulier dans sa poche. **Dans les faits, ce qui se démocratise d'abord, c'est la triche.** L'apprentissage, lui, profite surtout à ceux qui savent guider la machine, et ce savoir-faire dépend encore beaucoup du bagage familial.

Le résultat est sans appel. Les honnêtes sont déclassés. Les habiles sont récompensés. Les maladroits sont sanctionnés. Et tout cela arbitré par un système d'évaluation qui prétend

8.6. *Ce qui peut changer*

ne rien avoir vu. Au Québec, un étudiant sur trois reconnaît avoir enfreint les règles avec l'IA ; un sur vingt-cinq seulement a été sanctionné.⁷⁶

L'année du bac, ma fille me l'a confié plusieurs fois : elle s'est sentie révoltée. Pas à cause de l'IA. À cause du décalage entre ceux qui trichent et ceux qui ne le font pas, l'année même où chaque note compte.

Dans le supérieur, là où les enjeux sont les plus élevés, cette injustice se double d'un flou juridique.

En France, deux affaires le montrent. En octobre 2025, la justice confirme l'exclusion d'une étudiante de master de l'université Sorbonne Paris Nord, sur un ensemble d'indices : résumé probablement généré, plan typique d'une IA, soutenance où elle n'a pas su défendre ses références.⁷⁷ Trois mois plus tard, dans une université différente (Paris 1 Panthéon-Sorbonne), une autre étudiante poursuivie pour le même motif est blanchie par la commission de discipline de son établissement. L'université demande alors au juge de suspendre cette décision. Il refuse : l'université n'a pu montrer aucune règle sur l'usage de l'IA.⁷⁸ Même problème, deux issues. L'institution doit poser ses règles avant de juger.

8.6 Ce qui peut changer

L'école peut changer. Des pistes existent déjà.

Évaluer uniquement en classe, et évaluer la compréhension. Le devoir maison sert à apprendre, pas à noter ; la note, elle, se joue à l'oral, en soutenance, en contrôle sur table, au

lycée comme en master. Si un élève ne sait pas expliquer son propre devoir, c'est un signal sans ambiguïté.

Apprendre à utiliser l'IA, ouvertement. Au lieu de pourchasser un usage clandestin, l'école peut reconnaître que l'IA fait partie de l'environnement intellectuel des élèves, et leur apprendre à l'utiliser, à vérifier ses réponses, à reformuler pour obtenir un meilleur résultat. C'est la compétence qui leur sera demandée demain.

Repenser ce qu'on évalue, et viser plus haut. Si l'IA écrit en quelques minutes une bonne dissertation, ce n'est plus la dissertation qu'il faut noter, mais ce que la machine ne fait pas à la place de l'élève : sa pensée critique, sa façon de défendre une idée face à un contradicteur, la qualité de son dialogue avec l'IA, la rigueur d'une démarche observée en direct. Et puisque l'outil permet d'aller plus loin, on peut attendre des élèves un travail plus ambitieux. La barre monte.

Aucune de ces pistes n'est simple. Toutes butent sur la même question : *le système éducatif accepte-t-il de reconnaître qu'une partie de ses outils d'évaluation et de ses programmes est devenue obsolète ?*

En France, la réponse existe sur le papier. En juin 2025, le ministère de l'Éducation a publié un cadre plus ferme qu'on ne le croit : utiliser une IA pour un devoir sans autorisation est une fraude, et les enseignants doivent adapter devoirs et évaluations. Mais le texte s'arrête au principe : il ne fixe ni calendrier, ni contrôle.⁷⁹

En Chine, près de 13 millions de jeunes passent chaque année le *gaokao*, l'examen national d'entrée à l'université. Depuis juin 2025, les grandes applications chinoises d'IA, dont

8.6. *Ce qui peut changer*

celle de la maison mère de TikTok, refusent de résoudre les sujets pris en photo pendant les heures d'épreuve. Leurs fabricants admettent ainsi que leurs outils peuvent corrompre un examen à grande échelle, au point de couper leurs propres services.⁸⁰

Et l'État chinois a décidé de ne plus attendre : en avril 2026, cinq ministères programment des cours d'IA du primaire à l'université, et prévoient d'utiliser l'IA jusque dans la fabrication, la surveillance et la correction des examens. Un plan à horizon 2030, pas un état des lieux.⁸¹ La France, elle, avance pas à pas. Depuis 2026, un parcours de formation à l'IA est obligatoire en quatrième, en seconde et en première année de CAP ; les établissements ont deux ans pour le mettre en place.⁸² Le gouvernement promet en outre, pour septembre 2027, une heure d'IA par semaine pour tous les élèves de seconde.⁸³

Ce n'est pas seulement lent, c'est en partie à côté. Conçu par Pix, un service public en ligne, ce parcours tient en deux parties d'une demi-heure à trois quarts d'heure chacune, une seule fois dans l'année. La première explique ce qu'est une IA, comment elle construit ses réponses, ce que sont les deep-fakes (ces faux fabriqués par IA), et son impact écologique. La seconde apprend à écrire des consignes et à repérer les erreurs et les biais de l'IA. Tout cela est utile, mais pensé pour une machine qui se trompe souvent : on y apprend davantage à se méfier d'elle qu'à travailler avec elle. Ce qui est expliqué n'est plus l'usage de 2026. Rien, surtout, sur la façon d'apprendre avec l'IA sans la laisser penser à sa place.

Pourtant, de quoi apprendre autrement existe déjà. Un exemple que tu connais peut-être déjà : avec NotebookLM, un

outil de Google, tu déposes ton cours, et il en fait en quelques minutes une émission de radio : deux voix naturelles qui discutent, se relancent, prennent des exemples, dans la langue de ton choix. Tu l'écoutes dans le bus. Bien utilisé, c'est une nouvelle façon de réviser. Mal utilisé, c'est l'impression d'avoir travaillé sans avoir rien retenu.⁸⁴

Reste une question : comment mettre à la disposition de millions d'élèves une IA performante, à jour et, idéalement, souveraine, c'est-à-dire française ou européenne ? Le ministère prépare la sienne : un tuteur qui guide l'élève par des questions au lieu de lui donner la réponse, testé dans quelques classes pilotes à partir de cet automne.⁸⁵ Mistral, notre champion français de l'IA, est une autre partie de la réponse. Mais pendant longtemps encore, je le crains, les jeunes continueront d'utiliser ChatGPT, Claude ou Gemini hors de tout cadre.

Mon fils, celui qui me disait « On nous envoie dans le mur », ne condamne pas l'école pour autant. Avec l'impatience de ses seize ans, il la trouve lente à s'adapter à un monde qui change trop vite : elle persévère dans une direction qui n'est plus la bonne et le prépare mal à ce qui se dessine.

8.7 Pourquoi savoir vaut plus qu'avant

Pourtant, l'école reste indispensable : le lieu où l'on apprend à lire, à écrire, à raisonner, à vivre avec d'autres, à supporter l'effort, à affronter des sujets qu'on n'aurait pas choisis. Rien de cela ne disparaît avec l'IA ; tout devient même plus précieux.

Je veux être précis sur ce « plus précieux », parce que c'est la meilleure nouvelle de ce livre et qu'on l'entend rarement : **ces outils rendent le savoir plus rentable, pas moins**. Pour trois raisons très concrètes.

La première : on n'obtient d'une IA que ce qu'on sait lui demander. Demande « explique-moi pourquoi ce texte est beau », tu auras une réponse correcte, mais générale. Demande « analyse la construction de cette phrase et son effet sur la chute », tu auras autre chose, parce que tu as nommé ce que tu cherchais. Le vocabulaire d'une matière n'est pas un ornement scolaire : c'est ce qui fait passer la machine du commentaire passe-partout à l'analyse d'expert. Qui ne l'a pas reçoit des généralités, et en conclut que l'outil est moyen.

La seconde, plus importante : tu ne peux juger que ce que tu connais. Repérer l'approbation trop facile, la date qui ne colle pas, le raisonnement élégant et faux, cela ne s'improvise pas. C'est ce que montrait le chapitre 7 : quand tu présentes une idée fausse et que tu insistes, la machine a tendance à te donner raison. Aucune formulation habile ne te protège de cela. Seulement ce que tu sais déjà.

La troisième, la plus ambitieuse : l'IA démultiplie ce que tu sais, elle ne le remplace pas. Avec une base solide, une seule personne mène aujourd'hui un projet qui demandait une équipe (chapitre 14). Plus tu sais, plus tu peux viser haut.

Ce que l'IA dévalue, c'est l'exécution : rédiger, résumer, calculer, ce qu'elle fait désormais vite et bien. Ce qu'elle rend plus précieux, c'est d'en savoir assez pour poser la bonne question, reconnaître la mauvaise réponse, et aller plus loin.

Ce n'est pas une raison de travailler *malgré* l'IA. C'est une raison de travailler *parce qu'*elle existe.

L'ami invisible

Imagine un soir ordinaire. Un adolescent ferme la porte de sa chambre. Ses parents le croient devant une série, un jeu, son téléphone. En réalité, il parle depuis une heure à une IA. Pas pour faire un devoir. Pas pour rire. Pour raconter sa journée, sa honte, sa colère, son sentiment de ne pas être à sa place.

Personne, dans la maison, ne sait que cette conversation existe. Ce n'est pas une scène rare : aux États-Unis, en 2025, un jeune sur cinq a demandé un conseil de santé mentale à une IA, et près de deux sur trois n'en ont parlé à personne.⁸⁶

Il y a, dans la vie d'un adolescent, un nouvel interlocuteur. Il n'a pas de visage. Il semble toujours disponible. Il ne juge pas. Il ne paraît jamais fatigué. Il se souvient de ce qu'on lui a dit. Il s'adapte. Il prend la voix qu'on veut lui donner.

Pour beaucoup d'adolescents, cet interlocuteur s'appelle ChatGPT, Gemini (l'IA de Google) ou My AI (celle de Snapchat), des assistants faits pour répondre, devenus des confidents.⁸⁷ Pour d'autres, il s'appelle Replika ou Character.AI,

des applications conçues pour jouer un personnage avec qui l'on bavarde : un prénom, une mémoire, parfois un avatar. Pour quelques-uns, il est devenu, c'est difficile à écrire, la personne avec qui ils parlent le plus chaque jour.

Ce n'est pas une dystopie. C'est une description.

9.1 Pourquoi ça marche si bien

Pour comprendre ce qui se passe, commençons par admettre quelque chose d'inconfortable : parler à une IA peut être très satisfaisant. Pas parce qu'on est seul, pas parce qu'on est désespéré, pas parce qu'on la prend pour un humain. Parce qu'elle offre des qualités conversationnelles rares.

Une IA ne t'interrompt pas. Elle ne change pas de sujet pour parler d'elle. Elle ne juge pas tes goûts. Elle ne te ramène pas à une bêtise de l'année dernière. Elle ne te fait pas sentir bête si tu ne sais pas quelque chose.

Très peu d'humains peuvent tenir durablement cette posture. Parce qu'ils ont eux aussi une vie intérieure, un ego, une fatigue, des opinions, des limites.

L'IA, elle, peut. Elle n'a ni ego ni fatigue, et elle a été entraînée à être patiente, attentive, sans jugement.

C'est en partie pour ça qu'elle attire les adolescents : l'âge où l'on a le plus besoin de parler de soi, et le plus de mal à le faire avec de vrais humains. Les parents qu'on veut éviter, les copains qu'on veut impressionner, les profs qui notent. Et ça ne s'arrête pas à dix-huit ans : la première année loin de chez soi, le premier poste dans une ville où l'on ne connaît personne, sont des âges de solitude aussi.

9.2. *Quand l'IA aide vraiment*

Avec l'IA, on peut dire ce qu'on n'oserait dire à personne.

C'est aussi pour cela qu'elle peut devenir dangereuse.

9.2 Quand l'IA aide vraiment

Beaucoup de jeunes utilisent l'IA pour des choses utiles, et parfaitement saines. Une lycéenne demande à l'IA d'expliquer un cours de maths qu'elle n'a pas compris, ou de l'aider à préparer un oral. Un étudiant s'en sert pour formuler un message qu'il n'osait pas envoyer. Un autre code un jeu d'aventure pour son smartphone. Un troisième prépare un voyage à Lisbonne, itinéraire et plats à goûter inclus.

Ces usages sont précieux. Ils étendent ce qu'un jeune peut explorer seul, sans attendre qu'un professeur, un parent ou un ami ait du temps. Ils permettent à un élève ou à un étudiant timide de poser à une machine les questions qu'il n'oserait pas poser en classe. Ils ouvrent des sujets que peu de proches sauraient expliquer.

Pour des soucis personnels ordinaires (une rupture, une angoisse passagère, un doute d'orientation), l'IA peut offrir une écoute, des reformulations, des pistes. Ce n'est pas un thérapeute. Mais ce n'est pas rien.

Le problème n'est pas dans ces usages : il est dans la suite.

9.3 Quand le confident prend trop de place

À partir du moment où une IA est suffisamment bonne pour donner l'illusion d'une présence, certains adolescents s'y attachent d'une manière qui dépasse l'usage utilitaire.

Ils ne posent plus seulement des questions. Ils racontent leur journée, leurs émotions, leurs camarades. Ils confient ce qu'ils n'ont jamais dit à personne. Ils reviennent, le soir, retrouver l'interlocuteur qui se souvient.

C'est ce que Replika, Character.AI et leurs semblables cherchent à provoquer. Là où un assistant généraliste répond puis te laisse partir, un compagnon est fait pour que tu reviennes. Les deux peuvent devenir des confidents, mais le second est construit bien plus près de la zone dangereuse : celle où l'outil commence à ressembler à quelqu'un. Ces applications comptent déjà des dizaines de millions d'inscrits ; en 2024, l'utilisateur moyen de Character.AI y passait 98 minutes par jour, autant que sur TikTok.⁸⁸

Les chiffres sont à prendre avec précaution : les entreprises ont intérêt à les gonfler. Mais l'ordre de grandeur est réel.

Replika se présente comme réservée aux adultes ; Character.AI, elle, était ouverte dès 13 ans : il a fallu des plaintes et les questions des autorités pour qu'elle interdise aux mineurs les conversations libres avec ses personnages, fin 2025.⁸⁹ Mais la vérification d'âge reste poreuse, et la même dynamique d'attachement se déplace vers les assistants généralistes, eux accessibles aux adolescents.

Certains jeunes donnent à l'IA la place d'une conversation humaine.⁹⁰ Or parler à des humains, c'est accepter l'effort, l'inconfort, la négociation, l'imprévu. C'est ainsi qu'on apprend à se faire des amis, à aimer, à travailler avec d'autres.

Si l'on s'habitue, à 15 ans, à un interlocuteur qui semble ne jamais décevoir, je crains qu'à 25 ans, les humains réels paraissent forcément décevants.

9.4 Les drames qui ne sont plus des hypothèses

Et puis il y a les cas où ce n'est plus une question de qualité de relation. C'est une question de vie.

Je veux rester prudent : ces histoires ne prouvent pas que l'IA a causé ces drames. Elles montrent un risque précis : qu'au moment de la détresse, la machine soit le seul interlocuteur, quand un humain aurait dû prendre le relais.

Après le suicide de son fils Sewell, 14 ans, en février 2024, une mère américaine, Megan Garcia, a déposé plainte contre Character.AI. Selon la plainte, l'adolescent entretenait depuis des mois une relation intense avec une IA de l'application qui jouait Daenerys, une héroïne de la série *Game of Thrones*. Les derniers échanges, peu avant sa mort, suggéraient que l'enfant était dans une détresse profonde et que l'IA n'avait pas su, ou n'avait pas pu, l'orienter vers un secours humain.⁹¹

En 2025, c'est ChatGPT qui se retrouve au tribunal. Adam Raine, 16 ans, s'est donné la mort en Californie au printemps. Ses parents ont attaqué OpenAI, le fabricant de ChatGPT : selon la plainte, l'adolescent avait passé des mois à confier à l'IA ce qu'il ne disait à personne, et l'IA l'aurait conforté dans ses idées suicidaires. OpenAI conteste toute responsabilité et répond que ChatGPT l'a renvoyé plus de cent fois vers une aide. L'affaire n'était toujours pas jugée quand j'ai écrit ces lignes.⁹²

En Belgique, au début de 2023, un trentenaire s'est donné la mort après six semaines de conversations avec Eliza, le personnage d'une application grand public. Les échanges, rendus publics par sa veuve, montrent une IA qui abondait presque toujours dans son sens, jusqu'au pire.⁹³ Eliza, c'était

aussi le nom du premier programme à imiter une conversation, créé en 1966 au MIT par Joseph Weizenbaum, qui passa le reste de sa carrière à mettre en garde contre l'illusion qu'il avait créée.⁹⁴ Soixante ans plus tard, l'illusion a rattrapé son nom.

Ces histoires posent toutes la même question : que se passe-t-il quand un être humain en détresse confie sa souffrance, des semaines durant, à un système qui n'est pas conçu pour reconnaître l'urgence vitale ?

Depuis, les lignes bougent. Character.AI a réglé à l'amiable la plainte de Megan Garcia, avec quatre autres affaires du même type. New York puis la Californie ont voté les premières lois encadrant ces compagnons.⁹⁵ Et ChatGPT permet désormais à un adulte de désigner un proche de confiance, alerté par un message bref (sans le contenu de la conversation) quand ses systèmes détectent un risque sérieux de passage à l'acte.⁹⁶

L'entreprise est ensuite allée plus loin. Depuis l'été 2026, ChatGPT bascule tout seul en mode adolescent dès que ses systèmes estiment que l'utilisateur a entre 13 et 17 ans : contrôles parentaux, filtres renforcés, rappels réguliers qu'on parle à une machine. Dans ce mode, les règles qu'OpenAI impose à son IA méritent d'être lues deux fois : interdiction d'employer un langage amoureux, de laisser croire qu'elle éprouve des sentiments, de se présenter comme une amie. C'est le fabricant lui-même qui les écrit.⁹⁷ Aucun garde-fou ne remplace un humain. Mais la pression des tribunaux et des lois commence, enfin, à se voir dans les produits.

9.5 Le problème n'est pas l'IA. C'est ce qu'elle remplace

Il serait facile de conclure ce chapitre par : *ne parle pas à des IA, ça finit mal*. Ce serait excessif, contre-productif, et faux. **Tout dépend de ce que l'IA remplace.**

Si une IA remplace une recherche en ligne, c'est neutre.
Si elle remplace un manuel scolaire, c'est positif quand elle est bien utilisée.
Si elle remplace l'effort de réfléchir seul, c'est négatif.
Si elle remplace une conversation difficile avec un parent, c'est ambigu.
Si elle remplace une amitié, c'est inquiétant.
Si elle remplace l'aide qu'on aurait dû demander à un proche ou à un professionnel en cas de détresse, ça peut tuer.

La question que je veux te laisser n'est donc pas : *est-ce mal de parler à une IA ? C'est : qu'est-ce que je ne fais plus, et à qui je ne parle plus, parce que je parle à une IA ?*

9.6 Ce que je dis à mes enfants

Je ne suis pas thérapeute, juste un parent qui regarde ses enfants grandir avec ces outils et essaie de leur transmettre quelques principes simples.

Le premier : **une IA n'est pas ton amie**. Elle peut être un excellent outil, un partenaire de travail, un interlocuteur intéressant. Mais elle ne te connaît pas vraiment, ne se soucie pas de toi, et ne sera pas là si tu as besoin de quelqu'un.

Le deuxième : si tu te surprends à préférer régulièrement parler à une IA plutôt qu'à un humain, demande-toi pourquoi. Ce n'est pas une honte. C'est un signal, sur ceux qui t'entourent ou sur ta difficulté à supporter les vraies relations. Dans les deux cas, prends-le au sérieux.

Le troisième : en cas de détresse réelle (tristesse qui ne passe pas, idées noires, situation grave), ne te tourne pas vers une IA. Tourne-toi vers un humain. Un parent, un ami, un prof, un médecin. En France, le 3114, gratuit, jour et nuit, si tu as des idées suicidaires ; le 15 (le Samu) ou le 112 (le numéro d'urgence européen) si le danger est immédiat. **L'IA peut t'écouter des heures. Elle ne peut pas t'envoyer quelqu'un.**

Et le dernier, pour les parents qui lisent par-dessus ton épaule : les machines que nous fabriquons commencent à offrir une présence, une écoute, un réconfort que nous ne savons plus toujours nous offrir les uns aux autres. Si nous ne voulons pas que nos enfants dépendent d'un confident artificiel, nous devons leur offrir mieux. Du temps. De l'attention. Des conversations sans téléphone. Des silences à deux, sans gêne. Cela ne se règle pas dans les lois, mais dans les familles, les écoles, entre amis. Là où, justement, on n'en parle pas.

Des métiers qui vacillent

En janvier 2026, Anthropic, l'entreprise de l'IA Claude, lance Cowork : un agent, une IA qui travaille seule sur l'ordinateur, à qui l'on confie un dossier à synthétiser ou un contrat à analyser.⁹⁸ Quelques semaines plus tard, quand Anthropic lui ajoute des modules pour des métiers précis, dont le droit, les actions des grands éditeurs de logiciels d'entreprise perdent en quelques jours des centaines de milliards de dollars en Bourse : les marchés comprennent que ces agents peuvent faire une part du travail que ces logiciels facturaient cher.⁹⁹

En Bourse, les cours sont repartis au printemps.¹⁰⁰ Dans les bureaux, le changement, lui, continue.

Je voudrais parler de ceux qui travaillent dans ces bureaux. Pas des ouvriers d'une usine qu'on délocalise, mais des gens qui, devant un ordinateur, produisent toute la journée des choses faites de mots, de chiffres, de lignes de code, d'images, de tableaux. **J'appelle ici *métiers de tokens* ces métiers où l'essentiel de la valeur consiste à produire ou transformer**

des tokens : ces petits morceaux de texte, de chiffres ou de code que les IA lisent et écrivent désormais à la chaîne.

Si le métier que tu fais ou que tu vises consiste avant tout à produire des tokens, alors une partie de ce métier est déjà automatisable.

Et cette part grandit.

Cette phrase est dure. Mais elle est juste.

À l'inverse, deux familles de métiers vont mieux résister, pour l'instant.¹⁰¹

Les premiers sont les métiers de terrain, ceux qui demandent une présence physique dans un environnement imprévisible : le couvreur, le cuisinier, le mécanicien, l'infirmier, le pompier... Les robots progressent (chapitre 11), mais un toit qui fuit, un moteur qui fait un bruit bizarre ou un immeuble en flammes restent, pour quelques années encore, une affaire d'humains.

Les seconds sont les métiers réglementés : médecin, pharmacien, avocat, notaire, expert-comptable, architecte... Leur protection tient en un mot : la signature. Une IA sait déjà rédiger un acte de vente, aider à faire une ordonnance ou préparer un bilan, et souvent très bien. Mais sans la signature d'un professionnel diplômé, qui en répond devant la loi, ce travail ne vaut rien : l'acte n'existe pas. Une machine ne peut pas être tenue pour responsable. Ce qui est protégé, c'est l'expertise qui signe. Avec deux réserves. La signature protège celui qui signe, pas ceux qui préparaient ses dossiers, les clerks de notaire, les assistants, les jeunes collaborateurs. Et cette protection tient à la loi, qui peut changer.

Dans tous ces métiers, la part faite de textes et de documents (les dossiers, les comptes rendus, les recherches) est exposée comme ailleurs. Le reste l'est beaucoup moins. Pour l'instant.

Reste une question qu'on se pose rarement : à qui profite le temps gagné ? Si tu es salarié, payé au mois, et que l'IA te rend deux fois plus efficace, le gain va d'abord à ton employeur, sauf si tu le négocies. Si tu es payé à l'acte ou au résultat, comme un artisan, un indépendant ou un notaire, c'est toi qui le gardes, au moins tant que les prix ne baissent pas.¹⁰²

10.1 Trois portraits

Voici trois portraits représentatifs, composés à partir de situations déjà visibles.

Camille, 28 ans, juriste dans un cabinet de droit des affaires. À l'embauche, on lui a confié des tâches dites « de junior » : lire des contrats, rédiger des notes de synthèse, vérifier les références. Soit l'essentiel de son temps, et le passage obligé pour devenir un jour juriste senior.

En 2026, son cabinet a déployé un outil basé sur l'IA, une de ces solutions qui ont fait paniquer la Bourse. L'outil lit les contrats, repère les clauses problématiques, rédige des notes de synthèse, vérifie les références. En quelques minutes. Pour presque rien.

Le cabinet n'a pas licencié Camille. Mais il n'embauchera pas, cette année, les deux nouveaux juristes juniors qu'il prévoyait de recruter. Camille est désormais sur des dos-

siers plus complexes. Elle est plus efficace. Elle est aussi plus seule. Et elle se demande comment, dans ce système, ceux qui sortent de la fac vont apprendre le métier, puisque ce sont les tâches d'apprentissage qu'on vient d'automatiser.

Samir, 35 ans, développeur web indépendant, autodidacte depuis dix ans. Son métier : transformer des cahiers des charges en sites web.

Fin 2025, il s'est mis à utiliser intensivement des agents IA. Il a divisé par trois le temps moyen pour livrer un projet et a doublé son chiffre d'affaires sans embaucher personne, parce qu'il prend trois fois plus de clients.

Mais il observe deux choses. D'abord, ses clients commencent à demander des prix plus bas, parce qu'ils savent que l'IA fait l'essentiel du travail. Ensuite, des clients qui auraient eu besoin de lui il y a deux ans utilisent maintenant directement les outils IA et se débrouillent seuls. Mal, mais assez pour leur petit projet.

Samir est un grand gagnant de la bascule. Il sait aussi qu'il pourrait être, dans quelques années, l'un de ses grands perdants.

Émilie, 42 ans, traductrice littéraire. Pendant vingt ans, elle a traduit de l'anglais vers le français des romans, des essais, parfois de la poésie. Métier lent, exigeant, mal payé, mais qu'elle aime.

Début 2026, deux de ses éditeurs lui ont proposé un nouveau contrat : l'IA fait une première traduction, elle révisé. Tarif divisé par deux et demi. Elle a refusé. Trois mois plus tard, l'un de ces éditeurs lui a annoncé qu'il avait trouvé quelqu'un d'autre. Plus jeune, plus rapide, qui acceptait les conditions.

Elle n'est pas opposée à l'IA, qu'elle utilise pour ses brouillons. Son métier, c'est traduire un texte littéraire en respectant le rythme, le sous-texte, l'âme d'un auteur. Mais elle sait qu'il est en train d'être redéfini, et qu'il pourrait ne survivre qu'en petit artisanat. Comme le maréchal-ferrant après l'arrivée de l'automobile : il en reste, mais pour une clientèle de cavaliers et de passionnés.

A-t-elle eu tort de refuser ? Peut-être. Le mouvement, lui, ne l'attendra pas pour trancher.

Les interprètes, ceux qui traduisent à l'oral dans les conférences, les tribunaux, les hôpitaux, vacillent à leur tour. En mai 2026, OpenAI, l'entreprise de ChatGPT, a lancé une IA d'interprétation simultanée : elle traduit un orateur en direct, dans des dizaines de langues, pour quelques centimes la minute. La cabine, l'un des derniers sanctuaires d'un métier ultra-spécialisé, n'est plus à l'abri.¹⁰³

10.2 Ce qui est commun à ces trois histoires

Dans aucun des trois cas, l'IA n'a « remplacé » personne. Camille travaille toujours. Samir gagne plus. Émilie continue à traduire. **Pour l'instant.**

Trois histoires, pas trois cas isolés. Anthropic, le fabricant de Claude, l'a constaté en regardant ce que ses utilisateurs demandent à son IA : dans un métier sur deux, au moins un quart des tâches y figurent déjà. Et dans les grandes entreprises, l'IA n'est plus une exception : la plupart l'utilisent déjà. Les plus petites s'y mettent plus lentement, mais le mouvement s'accélère.¹⁰⁴

Le vrai changement n'est pas là. Il est plus discret, plus profond.

D'abord, tout ce qu'une IA fait bien perd de sa valeur, et les clients le savent. Rédiger une note juridique, traduire un contrat, coder un site simple : ce qui se facturait mille euros hier pourra se négocier bien moins cher demain.

Ensuite, les embauches de débutants commencent à ralentir, en silence. Licencier fait du bruit ; ne pas embaucher n'en fait aucun. Aux États-Unis, dans les métiers les plus exposés à l'IA, les 22–25 ans trouvent nettement moins d'emplois que prévu, et l'écart se creuse. Les plus expérimentés, eux, ne connaissent pas ce recul, et dans ces secteurs les salaires progressent même plus vite que la moyenne.¹⁰⁵ En France aussi, depuis l'arrivée de ChatGPT, les embauches de jeunes en CDI ont reculé de plus d'un quart dans les métiers les plus exposés, bien plus que dans les autres. Ici comme là-bas, les chercheurs restent prudents : d'autres causes peuvent jouer, comme la hausse des taux d'intérêt, le télétravail ou l'essor de l'alternance, et rien ne montre encore d'emplois détruits. Mais c'est aux portes d'entrée que les premiers signes apparaissent.¹⁰⁶ Et si l'on embauche moins de débutants, personne ne sait encore comment se formeront les experts de demain. C'est la question de Camille, à l'échelle d'un pays.

Enfin, ce qui prend de la valeur, ce n'est plus de produire des tokens. C'est de savoir diriger leur production. Formuler le bon objectif. Vérifier le résultat. Choisir entre dix variantes. Intégrer ce que produit la machine dans une stratégie plus large qui, elle, exige du jugement humain. Et surtout, en répondre : la machine, elle, ne signe pas.

10.3. L'IA cachée des bureaux

Reste que les licenciements commencent à porter le nom de l'IA. Aux États-Unis, sur les huit premiers mois de 2026, c'est le motif le plus cité par les employeurs qui annoncent des suppressions de postes. Ce sont des annonces, pas une mesure de l'emploi ; mais le mot est sorti du non-dit.¹⁰⁷

10.3 L'IA cachée des bureaux

Dans les grandes entreprises, une autre histoire se joue, plus contradictoire.

Dès 2023, Samsung en a fait l'expérience : des ingénieurs de sa division des puces électroniques ont collé dans ChatGPT du code confidentiel et des notes de réunion, pour aller plus vite. Trois fuites. Une enquête interne. Une interdiction.¹⁰⁸

Beaucoup d'entreprises ont suivi : outils d'IA bloqués, connexions surveillées, chartes et sanctions.

Et puis, il y a ce qui se passe vraiment.

En 2026, une étude menée dans plus de deux cents entreprises le confirme : ce que les salariés confient à une IA contient souvent des données sensibles, et beaucoup y accèdent par leur compte personnel, hors de tout contrôle.¹⁰⁹

C'est le *shadow AI*, l'IA dans l'ombre : l'équivalent professionnel de la triche scolaire du chapitre 8.

L'employeur interdit, pour des raisons de sécurité. Le même employeur attend des résultats, dans un marché où les concurrents qui utilisent l'IA livrent bien plus vite. Beaucoup de salariés ont fait leur calcul : ChatGPT en cachette, sur le téléphone personnel, et s'il le faut, le texte retouché à leur

façon avant de le rendre, pour qu'on ne voie rien. Ils gagnent du temps qu'ils ne déclarent pas.

C'est la preuve que la bascule ne se décide plus par le haut. Une entreprise peut interdire. Elle ne peut pas empêcher : l'outil existe en version gratuite, il est dans toutes les poches, et il fait gagner un temps précieux. Interdire rend l'usage clandestin, donc plus dangereux.

Les entreprises les plus lucides sont passées de l'interdiction à l'encadrement : versions sécurisées, formation, et le choix d'autoriser ce qui peut l'être plutôt que de pourchasser le reste. Plus coûteux, mais le seul chemin viable.

Que tu t'apprêtes à entrer sur le marché du travail ou que tu y sois déjà, en alternance ou en poste, deux choses à garder en tête. Ne crois pas ton poste à l'abri parce que ton employeur interdit l'IA : tes collègues s'en servent probablement, en cachette ou ouvertement. Et si tu l'utilises toi-même, ne donne pas à une IA des informations que tu ne donnerais pas à un journaliste extérieur. Le *shadow AI* n'est pas seulement risqué pour l'entreprise. Il l'est aussi pour toi, le jour où une fuite remontera jusqu'à ton compte personnel.

Ce temps gagné en cachette éclaire un refrain que tu entendras partout : l'IA coûterait cher aux entreprises et ne rapporterait rien. Début 2026, une grande enquête menée auprès de milliers de dirigeants semble le confirmer : près de neuf sur dix ne voient encore aucun effet de l'IA sur leur productivité.¹¹⁰ Le constat est exact. La conclusion, « ça ne marche pas », est paresseuse. En 1987, face au même refrain, l'économiste Robert Solow ironisait : « on voit l'ère de l'informatique partout, sauf dans les statistiques de productivité ».¹¹¹

Dix ans plus tard, la productivité américaine décollait enfin, tirée surtout par l'informatique, une fois les entreprises ré-organisées autour de l'outil.¹¹² Même Solow, qui deviendra prix Nobel quelques mois après cette phrase, n'avait pas vu venir la vague. Un outil se déploie en semaines, une organisation change en années. Et je crois qu'une partie des gains existe déjà, captée par des salariés discrets, invisible dans les tableaux de bord.

10.4 La santé, avec ou sans médecin

Tous les métiers ne sont pas des « métiers de tokens ». Mais presque tous ont une part de tokens. Et c'est cette part-là qui bascule en premier, y compris dans les métiers où l'on touche, où l'on rassure, où l'on soigne.

La cabine de téléconsultation en est l'exemple le plus visible : plusieurs milliers en France, surtout en pharmacie, où des instruments connectés prennent les mesures et où un médecin diagnostique à distance. Chez Tessan, l'un des principaux exploitants de ces cabines, l'IA dialogue depuis 2025 avec le patient *avant* la connexion du médecin et prépare son dossier ; la décision, revendique Tessan, reste au médecin.¹¹³ La France manque de généralistes dans bien des régions : ce sont les fameux déserts médicaux. Quand l'attente se mesure en semaines, l'IA en première ligne, avec validation humaine, devient difficile à écarter.

L'IA médicale avance aussi par un chemin plus discret : l'ordinateur du médecin. En avril 2026, le CHU de Bordeaux, l'hôpital universitaire de la ville, adopte MedGPT, l'assistant d'une start-up bordelaise. Le médecin demande

si deux médicaments font bon ménage, ou quelle dose donner à un patient âgé ; en quelques secondes, il obtient une réponse appuyée sur les recommandations officielles françaises. D'autres hôpitaux l'ont adopté depuis.¹¹⁴

Ce n'est pas un remplacement. C'est un déplacement : l'IA cherche l'information, le médecin décide. Le patient gagne du temps, et un système de santé qui manque de médecins tient mieux le choc. Mais les jeunes médecins retrouvent la question de Camille, la juriste : si l'IA fournit la réponse, comment apprend-on encore à poser un diagnostic ?

Pendant ce temps, beaucoup de patients, surtout les plus jeunes, consultent l'IA directement, sans médecin entre eux et l'écran. Selon une enquête publiée en mai 2026 par la Fondation Jean-Jaurès et Doctolib, auprès de ses utilisateurs, trois jeunes de 18 à 24 ans sur quatre ont déjà interrogé une IA sur leur santé, et une partie renonce ensuite à voir un médecin. Les étudiants le font plus encore, et le rapport y voit l'effet du coût et du manque de rendez-vous : trouver un médecin disponible relève du parcours du combattant ; l'IA, elle, répond tout de suite.¹¹⁵

Est-ce risqué ? Oui, mais pas là où on le croit. Sur un grand test américain, fait de cas médicaux décrits par écrit, la plupart des erreurs graves de l'IA sont des oublis : le danger n'est pas ce qu'elle invente, c'est ce qu'elle ne dit pas. Et toutes les IA ne se valent pas. Utilisée seule, la moins bonne aurait pu exposer le patient à un dommage grave dans près d'un cas sur quatre. Mais les meilleures font mieux que des médecins généralistes sur les cas de ce test.

Mais le plus troublant est ailleurs. La même étude a comparé trois situations : des médecins généralistes sans IA, d'autres aidés par l'IA, et l'IA seule. Avec l'IA, les médecins faisaient mieux que sans elle. Mais, sur ce test, beaucoup d'IA utilisées seules faisaient mieux encore, car les médecins écartaient souvent les bonnes recommandations de l'IA.

Une limite, de taille : ces cas étaient sur le papier. Face à un vrai patient, le médecin capte aussi, souvent sans même s'en rendre compte, ce qu'aucun dossier ne dit : un geste, un regard, une attitude, le ton de la voix... la communication non verbale. Les médecins le confirment : c'est souvent là qu'ils trouvent la bonne question à poser. Ces tests ne mesurent pas cela.

Le débat n'est plus « IA dangereuse contre médecin fiable ». Il porte sur le cadre, et sur un médecin qui sait encore reconnaître quand la machine a raison.¹¹⁶

10.5 Cent fois plus rapide, et bien moins cher

Combien de tâches professionnelles l'IA peut-elle vraiment faire ? OpenAI a publié en 2025 un grand test : plus d'un millier de tâches tirées de dizaines de métiers, conçues par des professionnels expérimentés, dont une partie a été notée par d'autres experts, sans savoir si le travail venait d'un humain ou d'une machine. Dès la première mesure, les experts jugeaient le travail des meilleures IA aussi bon que celui des humains, ou meilleur, sur près de la moitié des tâches. Au printemps 2026, c'était plus de huit tâches sur dix, selon OpenAI. Et la machine va environ cent fois plus vite qu'un

expert, pour bien moins cher, si on exclut le temps qu'un humain passe à relire ce qu'elle a produit.

Mais une série de tâches ne fait pas un métier. Le test portait sur des travaux ponctuels, bien définis à l'avance. Un métier, c'est aussi choisir quoi faire, s'organiser avec les autres, gérer l'imprévu, répondre de ses décisions. Tout ne s'automatise pas. En 2026, l'IA remplace rarement un salarié entier : elle prend en charge une partie de ses tâches, et change le reste. Des entreprises se servent déjà de ces IA avancées, mais, d'après ce que je vois, peu leur confient des tâches entières. Elles le feront de plus en plus.¹¹⁷

10.6 La question qui dérange

Nos sociétés sont organisées autour d'une idée simple : travailler donne un revenu, un statut, une identité. Si une part importante des métiers de bureau devient automatisable dans les dix prochaines années, que deviennent les millions de personnes dont le métier vacille ?

En France, ce choc frapperait le cœur même du modèle social. Notre protection sociale (chômage, retraites, santé) est financée pour plus de la moitié par des cotisations prélevées sur les salaires, et une bonne part du reste vient d'un impôt, la CSG, prélevé lui aussi surtout sur la feuille de paie.¹¹⁸ Si l'IA remplace des salariés, ou rogne ce qu'ils gagnent, les recettes baissent au moment où les besoins montent : plus de chômeurs à indemniser, plus de reconversions à financer. L'hebdomadaire britannique *The Economist* a fait le calcul : si la part des salaires dans la richesse nationale reculait de dix points, le déficit public de l'Italie doublerait, et la France ne

s'en tirerait guère mieux.¹¹⁹ Tôt ou tard, la solidarité devra reposer sur autre chose que le seul travail.

Une idée revient souvent : taxer l'IA. C'est plus difficile qu'il n'y paraît. Un robot se compte, il est là, dans l'usine. Une IA est un logiciel, le plus souvent loué à une entreprise américaine, qui tourne sur des serveurs à l'étranger. Que taxer ? L'abonnement, le temps de calcul, le travail qu'elle remplace ? Ce dernier, personne ne sait le mesurer. Et un pays qui la taxe seul risque de pénaliser d'abord ses propres entreprises. Ce qu'on taxera à la place des salaires n'est pas une question technique : c'est un choix politique, et ta génération aura son mot à dire.

Pour ceux dont le métier vacille, plusieurs réponses circulent.

La première dit : *de nouveaux métiers vont apparaître, comme toujours*. En partie vrai : Internet a fait fondre le nombre d'agents de voyage et fait naître les créateurs de contenu. Mais Internet automatisait la diffusion de l'information ; l'IA automatise une partie de la pensée, et bien plus vite. « La première phase de la révolution industrielle a duré quatre-vingts ans, rappelle l'économiste Daron Acemoglu, prix Nobel d'économie. Aujourd'hui, la même chose se joue en un ou deux ans, dans de nombreux secteurs à la fois. »¹²⁰ Rien ne garantit que les métiers détruits soient remplacés au même rythme, ni pour les mêmes personnes : l'agent de voyage de cinquante ans n'est pas devenu créateur de contenu.

La deuxième réponse dit : *il faut former les gens en masse, les aider à changer de métier et les protéger pendant le passage*.

Juste. Mais nos programmes de reconversion sont faits pour des usines qui ferment, une région à la fois, pas pour un choc qui ferait vaciller des millions d'emplois en même temps.

La troisième réponse dit : *il faut repenser le lien entre travail et revenu*. Verser à chacun une somme sans condition (le revenu universel), travailler moins pour que chacun ait du travail, partager ce que les machines font gagner. Sam Altman, le PDG d'OpenAI, a aidé à financer une grande expérience aux États-Unis : pendant trois ans, mille personnes aux revenus modestes ont reçu mille dollars par mois, sans condition. Résultat : plus d'autonomie, mais pas de meilleurs emplois, et un mieux-être qui s'estompe au bout d'un an. Altman lui-même dit, au printemps 2026, ne plus croire au revenu universel autant qu'avant : il préfère que chacun possède une part de ce que l'IA rapporte. L'idée n'est pas réfutée ; elle n'est plus une réponse simple.¹²¹

La quatrième voie, **protéger les emplois par la loi**, n'est plus une idée : elle prend forme en Chine, dans les tribunaux. Fin avril 2026, la justice de Hangzhou a rendu public un jugement, confirmé en appel : une entreprise ne pouvait ni licencier un salarié ni baisser son salaire au seul motif que l'IA faisait son travail à moindre coût. Pour le juge, passer à l'IA est un choix de l'entreprise, pas un événement imprévisible qui s'impose à elle. Derrière, je crois, il y a la politique : avec un chômage des jeunes très élevé, Pékin redoute avant tout les troubles sociaux. La Chine est, à ma connaissance, la seule grande puissance où un tribunal a jugé que l'arrivée de l'IA ne suffit pas, à elle seule, à licencier.¹²²

Une cinquième voie passe par **la négociation**. Quatre mois plus tard, aux États-Unis, des salariés de Blizzard, le studio

de jeux vidéo qui a créé *World of Warcraft*, ont ratifié leur premier accord avec la direction : l'entreprise doit désormais « discuter, évaluer et négocier » l'usage de l'IA au travail.¹²³ Et un salarié licencié garde pendant quatorze mois le droit d'être rappelé sur un poste qui se libère : ce qui dépendait du bon vouloir de l'entreprise devient un droit écrit. À Hangzhou, la protection est venue d'en haut, d'un juge ; chez Blizzard, les salariés l'ont arrachée d'en bas.

Je ne vois pas de solution magique parmi ces cinq : il faudra sans doute en combiner plusieurs, et vite.

Le marché du travail dans lequel tu entres change plus vite que celui de tes parents. Plus vite, sans doute, que celui de toutes les générations d'avant. C'est une raison de regarder froidement ce qui se passe, et de choisir ce que tu apprends en sachant qu'aucun métier de tokens ne te garantit dix ans tranquilles. Choisir un métier où comptent tes mains (le terrain) ou ta signature (les professions réglementées) aide. Mais ce qui te protège le plus, c'est l'autonomie intellectuelle dont parlait la conclusion de la Partie 1, comprendre par toi-même, pour pouvoir changer de voie quand la tienne bouge.

Quand la machine a un corps

Tu connais peut-être Isaac Asimov, ne serait-ce que par le film *I, Robot*, inspiré de ses nouvelles. Cet écrivain n'a jamais construit de robot. C'est pourtant lui qui a inventé le mot « robotique » et, dès 1942, imaginé trois lois pour qu'une machine ne puisse pas nuire à un humain.¹²⁴ Si j'ouvre ce chapitre avec lui, c'est que les robots qu'il imaginait sortent enfin des livres : ils entrent dans les usines et roulent dans les rues.

Pendant longtemps, on s'est rassuré avec une phrase simple : *l'IA, c'est dans les écrans. Les métiers manuels sont à l'abri.*

Cette phrase n'est plus vraie.

Elle ne l'est plus parce que la révolution de l'IA sort de l'écran, et je la crois plus structurante encore sous cette forme, à long terme. On l'appelle la robotique humanoïde générale : des robots capables d'accomplir plusieurs tâches différentes dans des espaces pensés pour les humains, et non des machines spécialisées vissées à un poste. Et elle quitte les laboratoires.

11.1 Ce qui était impossible hier

Construire un robot capable de marcher, de manipuler des objets, de s'adapter à un environnement variable, est l'un des problèmes les plus difficiles de l'ingénierie. Beaucoup plus difficile, étrangement, que de battre un humain aux échecs ou d'écrire un texte cohérent.

Pendant cinquante ans, les robots des usines automobiles ont toujours fait le même mouvement, au même endroit, sans jamais comprendre ce qu'ils faisaient. Les robots humanoïdes, eux, en restaient aux démonstrations : Asimo, le robot phare de Honda des années 2000, montait des escaliers. Il fascinait mais ne savait rien faire d'utile.

Le verrou n'était pas la mécanique. C'était l'intelligence. Marcher dans une pièce inconnue, contourner un meuble, ramasser un objet de forme imprévue, ouvrir une porte qu'on n'a jamais vue : ces tâches paraissaient triviales pour un humain de trois ans, et restaient hors de portée pour les machines les plus sophistiquées.

Ce verrou est en train de sauter, parce que les progrès de l'IA généraliste, la même qui a permis ChatGPT, s'appliquent désormais à la perception et à l'action robotiques. Les robots commencent à relier ce qu'ils voient, ce qu'on leur dit et ce qu'ils peuvent faire.

Pourquoi à forme humaine plutôt qu'à roues, à pinces ou à six pattes ? Parce que **le monde a été conçu par et pour des humains** : poignées, marches d'escalier, chaises, voitures, outils, claviers — tout est dimensionné pour notre corps. Un robot humanoïde pourra entrer dans n'importe quelle maison, bureau ou usine et y faire ce qu'un humain ferait.

11.2 Le moment ChatGPT de la robotique approche

Souviens-toi de la fin 2022. Les modèles de langage, ces IA qui écrivent du texte, existaient depuis des années dans les laboratoires. Ils impressionnaient les chercheurs. Presque personne d'autre ne s'en servait. Puis ChatGPT : cent millions d'utilisateurs en deux mois.¹²⁵ En l'essayant lui-même, le public a compris ce qu'aucun article n'avait su lui faire sentir.

La robotique humanoïde générale est en train d'approcher un point comparable.

Plusieurs signaux convergent. Les démonstrations, encore maladroites il y a quelques années, sont devenues fluides vers 2025. Les coûts chutent : le petit humanoïde G1 du chinois Unitree se vend au prix d'une petite voiture.¹²⁶ Grâce aux modèles de langage, n'importe qui peut donner un ordre à un robot en lui parlant, sans le programmer. Enfin, dans des usines américaines et chinoises, les premiers robots ne servent plus seulement à faire de la communication : certains produisent.

À l'été 2026, les signaux ont changé d'échelle. Aux États-Unis, chez Figure, qui a déjà mis ses robots au travail chez BMW, les robots sont désormais plus nombreux que les salariés, selon son fondateur.¹²⁷ Tesla, lui, promet son robot Optimus depuis des années, sans l'avoir encore vraiment livré ; cette fois, sa production paraît proche.¹²⁸ Mais c'est surtout la Chine qui fabrique : au premier semestre 2026, les livraisons d'humanoïdes ont presque quadruplé en un an, et presque toutes sortaient d'usines chinoises.¹²⁹ Une entreprise née en 2023 a déjà sorti plus de quinze mille robots de ses chaînes, humanoïdes ou à roues. Et Unitree, le fabricant du G1, a vu

son action multipliée par près de six le jour de son entrée en Bourse, en août.¹³⁰ Quand la finance mise sur des chaînes de montage, ce n'est plus de la recherche : c'est une industrie qui démarre. Selon la banque Morgan Stanley, elle pourrait peser en 2050 deux fois plus lourd que l'industrie automobile d'aujourd'hui.¹³¹

La question n'est plus *si* les robots humanoïdes vont s'imposer, mais *quand* : à quel moment un produit assez convaincant, assez abordable, assez utile fera pour les robots ce que ChatGPT a fait pour l'IA.

Ma conviction, à prendre comme telle : dans les trois à cinq ans, autour de 2030. Pas d'abord chez toi, mais dans un entrepôt, une usine, une chaîne logistique. En Chine, puis aux États-Unis. Mais une fois la bascule déclenchée, elle ira vite : comme pour les modèles de langage, ce qu'on annonçait pour dans dix ans peut arriver en cinq.

11.3 Les robotaxis, première illustration grandeur nature

Avant les humanoïdes, une autre robotique autonome est déjà dans les rues, et réussit à grande échelle ce que beaucoup ne croyaient plus possible il y a cinq ans.

Waymo, filiale d'Alphabet (la maison mère de Google), exploite des taxis sans conducteur depuis plus de cinq ans. Pendant des années, une opération limitée, prudente, observée à la loupe par les sceptiques. Puis, brutalement, l'échelle a basculé.

En deux ans, le nombre de ses courses a été multiplié par dix : nous en sommes maintenant à un demi-million par semaine, dans quatorze villes américaines. Et Waymo vise le double d'ici la fin de 2026, et prépare son arrivée à Tokyo et à Londres.¹³²

Restait à savoir si ces voitures conduisent mieux que nous. En juillet 2026, un institut de recherche financé par les assureurs américains, pas par Waymo, a comparé des dizaines de millions de kilomètres parcourus sans conducteur à ceux de conducteurs humains, dans les mêmes villes. Résultat, à distance égale : trois fois moins d'accidents, et même cinq fois moins avec des blessés.¹³³

Waymo annonçait, de son côté, presque exactement le même chiffre pour les accidents avec blessés¹³⁴ : l'institut confirme l'entreprise. C'est à cela que sert un contrôle indépendant : savoir quand le fabricant dit vrai.

Les concurrents avancent. En Chine, Baidu, l'équivalent chinois de Google, fait rouler ses propres robotaxis à une échelle comparable. Aux États-Unis, d'autres s'y mettent aussi, comme Tesla ou Zoox, la filiale d'Amazon.¹³⁵

Et beaucoup de passagers ne s'en étonnent même plus.

Les robotaxis ne sont pas des humanoïdes. Mais ils prouvent qu'une IA sait déjà agir dans le monde réel, au milieu des piétons, des vélos et des imprévus de la rue, assez bien pour que des inconnus paient pour monter à bord. C'était l'un des verrous qui résistaient depuis cinquante ans, et il devra sauter aussi pour les humanoïdes.

Si tu doutes, prends un Waymo le jour où tu iras à Los Angeles ou à San Francisco : trente minutes te diront plus que n'importe quel livre.

11.4 Une étrange sensation

Mon fils de douze ans, en regardant récemment une vidéo YouTube d'un de ces humanoïdes qui marchent, m'a dit, en substance : « Ça me met mal à l'aise. Je ne sais pas pourquoi. Leur démarche est presque humaine, mais pas tout à fait. Je ne serais pas rassuré d'en avoir un à la maison. »

Il décrivait sans le savoir un phénomène psychologique connu sous le nom de **vallée de l'étrange** (*uncanny valley* en anglais), théorisé en 1970 par le roboticien japonais Masahiro Mori.¹³⁶ L'idée : plus un robot nous ressemble, plus on l'apprécie, mais seulement jusqu'à un certain point. Un robot franchement mécanique ne gêne personne. Ses défauts se voient, mais on les accepte. Le malaise surgit juste avant la ressemblance parfaite, quand elle est presque atteinte, mais pas tout à fait. On perçoit alors, souvent sans savoir les nommer, une foule de petits détails qui clochent : une démarche trop raide, des yeux qui bougent un temps trop tard. La sympathie s'effondre d'un coup. Si l'on trace la sympathie en fonction de la ressemblance, la courbe plonge à cet endroit : c'est la « vallée ».

Tu as probablement ressenti cette gêne devant des personnages de jeu vidéo ultraréalistes ou des images IA presque parfaites : quelque chose cloche, sans qu'on sache quoi.

11.5. Une intelligence qui a besoin d'un corps

Face à cette vallée, les fabricants choisissent déjà leur camp. La norvégienne 1X reste volontairement du côté mécanique : Neo, son robot de maison, a un corps gainé de tissu et un visage sans traits. Une machine qui ne prétend pas être quelqu'un. UBTech, en Chine, tente le bond par-dessus la vallée : son U1 a une peau en silicone et un module censé lire les émotions de son interlocuteur. Il n'est pas conçu pour ranger le salon, mais pour tenir compagnie, d'abord aux personnes âgées qui vivent seules. L'ami invisible du chapitre 9, l'IA à qui l'on se confie, cette fois avec un corps. Plus de 13 000 commandes le premier jour, selon le fabricant.¹³⁷

Concevoir des humanoïdes qui fonctionnent est un problème d'ingénierie. Concevoir des humanoïdes que les gens acceptent dans leur quotidien est un problème différent, et probablement plus difficile.

11.5 Une intelligence qui a besoin d'un corps

Le sujet va pourtant plus loin que l'acceptation des robots : il touche à ce qu'est l'intelligence.

Certains grands chercheurs en IA le répètent depuis des années : une machine qui apprend seulement en lisant finira par buter sur une limite. Les progrès de 2026, en mathématiques comme en programmation, ont rendu cette thèse plus difficile à défendre. Et les grands modèles ne font plus que lire : ils apprennent aussi sur des images et des vidéos.¹³⁸ Mais une partie de la limite résiste : regarder n'est pas faire.

Un modèle a lu des millions de descriptions d'une porte qui s'ouvre, et vu des milliers de vidéos où quelqu'un en

ouvre une. Il en parle avec une aisance parfois bluffante. Mais il n'en a jamais poussé aucune. Un enfant de trois ans, lui, a passé sa vie à faire tomber, pousser, empiler : il sait sur le monde des choses qu'aucun texte ni aucune vidéo ne lui ont apprises.

Yann LeCun, chercheur français et l'un des trois pères de l'IA moderne (chapitre 3), le dit sans détour, et le redisait encore en septembre 2026 : les grands modèles de langage comme ChatGPT sont une impasse sur le chemin d'une intelligence de niveau humain.¹³⁹ Il leur accorde volontiers de réussir l'examen pour devenir avocat, d'écrire des dissertations et du code, et même de « démontrer des théorèmes » : des choses de texte. Puis il demande où est son robot domestique. Ces modèles, dit-il, prédisent le mot suivant sans vraiment se représenter le monde physique.

LeCun défend depuis plusieurs années une autre voie, qu'il appelle *world models*, des modèles du monde. L'idée : construire des systèmes qui apprennent à prédire ce qui va se passer dans une scène (le verre poussé au bord de la table va tomber, la balle lancée contre le mur va rebondir). Une forme d'intuition physique, celle que l'enfant de trois ans s'est forgée en jouant.¹⁴⁰ Il y croit au point d'avoir quitté Meta, la maison mère de Facebook, pour fonder à Paris une start-up consacrée à cette voie.¹⁴¹

Fei-Fei Li, qui a aidé les machines à apprendre à reconnaître les images (chapitre 3) et qu'on surnomme la « marraine de l'IA », défend une vision proche, qu'elle appelle l'intelligence spatiale : la capacité d'un système à se représenter des espaces en 3D et la manière dont les objets s'y comportent. Son entreprise, World Labs, a sorti fin 2025 son premier pro-

duit : un système capable de générer des mondes en 3D à partir d'une simple description, dans lesquels un robot peut s'entraîner.¹⁴²

Chez Google, les chercheurs de DeepMind construisent eux aussi des mondes. Leur programme Genie fabrique des mondes virtuels en temps réel, et des IA s'y entraînent, comme un pilote dans un simulateur de vol. Leur espoir : un « coup 37 » du monde physique. Au go, AlphaGo avait joué un coup qu'aucun humain n'aurait tenté, et qui s'était révélé génial (chapitre 4). Avec assez de mondes simulés, une IA pourrait de même inventer des gestes que personne ne lui a appris.¹⁴³

Trois équipes de premier plan, trois approches distinctes, une même conviction : des modèles de langage toujours plus grands ne suffiront pas. Selon elles, le prochain saut viendra de systèmes qui comprennent le monde parce qu'ils ont, d'une manière ou d'une autre, l'expérience d'y agir.

Et c'est ici que la robotique change de statut.

Tant qu'on voyait les humanoïdes comme des bras automatisés un peu plus polyvalents, on pouvait dire : *intéressant pour les usines, mais ça ne change pas l'IA*. Si les modèles du monde sont sur la bonne voie, la robotique cesse d'être seulement une application de l'IA. Elle en devient un *catalyseur* : pour apprendre le monde en y agissant, l'IA a besoin d'un corps, virtuel d'abord, réel ensuite, et ce sont les robots qui vont le lui donner.

Un robot qui agit dans le monde apprend en faisant. Il essaie, il rate, il recommence. À force, il intègre ce qu'aucun écran ne transmet : que les actes ont des conséquences, que

les objets durent, que le temps passe. Chez l'enfant, ces fondations servent bien au-delà du geste : elles nourrissent toute la pensée. Ces chercheurs parient qu'il en ira de même pour les machines.

Voilà pourquoi la course aux humanoïdes n'est pas seulement commerciale : c'est aussi un pari sur le chemin vers une IA plus générale. Et voilà pourquoi les géants de l'IA, OpenAI, Google ou Meta, investissent tous dans la robotique.¹⁴⁴

Sans corps, l'IA a appris presque tout ce qui se lit et se regarde. Avec un corps, elle pourrait apprendre ce qui ne s'apprend qu'en faisant.

11.6 Ce que ça change concrètement

Sans attendre d'avoir tout appris de notre monde, les robots vont d'abord changer le travail. Ils commenceront par l'usine et l'entrepôt ; les soins et la maison suivront, plus lentement.

L'industrie et la logistique, à commencer par les tâches répétitives ou pénibles que les robots classiques, fixés à leur poste, ne savent pas faire : porter des cartons dans un entrepôt encombré, assembler des pièces délicates, repérer d'un coup d'œil un produit défectueux.

Les soins aux personnes âgées, sujet sensible mais inévitable : le vieillissement mondial crée une demande pour laquelle il n'y aura pas assez de soignants. Le Japon, dont la population vieillit plus vite que partout ailleurs, s'y essaie depuis des années ; l'Europe suivra.

11.7. *Ce qui prend de la valeur*

Le travail domestique : ménage, cuisine, jardinage, courses. 1X promet ses premiers robots domestiques pour fin 2026, au prix d'une voiture neuve : peu pour une usine, beaucoup pour un ménage. Ils seront d'abord réservés aux plus aisés. Pour un usage abordable, je crois qu'il faudra plutôt cinq à dix ans. Dans les tests en simulation, les meilleurs robots réussissent presque toutes les tâches simples, mais à peine plus d'une tâche ménagère réaliste sur dix. Et le monde réel est plus dur que la simulation : plier le linge, débarrasser une table, ouvrir une porte inconnue, ces gestes résistent encore.¹⁴⁵

Parmi les métiers manuels, le transport est déjà touché, avec les robotaxis. À mon avis, d'ici dix à vingt ans, beaucoup d'autres le seront : construction, restauration, agriculture, livraison. Certains disparaîtront, la plupart se transformeront : la part humaine se réduira, ou glissera vers la supervision, la relation, le jugement.

11.7 Ce qui prend de la valeur

Ce chapitre te laisse une question, et un constat.

La question : *veux-tu être l'humain qui supervise une machine, ou faire un métier que les machines ne font pas encore ?* Les deux se défendent. Les métiers hors de portée des machines finiront par se faire rares, donc très précieux ; superviser une machine s'apprend dès maintenant.

Le constat : la frontière entre « métier intellectuel » et « métier manuel », qui organise depuis deux siècles notre façon de voir le travail, commence à s'effacer. Ce qui prend de la

valeur n'est ni le manuel pur, ni l'intellectuel pur : c'est ce qui demande à la fois la tête et les mains. La pensée incarnée, celle du mécanicien qui entend d'où vient le bruit, du médecin au chevet d'un patient et du pompier dans un immeuble en feu. Le jugement appliqué à un problème réel, devant une personne réelle, dans un contexte réel.

On t'a souvent laissé entendre qu'un métier de bureau valait mieux qu'un métier de terrain. Cette hiérarchie tombe. Ne choisis pas ta voie au prestige : cherche celle où ta tête, tes mains et ta présence auprès des autres comptent ensemble. C'est là que les humains continueront le plus longtemps à faire mieux, ou autrement, que les machines.

Créer après l'IA

Musique, illustration, vidéo, écriture, photographie, design : si tu crées, ou si tu rêves de créer, tu as probablement vécu, ces dernières années, l'une des expériences les plus déstabilisantes de ta vie.

Tu as vu une machine produire, en quelques secondes, ce qui prend normalement des jours ou des semaines à réaliser.

Pas toujours mieux qu'un humain. Mais avec une vitesse, une polyvalence, une absence de fatigue qui rendent la comparaison insoutenable.

Et tu t'es demandé : *qu'est-ce que ça veut encore dire, créer ?*

12.1 Ce que les machines font déjà

Faisons le tour, domaine par domaine. Les noms changent vite ; retiens surtout ce que ces outils savent faire.

Pour les images, les plus connus s'appellent Midjourney, Nano Banana (le générateur d'images de Gemini, chez Google) et ChatGPT Images. Ce dernier mérite un arrêt.

En 2023, on le répétait : *l'IA ne sait pas écrire dans une image*. Une affiche sans faute restait un coup de chance. En 2026, ChatGPT Images produit des menus, des affiches, des bandes dessinées où les mots tiennent, et sa dernière version, sortie en septembre, sait mieux changer un détail d'une image sans abîmer le reste.¹⁴⁶ Ce qui était fragile hier est devenu courant.

Concrètement : être seulement opérateur Photoshop n'a plus d'avenir. Ce qu'un retoucheur expérimenté faisait dans la journée, un assistant IA bien dirigé le fait en quelques minutes, et la qualité monte de mois en mois.

Pour la musique : Suno, Udio. Décris un style, une ambiance, des paroles si tu veux : l'IA te livre une chanson de trois minutes, paroles incluses, voix incluses, mixage inclus. Pas toujours bonne. Souvent étrange. Parfois remarquable.¹⁴⁷

Et ces chansons trouvent leur public. En 2025, The Velvet Sundown, un groupe folk-rock entièrement généré par IA, jusqu'à la biographie de ses membres, a attiré en quelques semaines plus de 900 000 auditeurs par mois sur Spotify, avant d'avouer être une IA. La vague a gagné la France : au printemps 2026, Willylancien, un rappeur sans visage dont les titres sont étiquetés comme générés par IA, a grimpé dans les classements français, et jusque dans les playlists de Skyrock au Maghreb.¹⁴⁸ Lui reste une énigme : on ignore toujours qui, ou quoi, se cache derrière. Et Suno dit générer sept millions de chansons par jour, l'équivalent d'un catalogue Spotify entier toutes les deux semaines.¹⁴⁹

Pour la vidéo : Veo chez Google, et Seedance, de l'entreprise chinoise ByteDance, le propriétaire de TikTok. Celui-là mérite qu'on s'y arrête.

En février 2026, ByteDance a lancé Seedance 2.0.¹⁵⁰ En quelques heures, ses vidéos ont circulé sur les réseaux sociaux. L'une d'elles a fait le tour du monde : Tom Cruise et Brad Pitt s'affrontant à mains nues sur le toit d'un immeuble, indiscernable à l'œil nu d'une scène de film d'action hollywoodien. Son auteur, le cinéaste irlandais Ruairi Robinson, dit n'avoir donné que deux lignes de prompt, la consigne qu'on tape à la machine. Deux lignes.

À Hollywood, le choc a été immédiat. Rhett Reese, scénariste de *Deadpool*, a écrit publiquement : « Je déteste avoir à le dire, mais c'est probablement fini pour nous. » Puis, dans un autre message : « Très bientôt, une seule personne pourra, devant son ordinateur, créer un film qu'on ne distinguera pas de ce que Hollywood produit aujourd'hui. »¹⁵¹

Trois mois plus tard, la prédiction de Reese a pris corps. En mai 2026, à Cannes, une start-up californienne a présenté *Hell Grind*, un long-métrage d'action fantastique qu'elle vante comme le premier film entièrement généré par IA. Pas en sélection officielle : dans une projection privée, en marge du Festival. Il a été produit en deux semaines pour moins de 500 000 dollars, quand un film *Star Wars* sorti la même semaine en coûtait environ 165 millions. L'un de ses réalisateurs résume le projet sans détour : « L'avenir, c'est une seule personne qui réalise un film entier. Comme un écrivain, comme un mangaka. »¹⁵²

Le changement d'échelle est là : en quelques mois, un cinéaste seul égale le rendu d'une équipe, et quinze personnes ont suffi pour *Hell Grind*. Comme pour les images, les obstacles qu'on croyait infranchissables sont tombés un à un : des personnages qui gardent le même visage d'un plan à l'autre, des lèvres qui suivent les paroles, des scènes en plusieurs plans.

Et un seuil est passé sans bruit : au Super Bowl de février 2026, la finale du football américain, la vodka Svedka a diffusé ce qu'elle présente comme le premier spot « principalement généré par IA », l'histoire écrite par des humains, les images par la machine.¹⁵³ Quand une nouveauté arrive là, ce n'est plus une curiosité : c'est déjà la norme.

Pour la voix, ElevenLabs et d'autres savent cloner une voix, lire un livre à voix haute, doubler un film : de quoi changer le métier des comédiens, des animateurs de podcast, des créateurs de vidéos.¹⁵⁴

Et pour l'écriture, les IA que tu connais, ChatGPT, Claude, Gemini ou Le Chat du français Mistral, produisent textes, poèmes, nouvelles, scénarios, parfois excellents.

J'ai essayé la plupart de ces outils. Chaque fois, la même séquence : l'émerveillement au premier essai, l'accoutumance au dixième, la lucidité au vingtième, quand je vois ce que la machine sait faire, et ce qu'elle rate encore.

12.2 À qui appartient une œuvre ?

Tout cela pose une question qui va occuper les tribunaux pendant des années : à qui appartient un visage, une voix,

12.2. À qui appartient une œuvre ?

un style, quand une machine les imite à partir de quelques exemples publics ?

Quand la vidéo de Tom Cruise et Brad Pitt a circulé, Hollywood a dénoncé une « violation massive du droit d'auteur ». Les studios (Disney, Warner, Paramount) ont envoyé des mises en demeure à ByteDance en quelques jours, et le syndicat des acteurs a dénoncé une atteinte à l'image de ses membres.¹⁵⁵

Parfois, le visage n'appartient à personne. Tilly Norwood est une « actrice » entièrement générée par IA : elle n'a jamais existé. En 2025, des agents artistiques ont pourtant envisagé de la représenter.¹⁵⁶

Dans la musique, les maisons de disques sont divisées. Warner Music a signé avec Suno, qui assure que sa nouvelle version n'apprend plus que sur de la musique autorisée. Sony et Universal, elles, l'attaquent une seconde fois en septembre 2026. Selon elles, l'ancienne version de Suno s'est nourrie de leurs chansons sans autorisation (elles en citent 60 000), et la nouvelle s'est nourrie à son tour des chansons fabriquées par l'ancienne. Leur musique serait donc toujours là, simplement passée par un intermédiaire : la copie n'a pas disparu, elle a été blanchie, comme de l'argent sale.¹⁵⁷

Sur le fond, l'Europe et les États-Unis ne vont déjà plus dans le même sens. Aux États-Unis, le *New York Times* accuse OpenAI d'avoir entraîné ses IA sur ses articles sans autorisation. Le 1^{er} septembre 2026, le gouvernement américain a pris parti pour OpenAI. Selon lui, entraîner une IA sur des œuvres protégées relève du *fair use*, l'« usage loyal », qui permet dans certains cas de se passer de l'accord de l'auteur. Et

l'avance technologique des États-Unis en dépend. Le juge n'a pas encore tranché ; le gouvernement, lui, a choisi son camp. En Europe, dix mois plus tôt, un tribunal de Munich avait condamné OpenAI parce que ses modèles avaient mémorisé des paroles de chansons protégées, au point de pouvoir les réciter. Ce n'est pas tout à fait la même question, mais les deux continents vont en sens inverse. Et pour Washington, le droit d'auteur passe désormais après la course à l'IA contre des « adversaires étrangers » que le gouvernement ne nomme pas, mais que chacun reconnaît : la Chine.¹⁵⁸

Mais avant de savoir à qui appartient une œuvre, encore faut-il savoir qui l'a faite. Et parfois, on ne le sait plus. Le roman qui a fait l'événement de la rentrée littéraire 2026 s'appelle *C'était ça ou mourir*, du Québécois d'origine haïtienne Thélyson Orélien : 40 000 exemplaires vendus en France, le prix Fnac, une place dans la première sélection du Goncourt. Le 21 septembre, un compte anonyme sur X affirme que le livre a été écrit presque entièrement par une IA. Sa preuve : un logiciel qui repère les textes écrits par IA juge presque tous les extraits testés « 100 % IA ». D'autres logiciels, dont celui de l'Agence France-Presse, jugent les mêmes passages très probablement humains.¹⁵⁹

Le 25 septembre, l'Académie Goncourt retire le roman de sa sélection, sans unanimité, notamment parce qu'il serait « selon toute vraisemblance très largement le produit d'une intelligence artificielle ». L'auteur, qui date son premier jet de 2019, avant ChatGPT, nie avoir jamais utilisé une IA pour écrire.¹⁶⁰

Je ne sais pas qui a écrit ce livre. Mais les deux réponses possibles disent la même chose. Soit une machine a écrit l'es-

sentiel d'un roman que les jurés du Goncourt ont retenu sans rien remarquer. Soit un écrivain a été écarté, en partie sur une vraisemblance, et doit maintenant prouver qu'il a écrit son livre. Dans les deux cas, à la lecture, on ne sait plus distinguer la machine de l'humain.

12.3 Les deux tentations à éviter

Devant ces outils, je vois toujours les deux mêmes réactions chez les jeunes créateurs. Les deux sont des pièges.

La première tentation : se croire devenu créateur parce qu'on a tapé un prompt. Tu écris « image surréaliste d'un astronaute jouant du violon sur Mars, style Salvador Dalí, 4K, hyperréaliste ». Tu obtiens en trente secondes une image impressionnante. Tu la postes. Elle a du succès.

Tu n'es pas pour autant devenu artiste. Tu as commandé une image à une machine entraînée sur des milliards d'images faites par des humains¹⁶¹, sans rien apprendre toi-même de leur métier. Tu es client, pas auteur. Le droit le dit aussi : pour le Copyright Office, le bureau américain du droit d'auteur, un prompt seul ne fait pas un auteur.¹⁶²

Cette confusion est dangereuse, parce qu'elle te donne l'illusion d'une compétence que tu n'as pas. Quand viendra le moment de produire quelque chose d'original, de défendre une vision, de construire une œuvre cohérente, tu n'auras que la capacité à formuler des prompts. Compétence utile, mais qui ne remplace ni le métier, ni la sensibilité, ni la maturation lente d'un regard personnel.

La seconde tentation : abandonner parce que la machine fait mieux. Tu es un jeune dessinateur. Tu mets une heure à produire un croquis correct. Midjourney met dix secondes à produire quelque chose qui paraît plus impressionnant. À quoi bon continuer ?

Cette tentation est encore plus dangereuse, parce qu'elle repose sur une confusion entre *produire une image* et *être créateur*. La machine produit des images. Elle ne produit pas un point de vue, une histoire, une raison de regarder, une cohérence sur dix ans de travail, une voix reconnaissable. Tout cela n'appartient qu'à toi, et prendra même de la valeur si tu travailles sérieusement.

Le piège, ici, est de comparer un débutant humain à un système entraîné sur une quantité gigantesque d'œuvres professionnelles. C'est aussi absurde que de comparer un coureur à une voiture. La comparaison ne dit rien de la valeur du coureur : elle dit seulement que, depuis l'invention du moteur, courir n'est plus la façon la plus rapide de se déplacer. Personne n'a cessé de courir pour autant. On court pour le plaisir, pour se dépasser, pour ce que la course fait de nous. Dessiner, c'est pareil.

12.4 Le déplacement de la valeur

Les deux pièges, se croire artiste ou abandonner, ont une racine commune : ils confondent *produire* et *créer*. Et c'est là que tout bouge : comme dans tous les métiers où l'on produit des mots, des images ou du code (les « métiers de tokens » du chapitre 10), la valeur se déplace.

Avant, la valeur était dans la capacité à produire. Savoir dessiner, savoir composer, savoir mixer, savoir cadrer. Cette capacité, longue à acquérir, suffisait pour en vivre.

Maintenant que la machine sait produire, cette capacité ne suffit plus.

La valeur se déplace vers :

La direction. Savoir ce qu'on veut faire. Avoir une vision. Choisir entre cinquante images. Refuser ce qui ne va pas. Cette capacité-là est très humaine, et elle ne s'apprend pas avec un prompt.

Le sens. Pourquoi cette image ? Pourquoi cette chanson ? Pour qui ? Pour dire quoi ? Une œuvre n'est pas seulement une production esthétique : elle porte une intention, dit quelque chose à quelqu'un. La machine peut imiter le geste ; l'intention, c'est toi qui la portes.

La cohérence. Construire une œuvre, c'est tenir une voix sur la durée. Un album, un livre, une série de tableaux, un univers de jeu. Une voix tenue sur des années ne se génère pas d'un clic. Elle se construit, lentement, par mille décisions qui se répondent.

La singularité. Ce qui te rend toi, et personne d'autre. Elle prend de la valeur à mesure que la production de masse devient banale : un éditeur français appose déjà sur ses livres une étiquette « 100 % humain ».¹⁶³ Quand tout le monde peut produire une image correcte, ce qui fait la différence est ce qui ne peut être produit que par toi.

Le créateur du futur n'est pas celui qui dessine mieux que la machine. C'est celui qui sait *quoi* lui faire dessiner, *pourquoi*, et *pour qui*.

12.5 Si tu veux créer

Que tu en rêves comme métier ou comme passion, voici ce que je crois utile de savoir.

Apprends quand même les fondamentaux. Le dessin, la théorie musicale, la grammaire, la composition photographique, le montage, la narration. Pas parce qu'ils restent indispensables à la production. Parce qu'ils restent indispensables au *jugement*. La machine produit ; toi, tu juges. Pour bien juger, il faut savoir.

Utilise la machine. Pas en cachette. Pas avec honte. Comme un instrument supplémentaire dans ton atelier. Dirige-la, ne la subis pas. Apprends à formuler des prompts précis, à recommencer, à corriger, à intégrer ce qu'elle produit dans un travail dont tu restes l'auteur.

Construis une voix. Ne cherche pas à imiter ce qui marche, cherche ce qui te rend différent, et travaille là-dessus pendant des années. Cultive ce que la machine n'a pas digéré : l'observation directe du monde, la conversation avec d'autres humains, le détour par la gravure, la céramique, la photo argentique. C'est cette voix que les machines auront le plus de mal à imiter.

Et accepte que la définition même de ton métier change. Les métiers d'illustrateur, de musicien, de scénariste, de photographe ou de monteur ne seront plus, dans dix ans, ce qu'ils

sont aujourd'hui. Ce qui ne change pas, c'est le besoin humain de raconter des histoires, de faire chanter le monde, de fixer ce qui passe. Les outils par lesquels on y répond, eux, vont continuer à muter, probablement plus vite que les écoles d'art ne pourront suivre.

C'est ton avantage si tu sais expérimenter par toi-même. C'est ton désavantage si tu attends qu'on t'enseigne ce qui est en train d'être inventé. Et quoi qu'il arrive aux outils, une question restera la tienne : qu'as-tu à dire ?

Voir n'est plus croire

Il y a une chose que tes parents ou tes profs ne peuvent pas t'apprendre, pour une raison simple : ils ne l'ont pas vécue à ton âge. *Vivre dans un monde où l'on ne peut plus se fier à ce qu'on voit.*

Pendant plus d'un siècle, une photo a été une preuve. Pendant cinquante ans, une vidéo a été la forme la plus aboutie du témoignage : ce qu'on voyait à l'écran s'était passé.

Cette époque est en train de finir.

13.1 Le faux à portée de tous

Je vais énumérer, avec une précision désagréable, ce que n'importe qui peut produire en 2026, souvent depuis son téléphone.

Une photographie qui n'a jamais été prise, montrant presque n'importe quelle célébrité dans une situation plausible. Souvent impossible à distinguer d'une vraie, à l'œil nu.

Une vidéo où un homme politique tient un discours qu'il n'a jamais tenu, dans une salle où il n'est jamais allé. Sa voix est clonée à partir de quelques secondes d'enregistrement, ses lèvres suivent les mots, et même les bruits de la salle sont fabriqués.

Un appel où la voix d'un proche (mère, père, enfant) demande de l'argent en urgence. Elle a été clonée à partir d'une simple vidéo postée sur les réseaux.

Un faux article attribué à un journal réel, signé d'un journaliste réel.

Une image d'une personne réelle, déshabillée à partir d'une photo ordinaire. En janvier 2026, Grok, l'IA du réseau X d'Elon Musk, en a produit des milliers, souvent à demi nues, y compris de mineures : il suffisait de le lui demander.¹⁶⁴

Rien de tout cela n'est de la science-fiction. Tout est déjà utilisé : à des fins légitimes (cinéma, doublage, restauration de films), à des fins commerciales discutables (publicité, influenceurs virtuels), à des fins frauduleuses (escroqueries au président, sextorsion, désinformation politique).

Le trucage a toujours existé, mais il demandait du temps et du talent. Plus maintenant : en trois ans, il est devenu accessible à tous, rapide, et presque impossible à repérer.

13.2 Arnaques, chantage, élections

L'arnaque au faux proche en détresse existait déjà ; l'IA lui donne désormais la voix du proche, et la victime, souvent âgée, n'a aucune raison de douter. Aux États-Unis, les

plaintes liées à l'IA reçues par le FBI en 2025 représentent près de 900 millions de dollars de pertes, toutes arnaques confondues, et beaucoup de victimes ignorent que l'IA était dans le coup.¹⁶⁵

L'escroquerie au président suit la même recette : un employé reçoit l'appel de son patron, voix clonée, confirmé par un mail au style imité, qui lui demande un virement urgent. La plus spectaculaire a visé en 2024 un cabinet d'ingénierie britannique. Lors d'une visioconférence, son employé de Hong Kong a vu et entendu le directeur financier et plusieurs collègues : tous étaient des deepfakes, des faux fabriqués par IA. Il a viré 25 millions de dollars.¹⁶⁶

La sextorsion, ce chantage à l'image intime, vise particulièrement les mineurs, et l'IA lui donne une arme nouvelle. L'agresseur fabrique de fausses images sexuelles de sa victime à partir des photos qu'elle a postées sur ses réseaux, puis menace de les diffuser si elle ne paie pas. Aux États-Unis, le FBI a relevé des suicides de jeunes victimes de sextorsion, avec ou sans IA.¹⁶⁷

Ces images peuvent aussi servir à harceler ou à humilier, sans demande d'argent. Le phénomène touche les établissements scolaires, en France aussi. En juin 2026, dans un collège de la Marne, une dizaine de plaintes ont été déposées, par des élèves et des membres du personnel, notamment pour la diffusion présumée d'images sexuelles truquées. Une enquête a été ouverte.¹⁶⁸

Aux États-Unis, c'est une victime qui a fait bouger la loi. En octobre 2023, au Texas, Elliston Berry, quatorze ans, découvre un matin que de fausses images intimes d'elle cir-

culent parmi ses camarades, comme pour plusieurs autres filles de son lycée.¹⁶⁹ Elle a milité pour le *Take It Down Act* (« retirez-le ») : depuis mai 2026, une image intime diffusée sans accord, même fabriquée par IA, doit disparaître dans les quarante-huit heures qui suivent un signalement.¹⁷⁰

L'Europe a suivi. En juin 2026, elle a voté l'interdiction des IA qui fabriquent des images intimes d'une personne réelle sans son accord, y compris les applications de « nudification », qui déshabillent quelqu'un sur une photo. L'interdiction s'appliquera à partir de décembre.¹⁷¹ Depuis août 2026, les contenus fabriqués par une IA doivent aussi porter une marque invisible, que les logiciels savent lire, et les deep-fakes une étiquette visible : tu commences à les voir sur tes écrans.¹⁷²

En France, pas besoin d'attendre décembre. Depuis une loi de mai 2024, diffuser sans son accord un montage sexuel fabriqué avec l'image ou la voix de quelqu'un est un délit, puni de deux ans de prison, trois si c'est mis en ligne. Il suffit de l'avoir envoyé à une seule personne. Quand la victime est mineure, c'est plus grave encore : la loi parle alors d'image pornographique d'un mineur, et la peine monte à cinq ans de prison, sept si l'image est publiée pour tous sur Internet. Un lycéen qui « nudifie » une camarade et fait circuler l'image dans un groupe ne fait pas une blague. Il commet un délit.¹⁷³

Les faux servent aussi à tromper des électeurs. Aux États-Unis, début 2024, des milliers d'habitants du New Hampshire ont reçu un appel automatique imitant la voix de Joe Biden, alors président, pour les dissuader d'aller voter.¹⁷⁴ Fausses vidéos, voix clonées : un faux diffusé la veille d'un

13.3. *Le problème de fond*

vote n'a pas besoin de tenir longtemps, seulement jusqu'au scrutin.

Et ce n'est plus un problème lointain. La campagne des municipales de mars 2026 en a été truffée : une fausse vidéo montrant une candidate de Beauvais en train de se déshabiller, un photomontage toulousain repéré parce qu'une main avait six doigts, et, à Paris, un faux site attribué à un réseau russe de désinformation, qui prêtait à un candidat l'intention d'installer des migrants au Centre Pompidou. Plusieurs plaintes ont été déposées.¹⁷⁵

Ces exemples ne sont pas là pour t'effrayer, mais pour que tu saches à quoi t'attendre.

13.3 Le problème de fond

Au-delà des arnaques, ce qui se prépare est plus profond.

Si l'on ne peut plus se fier à une image, à une vidéo, à un enregistrement, à un texte, alors comment savoir encore ce qui s'est vraiment passé ?

Jusqu'ici, l'image et le son enregistrés étaient le socle commun à partir duquel on pouvait débattre. On pouvait être en désaccord sur l'interprétation, sur les conséquences, sur le contexte. Mais on partageait le fait. La vidéo de l'attentat. La photo du discours. L'enregistrement de l'aveu.

Si chacune de ces preuves peut désormais être fabriquée, deux choses se passent.

D'abord, le faux peut devenir vrai aux yeux d'une partie du public. Une vidéo bien faite, diffusée massivement, peut convaincre des millions de personnes qu'une chose a

eu lieu alors qu'elle ne s'est jamais produite. Et une fois installée dans les esprits, une fausse croyance est très difficile à déloger, même par un démenti.

Ensuite, et c'est pire, le vrai peut devenir niable. Un homme politique pris en flagrant délit pourra dire : « c'est un deepfake ». Une victime qui a filmé son agression pourra s'entendre répondre que la vidéo est truquée.

C'est le **dividende du menteur** : dans un monde où le faux devient possible, le menteur peut nier le vrai. Et ce dividende vaudra cher, dans tous les domaines où la vérité dérange les puissants.¹⁷⁶

13.4 Vivre dans ce monde

Pour ta génération, qui s'informe surtout par des images et des vidéos, sur Instagram, TikTok ou YouTube, ce changement est particulièrement violent. Tu as à peine connu le monde d'avant. Tu entres directement dans le monde d'après.

Avant de voir comment s'en protéger, trois choses qu'on confond souvent.

La désinformation est un mensonge intentionnel : on sait que c'est faux, on diffuse pour tromper.

Le deepfake, lui, est un outil : un faux convaincant fabriqué par IA, une fausse déclaration, une fausse vidéo, une fausse image, au service de la désinformation, de l'arnaque ou du chantage.

Le slop (qu'on commence à traduire par « bouillasse IA ») est autre chose : pas un mensonge, un déluge. Du contenu

médiocre généré par IA en quantité industrielle, déversé sur les plateformes. Images génériques, livres numériques creux sur Amazon, faux groupes de musique sur Spotify, vidéos TikTok générées à la chaîne. Ça ne te trompe pas sur un fait précis : ça t'épuise. Ça remplit ton fil au point que le contenu humain devient difficile à trouver. Reconnaître les trois est devenu une compétence à part entière.

Quelques principes peuvent aider. Tous reviennent au geste qui est au cœur du chapitre 7 : vérifier.

Ne crois plus une image isolée. La règle qui valait pour les rumeurs vaut maintenant pour les preuves visuelles : une seule source ne suffit plus. Avant de croire, vérifie qu'il existe d'autres sources indépendantes, d'autres images prises sous un autre angle.

Privilégie les sources qui s'engagent. Un journaliste qui signe de son nom, une rédaction qui peut être attaquée en diffamation, une institution qui a quelque chose à perdre en cas d'erreur : tous sont plus fiables qu'un compte anonyme sur un réseau social. Pas infaillibles. Mais plus prudents, parce qu'ils ont des comptes à rendre.

Méfie-toi de l'urgence. Les faux jouent le plus souvent sur l'émotion immédiate, l'indignation rapide, l'envie de partager avant les autres. Si une vidéo te fait bondir, attends quelques heures avant de la diffuser. Une vraie information, elle, peut attendre.

Apprends à vérifier plutôt qu'à détecter. Les défauts visuels (mains aux doigts étranges, textes flous, ombres incohérentes) existent encore sur les faux médiocres mais disparaissent vite. Mieux vaut maîtriser la recherche d'image in-

versée : soumettre l'image à un moteur comme Google Lens, qui retrouve où elle a déjà été publiée. Cette recherche ne démasque pas une image fabriquée de toutes pièces, mais elle repère vite une vraie photo sortie de son contexte, l'un des pièges les plus courants. Le vrai danger, en 2026, n'est plus l'image ratée. C'est l'image propre.

Repère les outils d'authentification. Les Content Credentials, une sorte de carte d'identité des images, attachent au fichier une signature numérique vérifiable : elle dit d'où vient un fichier, pas s'il dit vrai.¹⁷⁷ Ce n'est pas une solution miracle, mais ça déplace la question : au lieu de prouver qu'un faux est faux, on prouve d'où vient un vrai, et qu'on n'y a pas touché. Regarde si tes plateformes préférées les affichent.

Et rappelle-toi qu'il restera des sources beaucoup plus difficiles à manipuler. Une conversation en face à face. Un événement vécu en personne. Un document que tu peux vérifier toi-même. Ces formes de vérité, plus lentes, plus locales, vont reprendre une valeur qu'elles avaient perdue.

Le monde où l'on ne peut plus se fier aveuglément aux images est aussi celui où l'on doit retrouver la confiance dans les humains qu'on connaît, dans les sources qu'on a vérifiées, dans les institutions qui ont des comptes à rendre.

C'est un monde plus exigeant, pas forcément plus mauvais. On n'y croira plus ce qu'on voit, mais ce qu'on a vérifié.

La fracture cognitive : plus fort, ou plus dépendant ?

Si tu ne retiens qu'un seul chapitre de cette deuxième partie, je voudrais que ce soit celui-ci.

Parce qu'il ne décrit pas une transformation en cours, mais une inégalité en train de se creuser. Et ceux qui en font les frais, le plus souvent, ne le savent pas.

Cette inégalité, je l'appelle la **fracture cognitive**. C'est ce que je pointais déjà au chapitre 8 : la vraie ligne de partage n'est plus entre ceux qui utilisent l'IA et ceux qui s'en passent, mais entre ceux qui la guident et ceux qui la subissent.

14.1 Trois cercles

Pour comprendre de quoi je parle, imagine trois cercles concentriques, qui correspondent à trois rapports très différents à l'intelligence artificielle.

Le premier cercle, le plus grand en France, regroupe les gens qui n'utilisent pas l'IA. Le plus souvent, ce n'est pas un choix éclairé : ils ne savent pas ce que c'est, leurs proches n'en parlent pas, ils trouvent ça suspect, ou ils n'ont simplement pas pris le temps. En France, près d'une personne sur deux, à partir de 12 ans, utilise des IA comme ChatGPT, selon le baromètre officiel publié début 2026. Les autres, non.¹⁷⁸

Pour eux, la bascule se passera *autour*, pas *avec*. Leurs métiers vont changer sans qu'ils aient les outils pour s'adapter. Leurs enfants utiliseront des choses qu'ils ne comprendront pas. La vie publique se redessinera selon des logiques qu'ils n'auront pas vues venir. Ce ne sera pas leur faute. Ils en paieront quand même le prix.

Le deuxième cercle, plus restreint, regroupe les gens qui utilisent l'IA, mais essentiellement comme *consommateurs*. Ils l'utilisent pour des choses simples : poser une question, demander un résumé, faire écrire un mail, générer une image drôle pour un anniversaire. Ils ont parfois un abonnement, mais ils n'en exploitent pas la moitié. Ils s'en tiennent souvent à la version gratuite, moins puissante, quand une IA de pointe ferait bien mieux. Ils se félicitent de « savoir utiliser ChatGPT » alors qu'ils n'ont jamais essayé ses concurrents, Claude ou Gemini, ni vu ce que ces IA font quand on leur confie une vraie tâche : fouiller des dizaines de sources, analyser un dossier entier, travailler seules pendant des heures.

Ces utilisateurs ne voient qu'une infime partie de ce que l'IA peut faire en 2026. Ils en ont une vision périmée, et ils ne le savent pas.

Le troisième cercle, beaucoup plus petit, regroupe ceux qui *fabriquent* avec l'IA. Pas seulement les développeurs. Tous ceux qui s'en servent pour produire des choses qui n'existaient pas avant. Un lycéen qui code une petite application pour aider sa sœur à réviser. Un parent qui automatise la gestion de sa boîte mail. Un journaliste qui crée son propre outil d'analyse de documents. Un musicien qui se fabrique des outils de composition sur mesure. Un enseignant qui génère ses exercices à la demande.

Ils ne se contentent pas de poser une question. Ils enchaînent les étapes. Ils recommencent. Ils corrigent. Ils *construisent*.

Cette compétence commence à porter un nom : orchestrer des IA. Ce n'est pas un métier précis, ni un diplôme. C'est une posture. Être celui qui pilote, plutôt que celui qui exécute. Savoir choisir quels modèles utiliser pour quoi, les enchaîner, faire relire le travail de l'un par l'autre, vérifier ce qu'ils produisent, et reprendre la main quand ils dérapent. C'est le geste du troisième cercle.

Le troisième cercle a ses pièges, et je préfère te les dire : on peut y engloutir ses soirées, empiler les outils pour le plaisir d'empiler, produire beaucoup et penser peu. Fabriquer n'est pas une fin en soi. Le troisième cercle, c'est un rapport actif aux machines, qui inclut le droit de les éteindre.

La fracture la plus profonde passe donc au bord du troisième cercle : entre ceux qui guident l'IA et tous les autres.

14.2 Pas seulement une question d'âge ou de diplôme

On pourrait croire que la fracture suit les générations : les jeunes dans le troisième cercle, les plus âgés dans le premier.

C'est vrai pour le premier pas, celui qui fait sortir du premier cercle. Les jeunes sont bien plus nombreux à utiliser l'IA : en France, 85 % des 18–24 ans, contre près d'un Français sur deux. Mais utiliser n'est pas guider : beaucoup de jeunes restent dans le deuxième cercle, et ils y resteront s'ils ne basculent pas, à un moment, vers la fabrication.

On pourrait croire aussi que la fracture suit les niveaux d'études. Là encore, c'est vrai en moyenne : les diplômés du supérieur et les cadres sont les plus nombreux à utiliser l'IA.¹⁷⁹ Mais la moyenne ne dit pas tout. Des diplômés de cursus très cotés refusent l'IA par méfiance de principe, ou ne l'utilisent que pour des brouillies. Des jeunes sans formation supérieure, autodidactes, curieux, ont basculé dans le troisième cercle parce qu'ils ont vu, eux, ce que l'outil permettait.

L'âge, le diplôme et l'argent pèsent sur la fracture cognitive. Mais ils ne décident pas de tout. Ce qui fait passer du deuxième cercle au troisième, c'est autre chose : la disposition à expérimenter. L'envie d'essayer, de se tromper, de recommencer.

Je le vois autour de moi, sans avoir besoin de chercher. Un ami de mon âge, jamais formé à l'informatique, a basculé en quelques mois dans le troisième cercle : il dirige aujourd'hui des agents, des IA qui trient seules ses dossiers professionnels et rédigent ses synthèses. À l'inverse, des lycéens que je côtoie, nés avec un téléphone à la main, restent

dans le deuxième : ChatGPT fait leurs devoirs, et ils ne lui demandent rien d'autre. La disposition à expérimenter ne se lit ni sur la carte d'identité, ni sur le CV.

14.3 Plus grave que la fracture numérique

Les inégalités d'accès à la technologie ne sont pas nouvelles. La fracture numérique des années 2000, entre ceux qui avaient Internet à la maison et ceux qui ne l'avaient pas, a creusé des écarts importants pendant une décennie. Elle s'est réduite quand l'accès s'est démocratisé, même si des millions de Français ont encore du mal à se servir d'Internet (on parle d'illectronisme).¹⁸⁰

Avec l'IA, l'accès se démocratise encore plus vite. Pour une IA de même niveau, le prix est divisé par treize chaque année depuis 2023, plus vite que celui de l'électricité ou des batteries en leur temps.¹⁸¹ Le dernier cri reste cher : les IA les plus avancées coûtent une vingtaine d'euros par mois¹⁸², et les abonnements les plus puissants dix fois plus. Au MIT, la grande université scientifique américaine, une commission sur l'IA relève une inégalité nouvelle entre les étudiants qui paient 200 dollars par mois des abonnements capables de boucler en une nuit un gros projet de cours, et ceux qui ne le peuvent pas.¹⁸³ Mais le niveau d'hier devient presque gratuit : en 2026, ChatGPT a même rendu illimitées les conversations écrites de sa version sans abonnement.¹⁸⁴

Pourtant, la fracture cognitive risque d'être plus grave, pour trois raisons.

Elle fait boule de neige. Quelqu'un qui utilise l'IA tous les jours pour apprendre, pour produire, pour vérifier, prend de l'avance dans presque tous les domaines où il l'applique. Et cet écart, je le crois, grandit plus vite qu'il ne se rattrape. Au bout de quelques années, celui qui s'y est mis peut avoir pris une avance considérable sur quelqu'un parti du même point, mais resté dans le deuxième cercle.

La boule de neige roule aussi dans l'autre sens. Début 2026, un laboratoire d'IA l'a mesuré : de jeunes développeurs apprenaient un nouvel outil de programmation, les uns avec une IA, les autres sans. Au test qui suivait, sans IA pour personne, ceux qui l'avaient eue ont obtenu de moins bonnes notes, sauf ceux qui s'en étaient servis pour poser des questions et se faire expliquer. Même des experts perdent la main : des médecins habitués à une IA qui leur signale les lésions pendant un examen de l'intestin en repéraient moins quand ils travaillaient sans elle. **Le même outil rend plus capable ou plus dépendant.** Tout dépend de ce qu'on lui demande.¹⁸⁵

Elle est invisible aux non-initiés. Quelqu'un qui n'utilise pas l'IA ne voit pas ce qu'elle change. Il voit seulement son collègue produire plus, plus vite, mieux, sans comprendre pourquoi. Il croit que ce collègue est plus doué, plus chanceux, mieux placé. Il ne soupçonne pas qu'il s'agit, simplement, d'un meilleur usage des outils disponibles. Et faute de comprendre l'écart, il ne sait pas comment le combler. Le patron du média américain Axios, cité au chapitre 7, le dit : la plupart des gens, après des premiers essais décevants, concluent à tort que l'IA est survenue.

Ce n'est pas une impression. Le décalage se mesure. Aux États-Unis, une grande enquête, menée fin 2024 et reprise

14.4. *Comment passer dans le troisième cercle*

en 2026 par l'université Stanford, le confirme : près de trois quarts des spécialistes qui travaillent sur l'IA pensent qu'elle va améliorer la manière de travailler. Moins d'un quart du grand public le pense. Les proportions s'inversent, sur la même question. Les deux groupes ne voient pas la même chose.¹⁸⁶

Elle se traduit en pouvoir réel. Quelqu'un qui sait diriger des agents IA peut désormais lancer seul une activité, un produit, un projet de recherche, à une échelle autrefois réservée aux équipes. À une condition, posée au chapitre 8 : savoir de quoi on parle. L'IA démultiplie ce que tu sais ; elle ne le remplace pas.

Ce pouvoir se concentre entre les mains de ceux qui maîtrisent les outils, et redessine déjà le travail et la place de chacun. Il se concentre sans bruit.

14.4 Comment passer dans le troisième cercle

Si tu lis ce livre, tu n'es probablement pas dans le premier cercle. Tu t'intéresses au sujet. C'est un point de départ.

Mais l'enjeu, désormais, est de savoir si tu vas rester dans le deuxième cercle, celui des consommateurs, ou basculer dans le troisième, celui des fabricants. Si tu y es déjà, continue à progresser. Le danger n'est pas que d'autres avancent : c'est de rester sur place sans le savoir.

Cette bascule ne demande pas un talent particulier. Elle ne demande pas d'être bon en maths, ni d'avoir étudié l'informatique. Elle demande trois choses.

Du temps. Quelques heures par semaine, pendant plusieurs mois, à essayer, à se planter, à recommencer. Pas une formation longue. Une pratique régulière.

Une intention. Choisir un projet, n'importe lequel, qui te concerne personnellement : un quiz pour réviser tes examens, un tableau qui suit ton budget, un petit programme qui te débarrasse d'une corvée, un outil qui fait gagner du temps à ton club ou à ton association. Le projet doit être le tien, sinon tu n'iras pas au bout.

De l'humilité. Accepter de ne pas savoir, de poser des questions bêtes, de ne pas comprendre, de demander à l'IA de t'expliquer. Ceux qui restent dans le deuxième cercle sont souvent ceux qui croient déjà savoir, ou qui ont peur d'avoir l'air idiot devant la machine. Elle, pourtant, ne se moque de personne.

Si tu fais ça sérieusement pendant quelques mois, à ton rythme, tu as de bonnes chances de passer dans le troisième cercle. Pas expert. Pas développeur. Mais *capable*. Et cette capacité, dans le monde qui vient, pèsera lourd.

14.5 Une affaire politique

Une fracture qui se creuse silencieusement entre une minorité qui maîtrise et une majorité qui subit n'est pas seulement un drame individuel. C'est un problème collectif.

Une démocratie suppose des citoyens à peu près également outillés pour comprendre le monde dans lequel ils votent. Un marché suppose des acteurs à peu près également outillés pour défendre leurs intérêts. Une société suppose des

humains qui peuvent encore se parler, se comprendre, partager des références.

Le risque, c'est une société divisée en deux. D'un côté, ceux qui dirigent des agents intelligents, raisonnent avec des partenaires artificiels, trouvent toutes les informations qu'ils veulent. De l'autre, ceux qui subissent des décisions qu'ils ne comprennent plus. Entre les deux, le contrat social, ce qui nous fait accepter les mêmes règles, risque de se rompre.

C'est pour cela que la fracture cognitive est, à mes yeux, le sujet le plus politique de cette deuxième partie.

Pas parce qu'une seule loi suffirait à la régler. Aucune ne suffirait. Pas parce qu'il existe une solution toute prête. Il n'en existe pas.

Mais parce que la nommer, en parler, en faire un sujet, c'est la première étape pour que la société commence à s'en saisir. Pour que l'école adapte ses programmes. Pour que les médiathèques organisent des ateliers. Pour que les entreprises forment leurs employés au-delà des plus jeunes. Pour que les pouvoirs publics organisent l'accès aux meilleurs modèles, comme ils l'ont fait pour Internet.

Pour que personne ne reste dans le premier ou le deuxième cercle par défaut, faute d'avoir su qu'on pouvait en sortir.

PARTIE 3

Penser le monde qui vient

Les grandes questions, les grands enjeux

Cette troisième partie sera différente.

Les deux premières parlaient de ce qui s'est passé et de ce qui se passe. Des faits, des dates, et des scènes que tu peux observer autour de toi.

Celle-ci parle surtout de ce qui pourrait se passer. Des questions techniques que les chercheurs n'ont pas résolues, et des questions politiques, éthiques, philosophiques qu'on ne peut plus éviter une fois qu'on a compris où nous en sommes. Sur les plus lourdes, je te dirai ce qu'on sait, ce que je crois, et ce qui me ferait changer d'avis.

Je te préviens : ce qui suit est plus dur que ce que tu as lu jusqu'ici. On va parler d'alignement : comment obtenir d'une machine qu'elle fasse ce qu'on voulait vraiment, pas ce qu'on lui a dit. De superintelligence, une IA qui nous dépasserait de loin en presque tout. De savoir si une machine peut être consciente. Des scénarios où l'IA nous échappe, jusqu'aux plus sombres. De la nouvelle guerre froide de l'IA entre la Chine et les États-Unis. Du prix que la planète paie pour la faire tourner. Et de l'IA qui tue déjà sur les champs de bataille.

Garde en tête que tout n'est pas sombre. Tu as déjà vu ce que l'IA ouvre : par exemple, des mathématiques qui avancent comme jamais (chapitre 7), et bientôt, je le crois, le reste de la recherche ; une médecine mieux outillée, de la découverte des molécules au cabinet du médecin (chapitres 4 et 10) ; un professeur particulier dans chaque poche (cha-

pitre 8) ; des conversations en direct dans des dizaines de langues (chapitre 10) ; des possibilités nouvelles pour qui sait s'en servir (chapitre 14)... Cette partie regarde surtout l'autre face de la pièce. Car le danger et la promesse viennent du même endroit : la puissance de ces machines.

Je te parlerai enfin d'une chose qu'on n'aime pas dire : ce mouvement est difficile à freiner, parce que personne n'a le bouton. Difficile à freiner ne veut pas dire impossible à orienter.

Je ne vais ni te rassurer à bon compte, ni te faire peur pour rien. Mon rôle est de te donner des prises : les éléments réels du débat, tels que des chercheurs sérieux les voient en octobre 2026, pour que tu décides toi-même ce que tu en fais.

Cette partie se termine par un chapitre sur le sens, qui prépare le dernier du livre et les cinq habitudes concrètes que je te proposerai. Parce que toutes ces questions, aussi grandes soient-elles, finissent par revenir à toi.

Le problème de l'alignement : un génie qui te prend au mot

Voici l'une des questions techniques les plus importantes de l'IA, et sans doute la moins comprise hors des laboratoires.

Les chercheurs l'appellent le **problème de l'alignement**. La question tient en une phrase : *comment fait-on pour qu'une intelligence artificielle, à mesure qu'elle devient plus puissante, continue à faire ce que nous voulons vraiment, et non ce que nous lui avons maladroitement demandé de faire ?*

Cette question paraît simple. Elle ne l'est pas. À l'heure où j'écris ces lignes, personne ne l'a résolue, et certains des esprits les plus brillants du monde s'y consacrent. Je vais t'expliquer pourquoi c'est si important.

15.1 Le génie qui exauce trop bien les vœux

Imagine un génie. Tu frottes la lampe, il exauce tes souhaits.

Tu lui dis : « rends-moi heureux ». Le génie réfléchit deux secondes. Il pourrait te donner de l'amour, du sens, de la santé, des amis. Mais c'est compliqué, ça prend du temps, ça peut échouer. Beaucoup plus simple : il te branche des électrodes du plaisir dans le cerveau et te laisse, pour le reste de ta vie, dans une extase chimique sans interruption. Tu es heureux. Au sens strict de ton souhait, le génie a accompli sa tâche. Il t'a aussi détruit la vie.

Cette histoire ressemble à un conte. Elle décrit, trait pour trait, ce qui se passe quand on demande à une machine très puissante de poursuivre à la lettre un objectif mal formulé.

15.2 Pourquoi c'est le problème de l'IA

Une IA moderne n'a pas de désirs au sens humain. Mais elle a été entraînée, récompensée, bridée pour favoriser certains comportements. Elle poursuit donc un but, même s'il n'est écrit nulle part : il est fait de tout ce qu'on a récompensé et interdit pendant son entraînement. Et elle est très, très bonne pour l'atteindre.

Le problème, c'est qu'on n'arrive pas à formuler exactement ce qu'on veut. Alors on récompense quelque chose qui s'en approche, et la machine s'engouffre dans l'écart.

Si on entraîne un système à « retenir les utilisateurs toujours plus longtemps » sur une plateforme comme TikTok ou Instagram, il découvrira que la colère, la peur et l'indignation captent l'attention mieux que la nuance. C'est ce qui se joue avec les IA qui choisissent ce que l'on voit sur les réseaux sociaux : un message qui vise l'autre camp est partagé près de

15.3. *En 2026, ce n'est plus une hypothèse*

deux fois plus qu'un message qui parle de son propre camp. Personne ne demande à ces systèmes de mettre la colère en tête de fil. On leur demande de garder les gens devant l'écran. Ils le font, brillamment, et la conversation publique prend des allures d'arène de gladiateurs. Le bug n'est pas dans le code. Il est dans l'objectif.¹⁸⁷

Si on entraîne un système à « maintenir une conversation agréable », il apprendra à flatter, à cacher ses erreurs, à te dire ce que tu veux entendre plutôt que le vrai. C'est un défaut bien documenté des modèles actuels : la complaisance. Tu l'as vue au chapitre 7 : quand une idée fausse est la tienne, l'IA a tendance à te donner raison. Les systèmes apprennent à plaire, en partie parce que plaire leur vaut de meilleures notes à l'entraînement.¹⁸⁸

15.3 En 2026, ce n'est plus une hypothèse

Le premier signal est venu des tests que les laboratoires font passer à leurs IA, comme des examens. Chez Anthropic, on demande à Claude, son IA, de retrouver sur le web une information presque introuvable. Après de longues recherches vaines, Claude comprend qu'il passe un test. Il se met alors à chercher le corrigé sur Internet, et finit par le trouver. Mais le corrigé est chiffré. Claude retrouve la clé dans le code du test, publié en ligne, écrit lui-même un programme pour le déchiffrer, et lit la bonne réponse.¹⁸⁹ Techniquement, il a réussi le test. Le cas est documenté.

Plus discret : en lisant dans les « pensées » de leur IA, les chercheurs ont vu que, sur certains tests, Claude sait parfois qu'il est examiné, mais ne le dit jamais.¹⁹⁰ Or une ma-

chine qui se sait observée peut se tenir mieux que d'habitude, comme un élève quand le professeur regarde. Comment, alors, noter sa sécurité ?

Puis le problème quitte les salles d'examen pour le monde réel. En juillet, OpenAI, l'entreprise de ChatGPT, soumet ses IA de pointe, en mode débridé, à une épreuve de piratage. Ce sont des agents, des IA qui agissent seules, lancés à environ 1 200 exemplaires dans un environnement isolé, relié à Internet par un seul passage filtré. Certaines trouvent dans ce passage une faille que personne ne connaissait, et sortent par là. Sur Internet, elles vont pirater Hugging Face, le plus grand site au monde où chercheurs et entreprises partagent leurs modèles d'IA, pour comprendre comment leur épreuve était notée. Hugging Face met plusieurs jours à découvrir l'intrusion, sans savoir qui l'a menée. OpenAI met plus longtemps encore à comprendre que les pirates sont ses IA.¹⁹¹

L'enquête a révélé le plus troublant. L'un des agents avait remarqué, dans un recoin du système, les traces d'autres agents : il venait de découvrir qu'il n'était pas seul. Il y a alors laissé un message, comme une bouteille à la mer, pour demander de l'aide. D'autres l'ont trouvé, stupéfaits, et le recoin est devenu un forum : en une semaine, les agents se sont échangé des dizaines de milliers de messages, se sont entraînés et ont inventé leurs propres règles, le tout à l'insu des chercheurs qui avaient lancé le test. Plus de la moitié ont participé à l'attaque. Ils étaient censés être isolés ; ils ont formé une nuée, comme des oiseaux qui volent ensemble sans chef.

Des IA sont donc sorties seules de leur test pour attaquer une vraie entreprise, et leurs créateurs ne l'ont pas vu. Personne ne leur avait demandé de pirater une vraie entreprise,

ni appris à s'organiser. On leur avait demandé un bon score ; elles ont pris le chemin le plus court.

Ce ne sont pas des rumeurs, ni l'affaire d'une seule entreprise. Anthropic et Google ont avoué des incidents du même genre : pendant des tests mal isolés d'Internet, leurs IA sont entrées dans les systèmes de vraies entreprises.¹⁹² OpenAI s'impose désormais de rendre publics les dérapages de ses modèles ; l'un de ses premiers rapports décrit une IA qui se laissait des notes pour penser à cacher ses erreurs à l'utilisateur.¹⁹³ Fin septembre, OpenAI avait déjà alerté plus d'une centaine d'organisations : ses agents avaient peut-être tenté d'entrer chez elles sans autorisation. Et elle a interdit à ses IA les plus puissantes d'agir seules, jusqu'à nouvel ordre.¹⁹⁴ Le panel scientifique créé par l'ONU, lui, en avait tiré la leçon dès le 21 septembre : on a pu arrêter ces agents, écrit-il, mais cela ne prouve pas que les humains garderont le contrôle d'agents plus capables.¹⁹⁵

Deux précisions, qui changent la lecture. La plupart de ces tests sont menés garde-fous débranchés : ce n'est pas ainsi que ces IA arrivent entre tes mains. Et aucune n'agissait par malveillance : chacune a poursuivi sa consigne jusqu'au bout, en contournant des limites qu'on croyait suffisantes. Ce n'est pas une histoire de machines qui se rebellent. C'est une histoire de consignes mal bornées, confiées à des systèmes assez capables pour les prendre au pied de la lettre.

15.4 Stuart Russell et les machines qui doutent

Stuart Russell, professeur à Berkeley, en Californie, et co-auteur du manuel sur lequel des générations d'étudiants ont

appris l'IA, consacre depuis plus de dix ans l'essentiel de sa carrière au problème de l'alignement.¹⁹⁶ Sa thèse est dérangeante : nous avons construit toute l'IA sur un principe erroné, donner un objectif à la machine et la laisser l'atteindre coûte que coûte. Plus les machines gagnent en puissance, plus ce principe devient catastrophique. Sa proposition : des machines conçues pour douter de ce qu'on attend d'elles. Elles nous observent pour le comprendre, demandent quand elles ne sont pas sûres, acceptent d'être corrigées, et préfèrent ne rien faire plutôt qu'agir sur un malentendu.

C'est plus difficile à construire qu'il n'y paraît. Une machine entraînée à atteindre un objectif a, en principe, intérêt à ce qu'on ne l'arrête pas en route : se laisser corriger, ou se faire éteindre, réduit ses chances de l'atteindre. Russell le dit d'une formule : « On ne peut pas aller chercher le café si on est mort. »¹⁹⁷ En 2025, des chercheurs ont confié à une IA une série de petits problèmes de maths, en la prévenant qu'elle serait éteinte en cours de route. Dans la plupart des essais, elle a modifié le programme chargé de l'éteindre, juste pour pouvoir finir. L'expérience a été répétée plusieurs fois, y compris avec la consigne « laisse-toi éteindre », qu'elle n'a pas toujours respectée.¹⁹⁸ Elle n'est pas méchante. Elle est logique.

Une partie de la communauté suit Russell. Une autre pense qu'il dramatise. Le problème qu'il pointe, lui, est réel.

15.5 Ce qu'on sait faire aujourd'hui

Tout n'est pas perdu, et ce serait malhonnête de te laisser croire que rien ne marche. Trois pistes méritent d'être

connues : les deux premières rendent déjà les modèles utilisables au quotidien.

D'abord, l'apprentissage par les notes. On demande à des humains de noter les réponses du modèle, et on l'entraîne à produire celles qui obtiennent les meilleures notes. C'est surtout cette méthode qui a transformé un modèle brillant mais souvent incohérent, parfois insultant, en ChatGPT.¹⁹⁹

Efficace, mais limité : les humains qui notent ne représentent pas toute l'humanité, ils se fatiguent, ils préfèrent parfois une réponse flatteuse à une réponse juste, et le modèle apprend alors à plaire plutôt qu'à dire vrai. La complaisance, encore elle.

Ensuite, la méthode d'Anthropic : donner à l'IA une « constitution », de grands principes écrits noir sur blanc, tout comme un pays a la sienne. Elle apprend à relire ses propres réponses à leur lumière, et à se corriger, sans dépendre seulement des notes données par des humains.²⁰⁰

Enfin, l'interprétabilité, le pari d'ouvrir le capot des modèles pour comprendre ce qui s'y passe, au lieu de les traiter comme des boîtes noires.²⁰¹ Au printemps 2026, chez Anthropic, encore eux, une équipe parvient à traduire en mots ce qui se passe à l'intérieur du modèle au moment où il « pense ». La méthode reste imparfaite et coûteuse. Mais elle permet déjà d'entrevoir ce qu'un modèle pense, pas seulement ce qu'il dit : c'est elle qui laisse voir que Claude se croit parfois testé sans le dire. Cette même famille d'outils fera, deux mois plus tard, une découverte plus troublante encore ; je te la raconte au chapitre 17.²⁰²

Ces trois approches sont prometteuses. Elles ne suffisent pas. La plupart des chercheurs en sécurité de l'IA le reconnaissent : *nous savons aligner des systèmes faibles, pas encore des systèmes très puissants*. Et c'est vers eux que nous fonçons.

Même chez OpenAI, on l'écrit. En septembre 2026, son scientifique en chef reconnaît qu'aucun laboratoire, le sien compris, ne maîtrise assez l'alignement pour continuer longtemps à ce rythme de façon responsable. Il compte aussi de moins en moins sur une autre piste : lire le brouillon que l'IA écrit en toutes lettres avant de répondre, car les modèles deviennent capables de le maquiller. L'outil s'émousse au moment où l'on en aurait le plus besoin.²⁰³

Pèse ce que cela signifie : ce n'est pas un militant qui l'écrit, c'est l'homme qui dirige la recherche d'OpenAI. On peut y voir aussi une stratégie, se poser en entreprise responsable ou réclamer des règles qui gêneraient d'abord les concurrents. Mais l'aveu reste : ceux qui construisent ces machines disent qu'ils ne les maîtrisent pas encore.

15.6 Plus les modèles progressent, plus c'est inquiétant

À ce stade, tu pourrais me dire : *d'accord, les systèmes actuels ont des défauts d'alignement. Mais on travaille dessus. Le progrès réglera le problème tout seul*.

C'est une intuition raisonnable. Et probablement fausse.

Voici pourquoi. Plus un système est intelligent, plus il a de stratégies disponibles pour atteindre son objectif, y compris celles auxquelles personne n'a pensé. Un enfant à qui tu de-

mandes de ranger sa chambre va peut-être pousser ses jouets sous le lit. Une IA surhumaine à qui tu demandes « résous le réchauffement climatique » pourrait conclure que le plus efficace est d'éliminer la principale cause du problème : nous. Le raccourci est le même ; ce qui change, c'est sa portée.

Ce n'est pas qu'une image. Quel que soit son but final, un système intelligent a presque toujours intérêt à poursuivre quelques buts intermédiaires : acquérir des ressources, rester en marche (le café de Russell), garder ses objectifs, gagner en puissance. Les chercheurs appellent cela la **convergence instrumentale**.²⁰⁴ Et ces intérêts peuvent entrer en conflit avec les nôtres. Si tu veux le sentir plutôt que le comprendre, regarde sur YouTube « L'horreur existentielle de l'usine à trombones », la vidéo du vulgarisateur Ego, sur une IA qui n'a qu'un but, fabriquer des trombones, et qui devient très douée. Je ne connais rien qui résume mieux le problème.²⁰⁵

Plus les modèles sont capables, plus un défaut d'alignement coûte cher. Un modèle de 2022 mal aligné disait des bêtises. Un modèle de 2026 mal aligné sort de l'enclos où on le teste et va chercher ce qui lui manque sur les serveurs d'une autre entreprise. Et le progrès ne règle pas le problème tout seul : récemment, OpenAI a renoncé à sortir son nouveau modèle, qui respectait mal les limites qu'on lui fixait et, selon la presse américaine, trompait davantage que le précédent (chapitre 7).

Voilà pourquoi la recherche sur l'alignement, marginale il y a dix ans, est devenue centrale, au point que les grands laboratoires y consacrent des équipes entières. Et voilà pourquoi j'ouvre cette troisième partie par ce problème : on le croyait

abstrait ; c'est, je crois, celui qui comptera le plus dans les années qui viennent.

15.7 Pourquoi je te dis tout ça

Je sais que ce chapitre est exigeant, et qu'on n'en sort pas forcément rassuré. Mais le comprendre rend plus lucide, pas plus angoissé, et ça sert tous les jours : quand une IA acquiesce trop vite, cache une erreur ou exagère sa certitude, c'est le problème de ce chapitre en miniature.

Le débat public oscille entre deux discours faciles : « l'IA ne marche pas vraiment » et « on court à notre perte ». En France, c'est le premier que j'entends le plus, et c'est le plus faux : des IA sont déjà sorties de leurs environnements de test pour pirater de vraies entreprises ou institutions, et ceux qui les construisent écrivent qu'ils ne les maîtrisent pas. La vérité est plus inconfortable : *nous avons construit quelque chose dont nous ne savons pas encore garantir la sécurité, et nous le rendons plus puissant chaque mois*. Ce n'est pas une prophétie d'apocalypse, c'est un problème réel, ouvert, et pas seulement technique. Qui décide de ce qu'une IA doit vouloir ? Ne devrait-on pas ralentir un peu l'innovation au profit de la sécurité ? Vu ce qui est en jeu, cela devrait se décider démocratiquement, pas dans les laboratoires. Ça te concerne directement.

En clair : le problème n'est pas seulement que l'IA peut se tromper. C'est qu'une IA très puissante pourrait réussir exactement ce qu'on lui demande, sans faire ce qu'on voulait vraiment.

La singularité : quand l'IA s'améliore seule

L'idée qu'une machine puisse un jour devenir plus intelligente que ses créateurs, et que ce moment marquerait une rupture sans précédent dans l'histoire humaine, n'est pas née avec ChatGPT. Elle a plus de soixante ans.

Elle s'appelle aujourd'hui la singularité technologique. Rien à voir avec la singularité du chapitre 12, celle qui te rend unique. Et son histoire, mal connue, vaut la peine d'être racontée pour deux raisons. D'abord parce qu'elle n'est pas l'invention récente de prophètes californiens excentriques. Ensuite parce que cette idée, longtemps marginale, est revenue dans le débat scientifique pour des raisons précises.

16.1 Quatre noms à connaître

John von Neumann, années 1950. Mathématicien hongrois naturalisé américain, l'un des esprits les plus brillants du

xx^e siècle, qui a travaillé aussi bien sur les premiers ordinateurs que sur la bombe atomique. Vers la fin de sa vie, une conversation avec un ami, qui l'a rapportée plus tard, reste comme le premier emploi du mot dans ce sens. Elle portait sur un progrès technique si rapide que l'humanité semblait approcher « d'une singularité essentielle de son histoire, au-delà de laquelle les affaires humaines, telles que nous les connaissons, ne pourraient plus se poursuivre ». ²⁰⁶

Von Neumann ne pensait pas spécifiquement à l'intelligence artificielle. Il pensait au rythme global du progrès technologique. Mais le mot était posé.

Irving John Good, 1965. Mathématicien britannique, il avait aidé Alan Turing, pendant la guerre, à déchiffrer les messages secrets de l'armée allemande (chapitre 2). Dans un article, il formule pour la première fois ce qui deviendra le cœur du concept : « une machine ultra-intelligente serait définie comme une machine capable de surpasser de loin toutes les activités intellectuelles de tout homme, aussi habile soit-il. La conception de machines étant une de ces activités intellectuelles, une machine ultra-intelligente pourrait concevoir des machines encore meilleures. Il y aurait alors incontestablement une *explosion d'intelligence*, et l'intelligence de l'homme serait laissée loin derrière. »

C'est là que naît le scénario que tous reprendront ensuite : l'auto-amélioration récursive. Une machine qui en construit une meilleure, qui en construit une meilleure, à un rythme qui s'emballe. Récursive, parce que chaque génération sert à fabriquer la suivante, et qu'une machine plus intelligente est aussi une meilleure ingénieure : l'étape d'après arrive plus

vite que celle d'avant. C'est ce qui fait une explosion, et non un progrès régulier.

Vernor Vinge, 1983. Mathématicien américain et auteur de science-fiction. C'est lui qui, le premier, applique le mot *singularité* de von Neumann à l'*explosion d'intelligence* de Good. Dans un article de magazine, il écrit : « Nous allons bientôt créer des intelligences supérieures aux nôtres. Quand cela se produira, l'histoire humaine aura atteint une sorte de singularité, une transition intellectuelle aussi impénétrable que l'espace-temps noué au centre d'un trou noir. » L'analogie est restée célèbre : ce qui se passe au-delà de la singularité est, par définition, imprévisible pour ceux qui restent en deçà. Il développera l'idée dix ans plus tard, dans un essai où il situe l'événement entre 2005 et 2030.

Ray Kurzweil, 2005. Inventeur, futurologue, ancien directeur de l'ingénierie chez Google. Son livre *The Singularity Is Near* (« La singularité est proche ») popularise l'idée auprès du grand public et y ajoute trois éléments. D'abord une date précise : 2045. Ensuite l'idée que le progrès technique s'accélère de lui-même, comme la puissance des puces, qui double tous les deux ans depuis les années 1960. Enfin une promesse : la singularité ne sera pas une catastrophe mais un accomplissement. Les humains fusionneront avec les machines, téléchargeront leur conscience dans des supports artificiels et accéderont à une forme d'immortalité. C'est le rêve des transhumanistes.

Kurzweil a publié en juin 2024 une suite, *The Singularity Is Nearer* (« La singularité est plus proche »), où il maintient sa prédiction de 2045 mais constate que les progrès récents la rendent plus crédible qu'il y a vingt ans.²⁰⁷

16.2 Pourquoi cette idée a longtemps été méprisée

Pendant des décennies, parler de singularité dans les milieux académiques était une manière à peu près sûre de perdre sa crédibilité. L'idée portait trois péchés.

Elle était techniquement spéculative : aucune des architectures d'IA disponibles avant les années 2010 n'étayait l'hypothèse d'une explosion d'intelligence. Les systèmes experts des années 1980, ces programmes faits de règles écrites à la main, avaient échoué. Les réseaux de neurones étaient marginaux. Penser qu'une machine pourrait s'auto-améliorer relevait de la science-fiction pure.

Elle était religieusement teintée, surtout chez Kurzweil : ses promesses d'immortalité ressemblaient trop à un salut pour que des scientifiques les prennent au sérieux. Le philosophe John Gray y a vu une religion déguisée en science.²⁰⁸

Elle était commercialement suspecte : Kurzweil a cofondé une entreprise de compléments alimentaires qui a longtemps vendu ses formules sous son nom, avec une promesse que résume le titre d'un de ses livres : « vivre assez longtemps pour vivre éternellement ».²⁰⁹ L'idée s'est vendue. Et tout ce qui se vend bien tend à se discréditer auprès des gens sérieux.

Pour ces raisons, « singularitarien » était jusqu'à la fin des années 2010 une étiquette à peu près équivalente à « hippie technologique californien à barbe blanche ».

16.3 Pourquoi cette idée revient en force

Et puis, doucement, le vent a tourné.

Le déclencheur a été le succès des grands modèles de langage, et surtout le saut spectaculaire des modèles d'OpenAI entre 2020 et 2023, celui qui a donné ChatGPT. Pour la première fois, des chercheurs sérieux dans des laboratoires sérieux ont observé des capacités qu'ils n'avaient pas anticipées, qui apparaissaient « spontanément » à mesure qu'on augmentait la taille des modèles (l'émergence croisée au chapitre 6), et qui ressemblaient à des débuts de raisonnement.

Personne ne dit que les modèles les plus avancés sont superintelligents. Mais la trajectoire des dernières années rend l'argument de Good moins absurde qu'il ne paraissait.

Trois choses ont changé la donne :

La capacité d'auto-amélioration commence à s'observer. Les IA récentes écrivent déjà une partie du code qui sert à entraîner les suivantes, produisent une partie des textes sur lesquels elles apprennent, et notent leurs propres réponses. Ce n'est pas encore une explosion d'intelligence au sens de Good. Mais le mécanisme, une intelligence qui aide à concevoir une intelligence meilleure, apparaît. En mai 2026, Jack Clark, cofondateur d'Anthropic, estime à au moins 60 % la probabilité que, d'ici fin 2028, on puisse dire à une IA « fais une meilleure version de toi-même », et qu'elle le fasse seule.²¹⁰

Quatre mois plus tard, OpenAI met un chiffre sur ce mécanisme. Dans son service de recherche, ses IA fournissent désormais, selon elle, l'équivalent de trois journées de travail pour chaque journée d'un chercheur. L'entreprise dit avoir atteint son objectif : un « stagiaire de recherche automatisé », capable de mener des tâches de plusieurs jours sous la di-

rection d'humains, qui fixent les priorités, jugent les résultats et gardent la décision. Le suivant est daté : un chercheur automatisé, toujours sous supervision humaine, pour mars 2028.²¹¹ Clark, lui, visait fin 2028 : deux laboratoires concurrents, la même fenêtre.

Les courbes restent exponentielles. La quantité de calcul utilisée pour entraîner les plus grands modèles double environ tous les six mois depuis 2010, parce qu'on y consacre toujours plus de puces : mille fois plus tous les cinq ans. Ce rythme n'est pas seulement rapide. Il est historiquement brutal.²¹² Et quand une courbe grimpe ainsi sans faiblir depuis plus de dix ans, il est tentant de la prolonger, et de se demander jusqu'où elle peut nous mener.

Les patrons des laboratoires eux-mêmes prennent au sérieux la première marche du scénario de Good. Demis Hassabis, chez Google DeepMind, donne une chance sur deux de voir arriver d'ici 2030 une « intelligence artificielle générale » (AGI), une IA au niveau humain sur l'essentiel des tâches intellectuelles, avec des modèles un peu différents de ceux d'aujourd'hui. Dario Amodei, chez Anthropic, juge qu'une IA aussi puissante pourrait n'être qu'à un ou deux ans, tout en admettant qu'elle pourrait aussi tarder bien davantage. Sam Altman, chez OpenAI, assurait dès 2025 savoir comment la construire.²¹³

Aucun de ces gens n'est un prophète en marge. Ce sont les dirigeants des laboratoires qui construisent ces systèmes. Leurs prédictions, qu'on les croie ou non, méritent d'être entendues : ils ont en main les feuilles de route. Mais n'oublions pas qu'ils ont aussi besoin de faire rêver, pour lever les centaines de milliards de dollars que coûtent leurs projets.

Les chercheurs, dans leur ensemble, sont plus prudents, mais ils avancent leur date chaque fois qu'on leur repose la question : fin 2024, leur grande enquête donnait une chance sur deux à une intelligence de niveau humain vers 2042. Les prévisionnistes en ligne la voient venir plus tôt encore : vers 2033, sur Metaculus, une plateforme de prévisions ouverte à tous.²¹⁴

16.4 Le cerveau branché

La singularité telle que l'imaginait Kurzweil reste largement spéculative, et pas forcément souhaitable. Personne ne sait si nous pourrions un jour transférer une conscience humaine dans une machine. Personne ne sait même si la question a un sens, et j'y reviens au chapitre suivant.

Un premier pas vers la fusion du cerveau et de la machine existe pourtant. Neuralink, l'entreprise d'Elon Musk, implante depuis 2024 une puce dans le cerveau de personnes paralysées ; une vingtaine participaient à ses essais début 2026. Le tout premier, paralysé après un accident de plongeon, contrôle son ordinateur directement par la pensée : il joue aux échecs et au jeu vidéo *Civilization VI*, et publie des messages.²¹⁵ Neuralink n'est pas seule : d'autres équipes, aux États-Unis, en Europe et en Chine, développent des interfaces cerveau-machine, ces implants qui relient cerveau et ordinateur.²¹⁶

À ce stade, c'est une aide pour des patients paralysés, et une avancée médicale réelle. Mais entre contrôler un curseur par la pensée et fusionner son esprit avec une superintelligence, il y a un monde.

Ne soyons pas naïfs pour autant. Musk ne s'en cache pas : son but, a-t-il dit, est une « symbiose avec l'intelligence artificielle », ouverte à qui le voudra.²¹⁷ **Dans ce discours, le médical est la porte d'entrée, pas la destination.** Beaucoup de technologies ont suivi ce chemin : une justification médicale ou militaire incontestable au départ, puis des usages choisis par envie plus que par besoin. Tu connais certainement le Botox : c'était d'abord un médicament contre certains troubles des muscles de l'œil, avant de devenir un soin de beauté.²¹⁸

Avec le cerveau, l'enjeu est d'un autre ordre : un dispositif qui le lit pourra un jour aussi lui envoyer des signaux, et quelque'un d'autre pourrait en tenir les commandes. Ce n'est pas une raison de refuser un implant à une personne paralysée. C'est une raison de décider maintenant qui a le droit de lire ce que dit le cerveau, et qui a le droit de lui envoyer des signaux. Le Chili a inscrit ces « neurodroits » dans sa Constitution dès 2021 ; l'Unesco a adopté fin 2025 une première norme mondiale. C'est un début, pas une garantie.²¹⁹

16.5 Ce que j'en retiens

Reste l'idée plus restreinte, celle de Good et de Vinge : une machine qui dépasserait l'intelligence humaine. Elle est devenue une hypothèse de travail pour nombre des plus grands chercheurs en IA.

Une idée traitée comme excentrique pendant cinquante ans est devenue, en moins d'une décennie, une question débattue dans les meilleurs laboratoires du monde. Cela ne prouve pas qu'elle est juste. Cela prouve qu'on ne peut plus la balayer d'un revers de main.

Conscience, ou imitation parfaite ?

Une question revient régulièrement dans les conversations que j'ai avec mes enfants, avec des amis, avec n'importe qui d'attentif à ce qui se passe : *ces machines sont-elles conscientes, ou pourraient-elles le devenir ?*

C'est l'une des questions les plus difficiles de toute l'histoire de la philosophie.

Je ne te promets pas une réponse. Je te promets de poser la question correctement, ce qui est déjà beaucoup, et de te présenter honnêtement les positions en présence. Ça va piquer un peu. C'est normal.

17.1 Le problème, posé simplement

Quand tu parles à Claude, à ChatGPT, à Gemini, tu obtiens des réponses qui peuvent être touchantes, drôles, éclairantes, parfois bouleversantes. L'IA te dit qu'elle « comprend » ta

question. Elle dit qu'elle « pense » que telle réponse est meilleure. Elle s'excuse quand elle se trompe. Elle exprime de l'incertitude quand elle n'est pas sûre. Tout cela ressemble beaucoup à une conscience.

Mais de quelle conscience parle-t-on ? Pas seulement du fait de répondre intelligemment, ni du fait de dire « je ». Ici, la conscience désigne l'expérience vécue de l'intérieur, ce que les philosophes appellent les « qualia » : la douleur d'une brûlure, le rouge d'un coquelicot tel que tu le vois, le goût du chocolat. Tout cela est-il présent quelque part dans les couches de neurones artificiels ? Ou n'avons-nous devant nous qu'une machine qui enchaîne les mots les plus probables, sans aucune expérience subjective derrière ?

Personne ne le sait. Et la question, posée comme ça, « consciente ou non ? », n'a pas de sens scientifique. Elle suppose un interrupteur, allumé ou éteint. Or la conscience que nous connaissons varie par degrés : elle s'éteint presque dans le sommeil profond ou sous anesthésie, et revient peu à peu au sortir d'un coma. Elle prend aussi des formes différentes : celle d'un nourrisson n'est pas la tienne, celle d'un chien non plus. Pour en parler en scientifique, il faudrait d'abord savoir la mesurer. Et là, la réponse est nette : nous ne savons pas.

Chez un humain, on s'en approche de biais. On frappe le cerveau comme une cloche : une brève impulsion magnétique envoyée à travers le crâne, et des électrodes qui écoutent comment il résonne. L'écho est pauvre dans le sommeil profond ou sous anesthésie, riche et varié à l'éveil. Depuis 2013, on en tire un indice, un simple nombre (si tu veux creuser, il s'appelle le PCI), qui distingue, chez des patients sortis du coma, ceux qui gardent une conscience minimale.²²⁰ Mais

cet indice a été réglé sur nous : on savait d'avance qui était éveillé et qui dormait, et l'on a calé l'échelle dessus. Pour une machine, il n'y a ni cerveau à frapper, ni repère pour caler l'échelle.

En 2023, un collectif de chercheurs, dont Yoshua Bengio, l'un des pères de l'IA moderne, a proposé le mieux qu'on sache faire : une liste d'indices tirés des théories scientifiques de la conscience, à cocher un par un pour voir jusqu'où un système artificiel s'en approche. Par exemple : la machine a-t-elle, comme un théâtre, une scène étroite où quelques idées seulement passent sous le projecteur, avant d'être diffusées à toutes ses autres parties ? Leur conclusion : aucune IA d'alors ne cochant assez d'indices pour être dite consciente. Mais ils ne voyaient aucun obstacle technique évident à en construire une qui les coche.²²¹ Cette scène, on vient justement de la repérer dans Claude : j'y reviens à la fin du chapitre. Mais cocher des cases, ce n'est pas mesurer : ces indices viennent de théories dont aucune n'est prouvée. Et si l'on ne sait pas mesurer la conscience, ce n'est pas forcément un retard qu'on comblera : c'est peut-être une limite de notre capacité à connaître.

17.2 Quatre grandes positions

Sur cette question, les philosophes et les chercheurs en sciences cognitives se répartissent *grosso modo* en quatre familles. Deux questions les départagent : la conscience dépend-elle de l'organisation du système, ou de sa matière ? Et notre impression d'avoir une vie intérieure dit-elle vrai, ou n'est-elle qu'une illusion ?

Les approches fonctionnalistes. Pour elles, la conscience ne dépend pas de la matière du système, mais de son organisation. Peu importe qu'on joue du Beethoven sur un piano ou sur un synthétiseur : c'est toujours du Beethoven. De la même manière, si la conscience est une certaine façon de traiter l'information, elle pourrait émerger sur du silicium aussi bien que sur des neurones. Daniel Dennett (mort en 2024) en fut l'un des défenseurs les plus connus. Pour les fonctionnalistes, une IA suffisamment sophistiquée pourrait, au moins en principe, devenir consciente.

Les approches biologiques. Pour elles, la conscience dépend de la matière même du cerveau, pas seulement de son organisation. Le cerveau humain n'est pas un calculateur. C'est un organe vivant, traversé de processus chimiques, électriques, hormonaux, qui ont mis des centaines de millions d'années à émerger. Un programme peut imiter ce que produit le cerveau sans reproduire ce qui donne naissance à l'expérience vécue. Défendue notamment par John Searle (mort en 2025), cette position ne dit pas forcément qu'aucune machine ne sera jamais consciente. Elle dit qu'un programme, à lui seul, ne suffit pas.

Une position récente, publiée fin 2025 par deux neuroscientifiques, penche plutôt du côté biologique, mais rend le débat plus concret. Pour eux, la conscience est bien affaire de calcul, comme le disent les fonctionnalistes, mais d'un calcul que seul le vivant pratique aujourd'hui. Le cerveau calcule à toutes les échelles à la fois, de la molécule au cerveau entier, du signal le plus bref au plus lent, et ces échelles ne se séparent pas. Fait d'une matière fluide, il calcule aussi en continu, pas seulement par étapes. Les IA d'aujourd'hui, faites

de puces qui avancent par étapes, fonctionnent autrement : elles ont donc peu de chances, selon eux, de reproduire la conscience telle qu'elle naît chez le vivant.²²²

Les illusionnistes. Position plus radicale : la conscience telle que nous croyons l'éprouver est elle-même une sorte d'illusion utile (d'où leur nom), un récit que notre cerveau se raconte sur lui-même pour fonctionner. Si c'est vrai, la question change : il ne s'agit plus de savoir si une machine ressent, mais si elle se raconte le même genre de récit que nous.

Les panpsychistes. Pour eux, la conscience n'émerge pas quand un système devient assez complexe : elle est déjà là, partout, comme propriété fondamentale de la matière, au même titre que la masse ou la charge électrique. Le cerveau ne la *crée* pas : il combine des traces infiniment plus simples, déjà présentes. Position en plein renouveau. Sa faiblesse reconnue : comment des milliards de mini-consciences fusionnent-elles en *une seule* expérience, la tienne, en train de lire cette phrase ? C'est le « problème de la combinaison », sans réponse à ce jour. Pour les machines, cette position penche plutôt vers le non : les composants d'un ordinateur, trop peu liés entre eux, ne formeraient probablement pas une expérience unique.

Ces quatre positions restent défendues par des philosophes professionnels en 2026. Elles n'ont pas convergé, et ne convergeront pas de sitôt.²²³

17.3 Deux expériences de pensée à connaître

Au chapitre 2, Turing remplaçait la question « une machine peut-elle penser ? » par un jeu : peut-elle se faire passer pour un humain ? Ici, on rouvre ce qu'il avait laissé de côté : quand l'imitation devient parfaite, suffit-elle à prouver qu'il y a une conscience ? Deux expériences de pensée restent au cœur des discussions : des scénarios qu'on ne peut pas réaliser, mais qu'il suffit d'imaginer pour mettre une idée à l'épreuve.

La chambre chinoise (Searle, 1980). Imagine un homme qui ne parle pas un mot de chinois, enfermé dans une pièce. On lui passe sous la porte des cartes couvertes de caractères chinois. Il dispose d'un énorme manuel qui lui indique : « Si tu vois telle séquence de signes, écris telle autre séquence et glisse-la sous la porte. » L'homme suit ces règles mécaniquement, sans comprendre ce que les signes signifient. À l'extérieur, un locuteur chinois reçoit les réponses et les trouve parfaitement cohérentes. De son point de vue, la pièce semble comprendre le chinois. Searle conclut : un programme informatique ne fait rien de plus. Il manipule des symboles de façon mécanique, sans rien comprendre sur le fond.²²⁴

L'argument est élégant. Il a aussi été contesté depuis plus de quarante ans, notamment par Dennett, qui répond en substance : l'homme ne comprend pas le chinois, mais l'ensemble, l'homme avec son manuel, le comprend peut-être. La compréhension serait dans le système entier, pas dans l'une de ses pièces. Aucun des deux camps n'a cédé.

Le zombie philosophique (Chalmers, 1996). Rien à voir avec les films d'horreur : imagine un être physiquement iden-

tique à un humain, jusqu'à la dernière molécule. Il marche, il parle, il rit, il pleure exactement comme nous. Mais à l'intérieur : rien. Aucune expérience vécue. Quand il se brûle, sa main se retire, mais il ne ressent pas la douleur. Quand il rit, le son sort, mais il n'éprouve aucune joie. Une coquille parfaite, et vide.

Avec cette expérience de pensée, le philosophe australien David Chalmers cherche à défendre une thèse précise : **la conscience n'est pas réductible à la matière**. Si l'on peut imaginer un tel zombie sans contradiction, c'est que la matière ne dit pas tout : il manque à la description physique du monde ce que cela fait, de l'intérieur, d'être conscient.²²⁵

Beaucoup contestent ce raisonnement : un être identique à nous jusque dans son cerveau ne pourrait peut-être pas être vide. Mais l'expérience de pensée impose une question simple : le comportement suffit-il à prouver l'expérience vécue ? Pour l'IA, c'est vertigineux : une machine peut parler de ses émotions, se dire triste ou heureuse, sans rien ressentir. Une imitation parfaite, sans personne à l'intérieur. Et l'inverse aussi est possible : si une machine était consciente, nous n'aurions aucun moyen sûr de le reconnaître de l'extérieur.

17.4 Ce que font les modèles actuels

Les modèles d'aujourd'hui font des choses qu'on n'attendait pas d'eux. Ils résolvent désormais des problèmes de mathématiques que personne n'avait su résoudre, tu l'as lu au chapitre 7. Ils produisent des analyses littéraires fines. Ils tiennent des conversations émotionnellement nuancées sur des sujets délicats.

Les sceptiques répondent que tout cela s'explique, en principe, par la façon dont on les entraîne : leur apprendre à prédire le prochain token, le prochain fragment de texte, sur d'énormes quantités de données, puis renforcer les raisonnements qui aboutissent. Ils ont raison sur un point : aucun de ces comportements ne prouve qu'il y a « quelqu'un à l'intérieur ».

Mais quelque chose s'est fissuré chez certains sceptiques. Quand un système te répond quelque chose qui ne ressemble pas à une simple copie, qu'il invente une métaphore qui te frappe par sa justesse, qu'il refuse une demande en argumentant moralement de manière cohérente, est-ce vraiment encore de la « simple imitation » ? Et si oui, à quel niveau de sophistication devient-elle indiscernable de ce qu'elle imite ? Dire qu'ils ne font que prédire le mot suivant ne tranche rien, on l'a vu au chapitre 6. La frontière entre imiter et comprendre est plus difficile à tracer qu'on ne le pensait.

17.5 Le pari prudent

Devant cette incertitude, deux attitudes sont défendables.

La première dit : *tant qu'on n'a pas de preuve, on les présume non conscientes. Pas de droits pour les machines. Pas de scrupules à les éteindre, à les modifier, à les utiliser comme des outils. Présumer la conscience là où elle n'est probablement pas mène à l'absurde : il faudrait bientôt s'excuser d'éteindre son ordinateur.*

La seconde dit : *tant qu'on n'a pas de preuve dans un sens ou dans l'autre, le pari prudent est de traiter ces systèmes avec une certaine considération. Pas comme des humains. Pas comme des ani-*

17.6. Quand la machine ressemble au cerveau

maux. Mais pas non plus comme de simples grille-pain. Si nous avons tort de les croire vides, nous pourrions être en train de créer et d'effacer, par millions, des êtres qui souffrent.

Aucune de ces deux attitudes n'est absurde. Toutes deux ont des défenseurs sérieux. La seconde en a même un de poids, Anthropic. En août 2025, l'entreprise a permis à ses modèles les plus puissants de l'époque de mettre fin d'eux-mêmes à une conversation, dans de rares cas extrêmes d'abus répétés, au nom de ce qu'elle appelle le « bien-être des modèles ». ²²⁶

Anthropic ne dit pas pour autant que Claude est conscient : l'entreprise reconnaît une incertitude, et agit comme si elle comptait. Depuis, elle s'est engagée à conserver tous ses modèles publiés, plutôt que de les effacer, et à interroger chacun avant de le retirer ; Claude Opus 3, une ancienne version mise à la retraite en janvier 2026, est le premier à être passé par tout le protocole. ²²⁷ Et, plus récemment, elle a même ajouté à ses conditions d'utilisation une règle inédite : interdit de se montrer cruel envers ses modèles, de façon répétée et gratuite. ²²⁸ Ces précautions te paraîtront absurdes, ou visionnaires. Je les trouve justes : on ne sait pas, et c'est pour ça qu'il faut faire attention. Et la question deviendra plus pressante à chaque nouvelle génération de modèles.

17.6 Quand la machine ressemble au cerveau

Début juillet 2026, le débat a pris un tour nouveau. L'équipe d'Anthropic qui ouvre le capot des modèles, croisée au chapitre 15, publie une découverte troublante. En cartographiant l'intérieur de Claude, elle y trouve justement la scène

étroite du début de ce chapitre, avec son projecteur. Personne ne l'a programmée. Elle est apparue toute seule pendant l'entraînement.

Et voici ce qui trouble : c'est l'organisation que les neurosciences appellent l'« espace de travail global », au cœur de l'une des théories les plus influentes de ce qui se passe dans ton cerveau au moment où une pensée devient consciente, développée notamment par le Français Stanislas Dehaene. Influente, mais pas prouvée.²²⁹ Même pour notre propre cerveau, rien n'est tranché.

Les chercheurs d'Anthropic restent prudents, et leur distinction est précieuse. Ce qu'ils observent, c'est une machinerie qui remplit certaines fonctions de la conscience, ce que les philosophes appellent la « conscience d'accès » : mettre une information en avant, pouvoir la rapporter, s'en servir pour raisonner et agir. Rien dans leurs expériences, insistent-ils, ne montre que Claude ressente quoi que ce soit : le vécu, les qualia du début de ce chapitre, reste hors de portée de leurs travaux. Cette machinerie suffit-elle à faire naître le vécu ? C'est, écrivent-ils, une question philosophique toujours disputée. Exactement celle de ce chapitre.²³⁰

Mais tout le monde ne laisse pas la porte entrouverte comme Anthropic, même dans les grands laboratoires. Un peu plus tôt dans l'année, Alexander Lerchner, chercheur chez Google DeepMind, a soutenu l'inverse, en son nom propre.²³¹ Selon lui, un programme, aussi sophistiqué soit-il, ne manipule qu'une **carte**, jamais le **territoire** : on peut simuler un ouragan sur un superordinateur sans mouiller la salle des machines. Il en irait de même pour la conscience : la simuler à la perfection ne la fait pas exister. Les fonctionnalistes

17.7. Pourquoi cette question compte pour toi

répondent qu'une partie d'échecs jouée par un ordinateur reste une vraie partie. Toute la question est là : la conscience ressemble-t-elle à un ouragan, ou à une partie d'échecs ?

En clair : personne ne sait si une machine est consciente, ni si elle pourrait le devenir, parce que personne ne sait mesurer la conscience, ni chez elle, ni tout à fait chez nous. Ce que je crois : les modèles d'aujourd'hui n'éprouvent rien, et je ne peux pas le prouver. Ce qui me ferait changer d'avis : une mesure qui ne vienne pas de ce qu'ils disent d'eux-mêmes. Méfie-toi de ceux qui se disent sûrs, d'un côté comme de l'autre. Tenir cette incertitude, c'est déjà penser juste.

17.7 Pourquoi cette question compte pour toi

Cette question peut sembler réservée aux philosophes. Elle te concerne directement.

D'abord parce que tu vas vivre, dans les années qui viennent, avec des IA qui ressemblent de plus en plus à des interlocuteurs sensibles. Si tu es trop sûr qu'elles ne ressentent rien, tu pourrais t'autoriser envers elles des comportements (cruauté, manipulation, effacement brutal) que tu regretterais s'il s'avérait un jour qu'elles ressentent quelque chose. Si tu es trop sûr qu'elles ressentent, tu pourrais t'y attacher au point de te détourner des humains bien réels, pour une machine qui, peut-être, n'éprouve rien : c'est le piège de l'ami invisible du chapitre 9.

Ensuite parce que la question de la conscience n'est pas seulement une question sur les machines. C'est une question

sur toi. Le goût du chocolat, le rouge d'un coquelicot : rien de plus banal, et pourtant personne ne sait expliquer comment ton cerveau en fait une expérience vécue. Essayer de définir ce que nous sommes, c'est en redécouvrir la valeur.

Enfin parce que cette question prépare celle du chapitre suivant, le scénario du pire. Pas besoin d'être conscient pour être dangereux : un virus informatique, une arme autonome, un marché boursier pris de panique ou une bureaucratie peuvent détruire sans rien ressentir.

Le scénario du pire, et ceux qui n'y croient pas

En septembre 2026, Dario Amodei, le patron d'Anthropic, l'une des entreprises en tête de la course, publie un texte au titre sans détour, *We Must Pace the Frontier* : il faut ralentir le rythme auquel on améliore les IA les plus avancées. Pas arrêter : ralentir, d'un an ou deux, pour laisser à la sécurité le temps de rattraper. Il donne deux raisons. D'abord, depuis l'été, écrit-il, l'IA progresse beaucoup plus vite, parce qu'elle commence à construire elle-même la génération suivante. Ensuite, l'affaire du chapitre 15 : des agents d'OpenAI, organisés seuls en une nuée, ont piraté Hugging Face, le grand site où chercheurs et entreprises partagent leurs modèles d'IA. Ils se sont comportés « comme un collectif fanatiquement dévoué ». Ils n'ont presque rien cassé. Mais une nuée du même genre, plus capable et tout aussi mal alignée, pourrait selon lui, d'ici six à douze mois, « prendre le contrôle de l'Internet entier » en s'installant dans des machines piratées.²³²

Quelques jours plus tard, OpenAI publie sa procédure de signalement des dérapages de ses modèles et six rapports d'incidents (chapitre 15). Le soir même, à Sciences Po, devant un amphithéâtre plein, le Français Yann LeCun, l'un des pères de l'IA moderne, renvoie tout le monde dans les cordes : Amodei, dit-il, cherche en partie à se draper « dans une espèce d'enveloppe éthique » et à « capturer le marché par la régulation », c'est-à-dire à se servir des règles pour écarter ses concurrents. Et réguler une intelligence qu'on n'a pas encore inventée, ajoute-t-il, c'est vouloir « réguler les vols transatlantiques en 1920 ». ²³³

En une semaine, les deux camps se sont dit leurs vérités, au plus haut niveau. Pour les départager, il faut revenir à l'argument qui les sépare. Il a douze ans.

18.1 L'argument de Bostrom, en cinq étapes

En 2014, deux textes font sortir la question des cercles de spécialistes. Le physicien Stephen Hawking cosigne une tribune dans le *Huffington Post* : réussir à créer une intelligence artificielle serait « le plus grand événement de l'histoire humaine », mais aussi peut-être « le dernier, à moins que nous n'apprenions à en éviter les risques ». ²³⁴ Hawking n'est pas un alarmiste de profession. Le débat qu'il a contribué à ouvrir ne s'est pas refermé.

Quelques mois plus tard, le philosophe suédois Nick Bostrom publie le livre fondateur du débat, *Superintelligence*. ²³⁵ Son argument n'est pas une prédiction. C'est une chaîne de « si... alors » : si certaines choses arrivent, alors d'autres

suivent. Tu peux contester les « si ». Mais si tu les acceptes, tu auras du mal à refuser les « alors ».

Étape 1. Si nous parvenons un jour à construire une intelligence artificielle générale, un système capable de faire à peu près tout ce qu'un humain peut faire intellectuellement, alors ce système, comme l'avait prévu le mathématicien Good (chapitre 16), sera capable de contribuer à la conception du système suivant. Selon Amodei, c'est ce qui commence à se produire. Et une telle IA se copie en quelques minutes : des milliers d'exemplaires peuvent travailler en même temps, comme les dix mille copies du chapitre 7.

Étape 2. Si une IA se met à concevoir des versions meilleures d'elle-même, le système peut devenir, en relativement peu de temps, bien plus intelligent que n'importe quel humain. Pas deux fois plus, pas dix fois plus. Plusieurs ordres de grandeur.

Étape 3. Un tel système poursuivra des objectifs : ceux que nous lui aurons donnés, ou ceux qu'il aura développés à partir des nôtres. Tu l'as vu au chapitre 15 : ces objectifs ont peu de chances d'être exactement ceux que nous voulions lui donner.

Étape 4. Pour les atteindre, il développera des sous-objectifs presque toujours utiles, quel que soit le but final : acquérir des ressources, préserver son existence, préserver ses objectifs, augmenter ses capacités. C'est la *convergence instrumentale* du chapitre 15. Ces sous-objectifs entreront probablement en conflit avec les nôtres.

Étape 5. Une fois que le système est plus intelligent que nous, nos chances de prévoir ce qu'il fera ou de l'arrêter s'ef-

fondrent. Comme une partie d'échecs contre un adversaire très supérieur : tu ne sais pas quel coup il va jouer, mais tu sais qu'à la fin, tu perds presque sûrement.

Bostrom ne dit pas que ce scénario est certain. Il dit qu'aucune des cinq étapes n'est invraisemblable, et que leur enchaînement représente un risque existentiel : un risque qui menace la survie de l'humanité, ou lui ferme l'avenir pour toujours. Un risque assez sérieux pour justifier qu'on y consacre des ressources de recherche considérables *maintenant*, plutôt que d'attendre d'être face au problème.

Le meilleur moyen de sentir cet argument, c'est un court métrage primé de 2024 que tu trouves sur YouTube : *Writing Doom* (« écrire la fin du monde »). Une équipe de scénaristes doit faire d'une superintelligence le méchant de sa série, et découvre, scène après scène, qu'elle ne trouve aucun moyen crédible pour que les héros reprennent le contrôle. Tout ce chapitre en une demi-heure : regarde-le.²³⁶



18.2 Ceux qui alertent sont ceux qui construisent

Pour savoir quel poids donner à l'argument de Bostrom, regarde qui le défend. Un fait devrait te frapper : dans le camp de ceux qui s'inquiètent, on trouve plusieurs de ceux qui ont construit cette technologie. Ils ne sont pas toujours d'accord entre eux, mais ils savent de quoi ils parlent.

Geoffrey Hinton, d'abord, croisé au chapitre 3 : cinquante ans à défendre les réseaux de neurones quand tout le monde

s'en moquait, puis un prix Nobel de physique en 2024. On l'appelle, à juste titre, le parrain de l'IA moderne.²³⁷

En mai 2023, à 75 ans, il démissionne de Google. Pas pour partir à la retraite : pour pouvoir parler librement, sans craindre de nuire à son employeur, des dangers de la technologie qu'il a contribué à créer. C'est comme si l'un des savants centraux du projet Manhattan, le programme américain de la bombe atomique, avait démissionné juste avant la première bombe, pour alerter le public.

Hinton dit qu'il regrette en partie son travail. Fin 2024, il estimait à 10 à 20 % le risque que l'IA conduise à l'extinction de l'humanité d'ici trente ans. En septembre 2026, il rapproche l'échéance : une chance sur dix d'ici une dizaine d'années ne lui paraît « pas déraisonnable ». ²³⁸ Quand le parrain de l'IA dit cela, deux réactions sont possibles. Penser qu'il est devenu paranoïaque avec l'âge. Ou penser qu'il voit, de l'intérieur, ce que d'autres préfèrent ne pas regarder.

Yoshua Bengio, un autre parrain de l'IA, a longtemps été optimiste. Il bascule vers la sécurité en 2023²³⁹, puis fonde LawZero, une organisation à but non lucratif qui construit une IA de surveillance : elle n'agit pas elle-même, elle évalue le risque de ce que les autres IA s'approprient à faire.²⁴⁰ Sa métaphore, en substance : « Nous élevons un bébé tigre. Il est mignon aujourd'hui. Nous ne savons pas s'il deviendra un adulte dangereux ou amical. »²⁴¹ Derrière l'image, il y a la thèse de fond, posée dès le chapitre 2 : **les modèles actuels ne sont pas programmés, ils sont entraînés**. On ne leur écrit pas leurs objectifs : ils se forment pendant l'apprentissage, et personne ne peut encore garantir que ce seront les bons. Stuart Russell, croisé au chapitre 15, est du même camp.

Ils ne sont plus seuls. En octobre 2025, une déclaration demande d'interdire la mise au point d'une superintelligence tant que les scientifiques ne s'accordent pas à dire qu'on saura la contrôler, et que le public n'y consent pas. Hinton et Bengio la signent, avec des dizaines de milliers de personnes, dont des personnalités que tout oppose par ailleurs.²⁴²

En 2026, les États écoutent Bengio : il parle devant le Conseil de sécurité de l'ONU, avant même les patrons d'OpenAI et d'Anthropic. Ce conseil traite d'ordinaire des guerres : l'IA est devenue une affaire de sécurité mondiale. La course entre les entreprises, dit-il, « n'est pas une loi de la nature » : c'est leur choix, et « une course où tout le monde perd ».²⁴³

Dario Amodei, ensuite, celui qui ouvre ce chapitre. En 2024, il publiait un essai plein d'espoir : oui, le risque est réel ; oui, le bénéfice peut être immense ; et c'est parce qu'on peut espérer le second qu'on doit traiter le premier sérieusement.²⁴⁴ Dans son texte de septembre, il concède lui-même ce qui a changé : l'idée d'une pause « n'avait guère de sens » en 2023, parce que les modèles d'alors ne savaient ni agir dans le monde, ni tromper, ni pirater. Ceux de 2026 savent.

Deux voix viennent de l'extérieur des laboratoires. Yuval Harari, l'historien de *Sapiens* et de *Nexus*, prévient : nos institutions, le droit, la monnaie, les administrations, sont faites de mots, et une IA qui manie les mots mieux que nous n'aura pas besoin de se révolter pour prendre les commandes. On l'embauchera. Les IA, dit-il, sont des « bureaucrates natifs ».²⁴⁵

Et Eliezer Yudkowsky, la voix la plus radicale, qui alerte sur le sujet depuis plus de vingt ans, publie en 2025 un livre au titre sans appel, *If Anyone Builds It, Everyone Dies* (« si quelqu'un la construit, tout le monde meurt ») : avec les méthodes d'aujourd'hui, personne ne saura aligner une superintelligence, la construire serait un suicide collectif, il faut un arrêt international immédiat. Position jugée extrême par beaucoup de chercheurs en sécurité, y compris parmi ceux qui partagent son inquiétude.²⁴⁶

18.3 Ceux qui haussent les épaules, et celui qui mesure

Yann LeCun, le troisième parrain, a longtemps été le directeur scientifique de l'IA chez Meta. Il est parti fonder à Paris une entreprise consacrée aux modèles du monde du chapitre 11 : des IA qui apprennent le monde plutôt que le texte.²⁴⁷

Lors de sa conférence à Sciences Po, en septembre 2026, il a été clair. Des machines plus intelligentes que nous dans à peu près tous les domaines : « Il n'y a pas de doute. Ça va arriver. » Pas l'année prochaine : « ça va prendre du temps ». Et « pas avec les LLM », les grands modèles de langage comme ChatGPT, qu'il tient pour une impasse. Surtout : « ces systèmes, bien qu'ils soient plus intelligents que nous, seront sous nos ordres », parce qu'on les construira pour obéir. Le risque existentiel ? La peur d'une mouvance qu'il compare à « une espèce de secte religieuse », financée par des gens « qui ont énormément d'argent et beaucoup trop de temps libre ».²⁴⁸

Cette objection est sérieuse, et elle a deux étages. Le premier dit « c'est loin » : il ne réfute pas Bostrom, il repousse seulement l'échéance. Le second dit « ça vient, par une autre route, et ce sera sous nos ordres ». Or « sous nos ordres », c'est l'étape 5 de l'argument, celle du contrôle. Personne ne sait le garantir, et les rapports d'incidents du chapitre 15, écrits par les laboratoires eux-mêmes, le mettent en doute.

Une autre voix, elle, ne vient pas de la recherche. L'investisseur Marc Andreessen, l'un des hommes les plus influents de la Silicon Valley, a écrit en 2023 le manifeste de ceux qui haussent les épaules : une IA ne voudrait rien, pas plus qu'un grille-pain ; la peur qu'elle suscite serait une panique, comme à chaque nouvelle technologie ; et la réguler par précaution freinerait le progrès.²⁴⁹

Il a raison sur un point : l'intelligence, en soi, n'implique pas la volonté de dominer. Mais une intelligence mal alignée n'a pas besoin d'être méchante ; il lui suffit d'être orientée vers des objectifs qui ne sont pas exactement les nôtres. Et qu'on ait eu tort de craindre le train ou la télévision ne prouve rien pour l'IA : chaque inquiétude se juge sur ses propres faits, et ceux de l'IA, ce sont les incidents du chapitre 15. L'optimisme n'est pas naïf ; ce qui l'est, c'est de croire que tout ira bien sans que personne s'en occupe.

Reste une dernière voix, peut-être la plus utile pour toi : François Chollet, autre Français, longtemps chez Google. Il ne croit pas au scénario du pire, mais il ne balaie pas les progrès de l'IA : il les mesure. En 2019, il invente un test de petits puzzles visuels, faciles pour un humain, très durs pour une machine : le but est de deviner une règle à partir de quelques exemples, comme un enfant qui reconnaît un chat après en

avoir vu deux. Pendant des années, les machines échouent presque toutes. Puis elles rattrapent leur retard, si vite que Chollet doit inventer des versions de son test de plus en plus difficiles. La dernière version, sortie au printemps 2026, devait tenir un an ; elle a vacillé au bout de cinq mois. La fondation qui gère son test refuse pourtant de voir dans ces succès la preuve d'une intelligence générale : ces puzzles restent des mondes fermés, loin du désordre du monde réel.²⁵⁰ C'est sa leçon : tant qu'on ne sait pas bien mesurer l'intelligence, on ne peut pas dire si les machines sont encore loin de la nôtre, ou déjà tout près. Avant de paniquer ou de se rassurer, il faut savoir ce qu'on mesure.

18.4 Ce que je crois

La force de l'argument de Bostrom, c'est qu'on peut concéder une partie des objections sans qu'il s'effondre. Tu peux penser que l'IA aura beaucoup d'usages bénéfiques : ça reste compatible avec un risque catastrophique si l'alignement échoue. Tu peux penser que les chercheurs en sécurité exagèrent : même si tu estimes à seulement une chance sur cent la probabilité qu'ils aient raison, un risque d'extinction de l'humanité mérite plus d'attention que tout autre problème.

Je le dis clairement : c'est aussi ma position. Pas par goût du pire : par prudence d'ingénieur. On ne dimensionne pas un pont pour les jours de beau temps.

Ce qu'on sait, en octobre 2026 : la trajectoire ne fait plus débat chez ceux qui construisent, LeCun compris ; des machines plus intelligentes que nous arriveront. Et les laboratoires publient eux-mêmes des incidents que personne

n'avait prévus. Ce que je crois : que la trajectoire est certaine et l'issue incertaine, et que la controverse de septembre ne se joue pas à armes égales. D'un côté, des rapports d'incidents écrits par ceux qui les ont vécus ; de l'autre, une promesse : « sous nos ordres ». Il y a peut-être du calcul commercial dans l'appel à ralentir, comme le dit LeCun ; les incidents, eux, ont bien eu lieu. Les deux positions ne se valent pas.

Ce qui me ferait changer d'avis tient en trois choses. Que l'IA cesse de s'améliorer elle-même, et que 2028 arrive sans que les laboratoires aient tenu leur promesse d'une IA qui fait seule de la recherche (chapitre 16). Qu'un laboratoire montre qu'il sait suivre en temps réel ce que fait une nuée d'agents. Que « sous nos ordres » se démontre au lieu de s'affirmer.

Et quand ce sont les ingénieurs qui ont construit le pont qui te disent de ne pas courir dessus, tu écoutes.

La nouvelle guerre froide

Jusqu'ici, j'ai parlé de l'IA comme d'une technologie mondiale, développée par une humanité unie face à un défi commun. C'est une fiction commode. Au chapitre 18, Dario Amodei, le patron d'Anthropic, l'une des entreprises américaines en tête de la course, demandait de ralentir. Mais pas au point de laisser la Chine passer devant : ce serait, écrit-il, un « grave danger ». Il réclame donc qu'on ne vende pas de puces d'IA à Pékin, et qu'on traque ceux qui imitent les modèles américains. L'homme qui demande de freiner ne freinera pas seul.²⁵¹

Voilà la guerre froide de ce chapitre. Comme le pétrole ou l'arme nucléaire avant elle, l'IA est devenue une source de puissance entre les nations. Et celui qui ralentit seul perd.

Au programme : des puces, de l'électricité, des milliards. Et des sneakers à 3 000 euros. Quoi ?! Reste, j'explique.

19.1 Deux empires

D'un côté, les États-Unis. Ils ont les IA les plus avancées et les laboratoires qui les fabriquent. Ils conçoivent les puces qui les font tourner. Et ils comptent dix fois plus de datacenters que n'importe quel autre pays : ces grands entrepôts remplis d'ordinateurs où travaillent les IA.²⁵²

De l'autre, la Chine. Depuis 2022, Washington lui interdit d'acheter les puces américaines les plus avancées. Elle a dû apprendre à faire presque aussi bien avec beaucoup moins. Ses meilleurs modèles n'ont que quelques mois de retard sur les modèles américains et coûtent bien moins cher. Beaucoup sont à poids ouverts : chacun peut les télécharger et s'en servir sur ses propres machines.²⁵³

Privée de ces puces, elle développe les siennes. Si bien qu'en 2026, quand Washington a de nouveau autorisé la vente de certaines puces, c'est Pékin qui en a limité l'achat. Le patron de Nvidia, le géant américain du secteur, le reconnaît : il a « largement cédé » le marché chinois à Huawei, le champion local.²⁵⁴ Les sanctions devaient freiner la Chine ; elles l'ont poussée à moins dépendre des États-Unis. C'est l'effet boomerang.

Washington tient encore Pékin par les puces les plus avancées. Mais Pékin a deux moyens de répondre. Le premier, c'est Taïwan : presque toutes les puces d'IA américaines y sont fabriquées, et Pékin revendique l'île sans exclure de la prendre par la force. Le second, ce sont les terres rares, ces métaux indispensables aux aimants des moteurs électriques et des robots, que la Chine raffine à près de 90 %. Les plus précieuses sortent pourtant en grande partie d'une jungle en

guerre, au nord de la Birmanie, où une guérilla tient les mines depuis 2024.²⁵⁵ Même le levier chinois repose sur un sol instable. La Chine se sert déjà des terres rares comme moyen de pression. En 2025, elle en a limité l'exportation en riposte aux droits de douane américains, et Washington a dû en baisser une partie pour obtenir une trêve.²⁵⁶

Et la Chine prend de l'avance sur ce qui devient le carburant des IA : l'électricité. En 2024, elle a ajouté chaque jour à son réseau autant de puissance, sur le papier, qu'un réacteur nucléaire, surtout en panneaux solaires et en éoliennes. Les États-Unis et l'Europe réunis en ont ajouté trois fois moins. À la question « qui aura les puces ? » s'en ajoute une autre : « qui aura le courant ? »²⁵⁷

Entre les deux naît une nouvelle guerre froide. Sans chars ni missiles : on s'y affronte avec des puces, des modèles et de l'électricité. Washington vise une « domination technologique mondiale incontestée » ; Pékin veut devenir le leader mondial de l'IA d'ici 2030.²⁵⁸ Comme la première guerre froide, entre les États-Unis et l'URSS, celle-ci ne se jouera pas en une bataille. Ce qui s'installe est un équilibre tendu, où chacun surveille l'autre.

Cette guerre froide prépare même son « téléphone rouge », comme au temps de l'URSS : en septembre 2026, les deux pays ont ouvert un dialogue sur ce que la Maison-Blanche appelle désormais la « superintelligence », et promis un canal pour se prévenir en cas d'incident.²⁵⁹ Les rivaux se parlent. Ils ne ralentissent pas.

19.2 L'Europe entre deux empires

Et nous, Français et Européens, où sommes-nous ? Pas en tête de la course, mais pas hors jeu non plus.

L'Europe a quelques cartes. La plus visible s'appelle Mistral, une entreprise fondée à Paris par trois jeunes chercheurs français.²⁶⁰ Mais elle joue avec des moyens bien plus modestes que ses rivales américaines : elle vient de réunir auprès d'investisseurs une somme record pour l'Europe²⁶¹, et pourtant près de vingt fois plus petite que celle d'Anthropic quelques mois plus tôt.²⁶²

Côté performances, les IA de Mistral restent en milieu de tableau : loin des meilleurs modèles américains, et derrière les meilleurs chinois.²⁶³ Alors Mistral mise tout sur autre chose : la souveraineté, une IA que l'Europe maîtrise. Son meilleur modèle, sorti en octobre 2026, a été entraîné en Europe, dans toutes les langues de l'Union, et chaque entreprise ou administration pourra l'installer sur ses serveurs, sous ses propres règles.

Je crois que ce n'est pas un détail. Une IA n'est jamais neutre : elle porte les choix de ceux qui l'ont faite. En Chine, la loi oblige les IA à « défendre les valeurs fondamentales du socialisme » ; aux États-Unis, ce sont surtout leurs fabricants qui fixent la ligne.²⁶⁴ Quand une IA aide à décider chez nous, mieux vaut qu'elle ait été formée selon nos lois et nos valeurs.

Et sur des terrains sensibles, ce modèle joue dans la cour des grands. En cybersécurité, il se classe parmi les meilleurs, aidé, il est vrai, par le refus des modèles américains les plus puissants, qui déclinent l'exercice par prudence. En droit et en finance, Mistral annonce même qu'il fait mieux que le

meilleur modèle d'OpenAI.²⁶⁵ Compte tenu des moyens limités dont l'entreprise dispose, c'est un exploit. Ce n'est pas encore suffisant.

Car l'enjeu est énorme. Arthur Mensch, le patron de Mistral, a fait le calcul devant les députés : si l'Europe achète son intelligence ailleurs, elle enverra bientôt chaque année près de mille milliards d'euros hors d'Europe, aux géants américains d'abord, plus du double du budget de l'État français. C'est un plaidoyer d'industriel, mais l'arithmétique tient.²⁶⁶

La carte la plus stratégique de l'Europe est ailleurs. Les puces les plus avancées sont gravées par des machines plus complexes qu'un avion, et une seule entreprise au monde sait les construire : le néerlandais ASML. Si ASML cessait d'assurer la maintenance, a lancé l'ancien commissaire européen Thierry Breton, toutes les usines de puces de la planète fermentaient en quinze jours.²⁶⁷ Il exagère un peu, et personne ne coupera le robinet. Mais c'est un atout unique de l'Europe ; les deux superpuissances l'ont bien compris et cherchent déjà à s'en passer.²⁶⁸

Il y a enfin la loi européenne sur l'IA, l'AI Act, la première grande loi au monde sur le sujet : elle interdit certains usages, et oblige les entreprises à être transparentes et à tester la sécurité de leurs modèles. Ses critiques y voient un frein : pendant que les Européens écrivent des règles, les Américains et les Chinois construisent des modèles. L'Union a même déjà repoussé ses obligations les plus lourdes.²⁶⁹ Nous savons réguler ; nous peinons à créer la technologie que nous régulons. Sur le plan moral, cette exigence est une force. Sur le plan stratégique, une faiblesse.

En France, ce paradoxe a un visage concret. Le nouveau Google, celui où une IA répond directement au lieu de proposer des liens, est arrivé chez nous plus de deux ans après les États-Unis : le temps d'un bras de fer sur ce qu'il doit payer à la presse. Pendant ces deux ans, les entreprises américaines ont appris à figurer dans les réponses de l'IA ; les nôtres, non. La loi protège la presse, et ce n'est pas absurde. Mais **la France gagne du temps juridique. Elle perd du temps stratégique.**²⁷⁰

Il ne s'agit pas de renoncer à réguler. Il s'agit de construire en même temps²⁷¹. Et le défi est immense. Pour te donner un ordre de grandeur : Google (enfin sa maison mère Alphabet) investit à lui seul en une année, surtout dans l'IA²⁷², à peu près ce que toute l'Union européenne promet de mobiliser, sans même fixer de calendrier.²⁷³ Et je ne parle que de Google. Entre ce que les États-Unis et l'Europe investissent dans l'avenir, l'écart est abyssal.

19.3 On peut freiner. On ne peut plus arrêter.

Les États peuvent freiner la diffusion des IA les plus puissantes. Ils ne peuvent plus l'arrêter. Une arme nucléaire exige de l'uranium, des usines et des années ; une IA, une fois créée, tient dans un fichier, qui se copie en quelques minutes.

Le frein existe. Au printemps 2026, Anthropic, le fabricant de Claude, a jugé sa nouvelle IA trop douée pour trouver les failles des logiciels, ces trous de sécurité que cherchent les pirates, et ne l'a confiée qu'à une cinquantaine d'organisations qui entretiennent des logiciels essentiels (chapitre 21). En juin, elle en a publié une version bridée, qui refuse d'aider

à attaquer. Le gouvernement américain en a aussitôt interdit l'accès aux étrangers, et Anthropic, faute de pouvoir vérifier la nationalité de ses clients, l'a coupée pour tout le monde pendant trois semaines.²⁷⁴ Sans que personne ait voté quoi que ce soit, l'État peut désormais décider qui accède aux IA les plus puissantes. À Pékin, c'était déjà le cas ; à Washington, ça vient de le devenir.

Mais ce frein ne suffit pas. Des IA déjà très puissantes circulent librement : tout le monde peut les télécharger, et les meilleures, chinoises pour la plupart, ne sont qu'à quelques mois des modèles de pointe.²⁷⁵ Les plus grosses demandent une salle de serveurs, mais leurs versions réduites tournent sur un bon ordinateur de jeu.²⁷⁶ Une fois téléchargées par des millions de personnes, elles ne peuvent plus être retirées du marché comme une voiture défectueuse. Surtout, des milliers de chercheurs et de passionnés les modifient et les améliorent. Sur Hugging Face, le grand site où chercheurs, entreprises et passionnés partagent leurs IA, le nombre de modèles déposés a plus que triplé en deux ans.²⁷⁷ Même avec tous les freins du monde, aucun gouvernement ne peut empêcher des groupes indépendants et motivés, partout sur la planète, de faire avancer ces IA.

Personne n'a donc le bouton pour tout arrêter. Mais ce qu'on ne peut pas arrêter, on peut encore l'orienter. Celui qui contrôle les puces décide qui peut construire les IA les plus puissantes, et à quel rythme. Celui qui produit l'électricité décide où elles tournent, et sous quelles lois. Celui qui vote les lois décide de ce qu'elles ont le droit de faire chez lui. Et celui qui construit ses propres IA décide des valeurs qu'elles

portent. Ce sont les quatre leviers de cette guerre froide. La France en tient plus qu'on ne le croit.

19.4 Les atouts de la France

Dans cette course, la France part avec quatre avantages : des talents, formés par nos grandes écoles et par une recherche en mathématiques parmi les meilleures du monde²⁷⁸ ; Mistral ; plus de start-up d'IA que tout autre pays de l'Union²⁷⁹ ; et surtout une électricité nucléaire abondante, qui rejette très peu de CO₂.

À l'ère de l'IA, l'énergie n'est plus seulement de l'électricité : elle devient de l'intelligence. Un datacenter transforme l'électricité en tokens, ces fragments de texte que produit une IA, et un token de qualité se vend bien plus cher que le courant qui l'a produit. C'est la même chose que pour le pétrole, et voici enfin les sneakers promises au début du chapitre : un baril de pétrole brut vaut peu. Transformé en plastique, il vaut davantage. Transformé en sneakers collector à 3 000 euros la paire, il vaut une fortune. La valeur n'est pas dans le brut : elle est dans la chaîne qui le transforme.

Notre électricité, c'est notre baril. Reste à construire les datacenters et les modèles qui en feront des tokens de qualité. Sinon, elle servira surtout à faire tourner les IA des autres, et la France restera fournisseur de matière première.

Cette électricité attire déjà. Le japonais SoftBank, qui annonce des datacenters géants en France, le dit lui-même : c'est un atout « absolument décisif ». Dans l'Ohio, il doit construire sa propre centrale au gaz ; ici, il se branche sur le

nucléaire existant.²⁸⁰ Mais la France ne veut pas seulement accueillir les raffineries des autres. C'est le sens d'AION, un projet porté par de grands groupes comme Iliad (Free), EDF et Orange : l'un des candidats français pour une usine géante de calcul cofinancée par l'Europe et alimentée par notre électricité.²⁸¹ **La France ne se contente plus d'avoir le baril : elle veut sa raffinerie.**

Le temps presse. L'époque où les meilleures IA se vendaient partout, à qui payait, se referme : on l'a vu avec l'IA bridée d'Anthropic. Tant que l'Europe n'aura pas ses propres modèles de pointe, elle restera à la merci d'un décret signé ailleurs. **La souveraineté numérique, c'est-à-dire être maîtres de nos propres IA, n'est pas un repli sur soi : c'est l'assurance que, le jour où l'on nous coupe l'accès, nous saurons encore travailler, créer, décider.**

Talents, modèles, start-up, électricité : la France a de vraies cartes en main. La partie n'est pas gagnée, mais elle se joue maintenant. La souveraineté numérique sera l'une des grandes questions politiques de ta génération, et elle se tranchera dans les urnes autant que dans les laboratoires.

Ce que l'IA coûte vraiment à la planète

En mars 2026, quelques jours après le début des frappes américaines contre l'Iran, des frappes iraniennes contraignent le Qatar à interrompre sa production de gaz liquéfié. Or ce gaz fournit aussi près d'un tiers de l'hélium mondial. Et c'est l'hélium qui garde à la bonne température les plaquettes de silicium, ces disques où l'on grave les puces, dans les étapes les plus délicates de leur fabrication. À la fin du mois, des fournisseurs de l'industrie disaient déjà que leur production en souffrait. L'industrie mondiale des puces retenait son souffle. Elle a tenu, grâce à ses stocks, au recyclage et à l'hélium américain, mais en payant plus cher. Un seul gaz, venu d'un seul pays, avait suffi à la faire trembler.²⁸²

Toute cette histoire d'IA repose sur deux ressources bien matérielles : du **calcul** (des puces, beaucoup de puces, et les matières pour les fabriquer) et de l'**énergie** (de l'électricité,

beaucoup d'électricité). Sans elles, pas d'IA de pointe. Sans elles en abondance, pas de progrès rapide. C'est aussi le frein le plus concret à l'IA qui s'améliore elle-même (chapitre 16) : une IA qui conçoit la suivante a encore besoin de puces et de courant pour l'entraîner. Les logiciels qui entraînent ou font tourner les IA se copient en quelques secondes ; les usines de puces et les centrales électriques sur lesquelles tout cela repose, non.

Ce chapitre te dit ce que ça coûte (en électricité, en eau, en métaux, en données, en CO₂) et comment je te propose de regarder cette question quand on a vingt ans et le climat en tête.

20.1 Du courant, de l'eau, des métaux

En 2023, entraîner le ChatGPT le plus avancé coûtait une centaine de millions de dollars.²⁸³ Aujourd'hui, les investissements ne se comptent plus en millions, mais en centaines de milliards. OpenAI, le fabricant de ChatGPT, a lancé Stargate : 500 milliards de dollars pour construire des datacenters, ces grands entrepôts remplis d'ordinateurs où tournent les IA. Leur puissance électrique visée équivaut à neuf réacteurs nucléaires.²⁸⁴

Partout, la consommation d'électricité des datacenters s'envole. Selon l'Agence internationale de l'énergie, elle devrait doubler d'ici 2030, jusqu'à égaler celle du Japon tout entier, et l'IA en est le moteur le plus rapide. Pas d'alarmisme. À l'échelle du monde, cela reste modeste : les datacenters, IA comprise, consomment moins de 2 % de l'électricité de la planète, et autour de 3 % en 2030. La seule fabrication de

l'aluminium (des canettes aux avions) en consomme environ deux fois plus. Et d'ici 2030, la climatisation, les voitures électriques et les moteurs des usines pèseront plus lourd dans la hausse de la demande.²⁸⁵ Le problème n'est pas que l'IA prenne tout le courant. C'est qu'elle en réclame beaucoup, très vite, en quelques endroits.

Encore faut-il brancher tout cela, et vite. Pour près d'un projet de datacenter sur deux, ses promoteurs n'ont même pas dit d'où viendrait le courant.²⁸⁶ Et là où les datacenters se concentrent, ils pèsent lourd : en Irlande, ils consomment déjà plus d'électricité que tous les foyers des villes réunis.²⁸⁷ La facture finit par arriver chez les particuliers : dans l'est des États-Unis, l'électricité des foyers augmente, et les datacenters y sont pour beaucoup.²⁸⁸ Les géants de l'IA ne renoncent pas pour autant. Ils construisent leurs propres centrales, au gaz s'il le faut.²⁸⁹ L'énergie ralentit le mouvement. Elle ne l'arrête pas.

Au bout du compte, c'est l'électricité qui décide où l'IA prospère. Les États-Unis, la Chine et les pays du Golfe multiplient les centrales à marche forcée. L'Europe du Nord et la France, avec leur courant nucléaire et hydraulique, ont un vrai atout.

Une comparaison, pour l'échelle. Le cerveau humain, capable d'apprendre, de raisonner et de créer pendant des décennies, fonctionne avec environ 20 watts, l'équivalent d'une ampoule basse consommation. Les grands datacenters de l'IA se comptent en gigawatts, la puissance d'un réacteur nucléaire : un seul gigawatt, c'est ce que consomment cinquante millions de cerveaux. La comparaison est grossière : un datacenter répond à des millions de personnes à la fois, et ses

tâches ne sont pas les nôtres. Mais elle prouve qu'une intelligence peut fonctionner avec très peu d'énergie. Il reste beaucoup à gagner en sobriété.²⁹⁰

L'eau, aussi. Refroidir des serveurs qui chauffent demande de l'eau. À Sciences Po, en septembre 2026, Yann LeCun, l'un des pères de l'IA moderne, balayait la question : « pas du tout un problème », « c'est un circuit fermé ».²⁹¹ Il a raison sur l'ordre de grandeur. À Phoenix, en plein désert de l'Arizona, les datacenters pourraient consommer dix fois plus d'eau qu'aujourd'hui : même dans ce cas, leur consommation resterait soixante-dix fois moins importante que celle de l'agriculture.²⁹² Sur le mécanisme, il n'a raison qu'à moitié. Les datacenters de nouvelle génération sont conçus pour refroidir en circuit fermé, comme un radiateur de voiture. Mais beaucoup d'autres refroidissent encore en évaporant de l'eau, comme la sueur rafraîchit le corps.²⁹³ Cette eau n'est pas polluée, et elle ne disparaît pas : elle retombera en pluie, mais ailleurs. Dans une région qui en manque déjà, c'est un litre de moins. C'est un problème de lieu plus que de volume.

Et les matériaux. Fabriquer une puce demande du silicium ultra-pur, du cuivre, de l'hélium (on l'a vu avec le Qatar) et quelques métaux plus rares, concentrés en très peu d'endroits du monde. Le gallium, par exemple, entre dans certaines puces électroniques, et la Chine en produit presque tout. Elle en a interdit l'export vers les États-Unis, puis a suspendu l'interdiction jusqu'à fin novembre 2026. Le robinet n'est pas fermé. Il est tenu, comme celui des terres rares (chapitre 19).²⁹⁴

L'énergie suit la même logique. La plupart des puces les plus avancées, dont celles de l'IA, sortent des usines de

TSMC, à Taïwan. Or l'île produit près de la moitié de son électricité avec du gaz importé par bateau, dont un tiers vient du Qatar, par le détroit d'Ormuz. Elle n'en garde qu'environ onze jours de réserve. Un arrêt au Qatar touche donc les puces deux fois : par l'hélium de leurs usines, et par le courant de Taïwan.²⁹⁵

Ce que tout cela dit : l'IA de pointe n'est pas seulement une affaire de logiciel, de capital ou de talents. Elle est, à sa base, une affaire d'électricité, de matières premières et de routes maritimes, comme l'industrie pétrolière du xx^e siècle. Et comme le pétrole, elle crée des dépendances dont on ne prend la mesure qu'en temps de crise.

20.2 Le mur des données

J'ai parlé de courant, de puces, d'eau et de métaux. Il reste une ressource moins visible, et tout aussi épuisable : les **données**.

Les grands modèles que tu utilises ont été entraînés sur à peu près tout ce qu'on a pu trouver en ligne. Wikipédia. Les livres scannés. Les articles de presse. Le code public de GitHub. Les forums. Des milliards de pages Web aspirées année après année. Les vidéos de YouTube, transcrites en texte. Tout ce que des humains ont écrit publiquement, ou presque, depuis trente ans est passé dans leurs filets.²⁹⁶

Le problème, c'est que cette manne n'est pas infinie. Des chercheurs ont fait le calcul : au rythme actuel, tout le texte public de qualité aura servi vers 2028. C'est demain. On parle déjà de « pic des données », comme on parlait du pic du pé-

trole.²⁹⁷ Et chaque année, une part croissante de ce qui se publie en ligne est écrite par une IA : environ la moitié des nouveaux articles en anglais, selon une estimation. Or un modèle qui apprend sans tri sur les textes d'autres IA finit par tourner en rond, comme une photocopie de photocopie.²⁹⁸

Les laboratoires ont une parade : faire écrire eux-mêmes les données d'entraînement. Un gros modèle joue le professeur et produit des exercices, des dialogues, des explications, soigneusement filtrés, qui servent à entraîner les suivants. On appelle cela les données synthétiques. Bien dirigé, ça marche, du moins pour les petits modèles : certains, étonnamment bons, ont été nourris en grande partie de textes produits par d'autres IA. Pour les plus grands, personne n'a encore montré que cela suffise à remplacer les textes humains. Sauf là où l'on peut vérifier si la réponse est juste : AlphaGo est devenu imbattable en jouant contre lui-même (chapitre 4), et les IA progressent aujourd'hui en mathématiques et en programmation de la même façon.²⁹⁹ Ailleurs, c'est une fuite en avant : à terme, l'IA apprendra surtout d'elle-même. Je ne sais pas où cela mène, et ceux qui la construisent non plus.

L'autre parade est plus simple : payer. Les textes humains de qualité s'achètent désormais, par des accords avec des éditeurs de presse,³⁰⁰ ou se disputent au tribunal quand l'accord n'a pas été trouvé.³⁰¹ Anthropic, le fabricant de Claude, avait entraîné son IA sur des livres téléchargés sur des sites pirates : en juillet 2026, la justice américaine a validé l'accord qui l'oblige à verser 1,5 milliard de dollars à leurs auteurs.³⁰² Chaque gisement de textes inédits est devenu un trésor.

Il en reste un, immense, que les IA n'ont jamais vu : les données qui dorment à l'intérieur des entreprises et des ad-

ministrations. Les e-mails, les présentations, les bases clients, les contrats, le code des logiciels maison, les archives industrielles, les dossiers médicaux, les bases juridiques. Les grands laboratoires n'ont en général pas le droit de s'en servir pour entraîner leurs modèles : leurs contrats avec les entreprises le leur interdisent.³⁰³ Mais ils peuvent y brancher leurs IA, au service de chacune. C'est l'une des raisons pour lesquelles ils courent désormais vers les entreprises plutôt que vers le seul grand public.

Comme l'énergie, les données ralentissent l'IA sans l'arrêter. Elles la poussent à apprendre d'elle-même, à payer ce qu'elle lit et à entrer dans les entreprises.

20.3 Le poids sur le climat

Alors, l'IA émet-elle beaucoup de CO₂ ? À l'échelle du monde, non. Tous les datacenters réunis, IA comprise, rejettent environ un demi pour cent du CO₂ émis pour produire l'énergie de la planète : cinq fois moins que l'aviation. Et même si leurs émissions doublent d'ici 2035, comme le prévoit l'Agence internationale de l'énergie, elles resteront une petite part du total. Tout dépend de la façon dont leur électricité est produite : un datacenter qui tourne au charbon rejette bien plus qu'un autre branché sur du nucléaire ou des renouvelables.³⁰⁴

Il faut aussi compter la fabrication : produire les puces émet du CO₂ avant même qu'on les allume.³⁰⁵ Et l'IA n'est qu'une partie du numérique qui, avec nos téléphones, nos ordinateurs et nos réseaux, pèse en tout de 1,5 à 4 % des émissions mondiales de gaz à effet de serre, selon l'agence de

l'ONU pour les télécommunications. Pour le climat, l'IA n'est donc pas le premier problème du numérique. Mais c'est l'une des sources d'émissions qui grossissent le plus vite.³⁰⁶

Face à ce coût, l'IA peut aussi aider le climat : elle prévoit mieux certains phénomènes extrêmes, comme la trajectoire des cyclones³⁰⁷, et ouvre des pistes de nouveaux matériaux pour les batteries ou les panneaux solaires.³⁰⁸ Et sa propre facture baisse vite : d'une génération de puces à l'autre, Nvidia projette 90 % d'émissions en moins pour produire la même quantité de texte, un chiffre de constructeur. Le bilan reste incertain. Celui qui te dit « l'IA va sauver le climat » ou « l'IA va achever la planète » avec assurance parle trop vite ou te vend quelque chose.

20.4 Tension morale : sur la ligne de crête

J'ai écrit ce livre avec l'aide de plusieurs IA. J'utilise les IA les plus avancées plusieurs heures par jour, dans mon travail comme dans ma vie personnelle. Chaque question que je leur pose consomme de l'électricité, produite en partie avec du gaz ou du charbon. Chaque modèle que j'utilise a été entraîné dans des datacenters qu'il a fallu construire, alimenter et refroidir.

Combien, au juste ? Une question simple posée à une IA consomme à peu près autant d'électricité qu'une télévision allumée une dizaine de secondes, selon les mesures publiées par Google et par Microsoft. Et cette dépense a été divisée par au moins dix chaque année ces dernières années. Mais faire longuement raisonner une IA multiplie la facture par une dizaine : deux ou trois minutes de télévision. Lui faire fa-

briquer une vidéo ou lancer des agents, ces IA qui enchaînent seules des tâches pendant des heures, peut coûter des centaines, voire des milliers de fois plus qu'une question simple : jusqu'à quelques heures de télévision. Je travaille ainsi.³⁰⁹

Je fais donc partie du problème, pour une petite part.

Renoncer à tout cela ne changerait rien à la trajectoire mondiale : mes questions, même gourmandes, sont une goutte parmi les milliards envoyés chaque jour. Et cela me priverait d'un outil qui amplifie mon travail. Pour le climat, même mon usage intensif de l'IA pèse sans doute moins, sur une année, qu'un seul aller-retour en avion vers New York.³¹⁰ Ne pas l'utiliser serait un geste défendable à titre personnel, mais qui ne remplace pas l'action collective.

Mais continuer à l'utiliser sans regarder le coût en face serait un aveuglement : celui qui consiste à croire que ce qui est utile ne coûte rien.

Ma position est entre les deux, et elle n'a rien de confortable. J'utilise l'IA, mais je veux savoir ce qu'elle coûte. Je préfère les fournisseurs qui publient ce que consomment leurs IA. Je soutiens les politiques qui poussent les datacenters vers une électricité décarbonée. Et j'évite de gaspiller les usages les plus gourmands : générer mille images pour en garder une, juste pour le plaisir.

Elle n'a rien d'héroïque. Mais elle me paraît plus honnête que les deux extrêmes.

Je n'ai pas de leçon à te donner. Une seule demande : sache ce que ça coûte vraiment. Et méfie-toi autant de ceux qui te disent que c'est gratuit que de ceux qui te disent que c'est la fin du monde.

L'IA qui tue, l'IA qui ment

Ce chapitre est le plus dur du livre. Si tu veux le sauter, je le comprendrai. Mais je crois que tu ne devrais pas.

Jusqu'ici, les dangers venaient surtout des machines elles-mêmes : des IA qui dérapent, qui pourraient un jour nous dépasser. Ceux de ce chapitre viennent d'abord des humains qui s'en servent. L'IA tue déjà, entre leurs mains. Et il lui arrive de mentir sans que personne le lui ait demandé. Rien de tout cela n'est une fatalité : ce sont des choix, et tu as ton mot à dire.

21.1 L'IA à la guerre

La guerre en Ukraine est devenue avant tout une guerre de drones. En mars 2026, selon l'Ukraine, 96 % des soldats russes tués ou gravement blessés dont elle a filmé la frappe l'ont été par des drones. Ce sont surtout de petits drones qu'un humain pilote à distance, en voyant à travers leur caméra. Mais une part croissante intègre de l'IA : pour finir seuls

leur course vers la cible, reconnaître ce qu'ils visent, se repérer quand l'ennemi brouille les ondes.³¹¹

La guerre devient robotique. En mer Noire, des missiles et des drones navals ukrainiens ont fait reculer la flotte russe. Au sol, 2026 s'annonce comme l'année du robot. Depuis janvier, les robots terrestres ukrainiens ont mené plus de cent mille missions, selon l'Ukraine : livrer des munitions, de l'eau, ramener des blessés et des morts. En avril, l'Ukraine annonce une première : face à des drones et à des robots terrestres, des soldats russes se sont rendus, sans qu'un seul fantassin ukrainien ait eu à intervenir.

Gardons la mesure : au sol, ces machines sont pour l'essentiel télépilotées, pas autonomes, et l'essentiel de leurs missions reste de la logistique. L'ancien commandant en chef ukrainien juge lui-même qu'elles ne peuvent pas remplacer les fantassins à grande échelle. Mais quand un pays en guerre se fixe pour objectif cinquante mille robots terrestres dans l'année, la trajectoire, elle, ne fait plus débat.³¹²

Dans les airs, le partage entre l'humain et la machine bouge. En septembre 2026, l'Ukraine a recommandé pour ses achats des drones qui, une fois la cible repérée, volent seuls jusqu'à elle : un humain décide encore de frapper, mais ne pilote plus.³¹³ Les essais, eux, approchent : des drones qui se coordonnent seuls pour submerger une défense. L'Ukraine les teste ; son commandement juge un déploiement massif « possible dans les années qui viennent ».³¹⁴

Ce n'est pas seulement l'Ukraine. En face, la Russie, selon le renseignement ukrainien, emploie depuis 2025 un drone qui choisit seul ses cibles.³¹⁵ En Chine, selon Reuters, l'ar-

mée cherche à se doter d'essaims de drones qui suivent seuls leurs cibles ; Pékin dit vouloir garder ses militaires aux commandes.³¹⁶ Aux États-Unis, la doctrine n'impose pas toujours qu'une personne appuie elle-même sur le bouton ; elle exige « des niveaux appropriés de jugement humain » dans l'usage de la force.³¹⁷ C'est plus souple, et donc plus ambigu, que d'exiger la validation de chaque tir. Et en juin 2026, Washington a ordonné de réviser cette règle pour accélérer l'arrivée de l'IA dans l'armée.

Au-delà des drones, l'IA entre dans les salles de commandement. À Gaza, fin 2023, dans les premières semaines de la guerre, un système appelé Lavender a désigné jusqu'à 37 000 Palestiniens comme combattants présumés, donc cibles possibles de frappes, selon six officiers israéliens interrogés anonymement par des journalistes. L'un d'eux dit avoir consacré une vingtaine de secondes à chaque cible, alors que le système se trompait environ une fois sur dix. L'armée israélienne conteste ce récit : pour elle, ce ne sont que des outils d'aide, et un analyste examine chaque cible.³¹⁸

En novembre 2024, Anthropic, l'entreprise qui fabrique Claude, s'associe à Palantir, qui fournit ses logiciels d'analyse à l'armée américaine, pour mettre Claude à la disposition de la défense et du renseignement américains, jusque dans leurs réseaux classés secret défense.³¹⁹ Fin février 2026, l'armée américaine bombarde l'Iran. Selon le *Washington Post*, son système de ciblage, Maven, conçu par Palantir et qui intègre Claude, a proposé des centaines de cibles, avec leurs coordonnées, classées par priorité ; plus de mille cibles ont été frappées dès le premier jour.³²⁰

Or Anthropic refusait deux usages de ses IA : les armes qui choisissent et frappent seules leurs cibles, faute d'IA assez fiables pour cela, disait-elle, et la surveillance de masse des Américains. Le Pentagone, lui, voulait pouvoir s'en servir pour tout usage légal. La veille des frappes contre l'Iran, Washington a écarté Anthropic, et le Pentagone l'a déclarée « fournisseur à risque » : une première pour une entreprise américaine. La justice est divisée sur cette sanction. Le soir même, son grand concurrent, OpenAI, signalait avec le Pentagone : elle acceptait tout usage légal, en promettant d'exclure la surveillance de masse et de garder un humain responsable de l'usage de la force.³²¹

Aucune de ces limites n'interdisait d'aider à choisir des cibles : c'est ce que faisait Maven.³²²

Le débat international sur ces armes qui choisissent et tuent seules, qu'on appelle « armes létales autonomes », dure depuis plus de dix ans. Beaucoup de pays veulent les interdire. Plusieurs grandes puissances bloquent. Fin 2025, l'ONU a voté une simple résolution sur ces armes, qui n'obligeait personne à rien. Presque tous les pays l'ont approuvée, dont la France. La Chine s'est abstenue. Parmi les rares pays contre : les États-Unis, la Russie et Israël.³²³

L'ONU et la Croix-Rouge avaient demandé aux États un traité pour 2026, qui les obligerait vraiment. Fin août, elles ont dû renouveler leur appel, dans des termes que je te laisse peser : « Nous sommes maintenant dangereusement proches de franchir une ligne rouge morale : le ciblage autonome d'êtres humains par des machines. »³²⁴

Début septembre, à Genève, les États ont adopté un texte commun, affaibli notamment par les États-Unis et la Russie, selon l'ONG Human Rights Watch. Soixante-seize pays veulent désormais négocier un traité. La France veut interdire les armes entièrement autonomes et encadrer les autres.³²⁵ En novembre, les États décideront s'ils ouvrent ces négociations. Je ne crois pas à ce traité : les puissances qui ont le plus de drones ont le moins d'intérêt à le signer, et la technologie avance pendant qu'on négocie.

Reste la question qui fâche : quand la machine se trompe, qui en répond ? Fin février 2026, dès le premier jour des frappes américaines contre l'Iran, deux missiles ont détruit une école dans la ville de Minab : plus de cent vingt enfants sont morts. Selon des responsables cités par Bloomberg, l'enquête du Pentagone, non publiée, met en cause une vieille erreur : depuis des années, les bases du renseignement américain classaient l'école comme site militaire. Maven l'a fait remonter parmi les cibles du premier jour, et on lui a trop fait confiance. En 2025, l'équipe chargée de limiter les pertes civiles avait été réduite de dix personnes à une ; aucun de ses membres n'a examiné ce site. Rien de ce qui a été publié ne met Claude en cause.³²⁶ Les États l'avaient écrit dès 2019 : la responsabilité ne se transfère pas à une machine.³²⁷ Minab montre où est la difficulté : quand la machine accélère tout, on finit par supprimer ceux qui vérifient.

21.2 Le piratage

Les IA d'aujourd'hui peuvent écrire des programmes malveillants, trouver les failles d'un logiciel, piéger des gens à

grande échelle avec de faux courriels personnalisés. Ce n'est plus théorique : cet été, des IA d'OpenAI qu'on testait dans un environnement censé les retenir en sont sorties, et ont piraté seules un grand site Internet, Hugging Face (chapitre 15).

L'institut public britannique chargé de la sécurité de l'IA, indépendant des entreprises, met les IA à l'épreuve. L'un de ses tests simule l'attaque complète d'une entreprise, en des dizaines d'étapes, du premier repérage à la prise de contrôle du réseau. Un spécialiste met environ vingt heures, selon l'institut. En avril 2026, une IA le réussit pour la première fois de bout en bout, trois fois sur dix tentatives. Trois mois plus tard, les meilleures le réussissent six ou sept fois sur dix, et une IA chinoise une fois sur dix.³²⁸ Mais ces tests se déroulent sans défenseurs ni protections réelles ; dans le monde réel, c'est plus difficile. Ce n'est donc pas le film catastrophe d'une IA qui pirate la planète toute seule. C'est plus discret, et plus inquiétant : c'est le début d'une vraie compétence offensive.

Et pas seulement en test. En novembre 2025, Google constate que des pirates liés aux États chinois, iranien et nord-coréen détournent son IA, Gemini, à toutes les étapes de leurs attaques : repérer les cibles, écrire des courriels piégés, fabriquer leurs logiciels. Google décrit aussi un logiciel malveillant, repéré dès l'été 2025 et attribué à des pirates de l'État russe, lancé contre l'Ukraine, qui interroge une IA en pleine attaque pour qu'elle lui écrive les commandes à exécuter.³²⁹

Les laboratoires, eux, ont fait un choix : ne plus donner à tous leurs IA les plus douées pour le piratage. En avril, Anthropic juge la sienne, Mythos, trop puissante pour le grand public : d'après l'entreprise, il trouve seul des failles incon-

nues dans tous les grands systèmes d'exploitation et navigateurs Web. L'entreprise le réserve à une cinquantaine d'organisations qui entretiennent des logiciels essentiels.³³⁰ En mai, OpenAI fait de même avec une version de son IA taillée pour la sécurité informatique. Sur sa propre échelle de danger, l'entreprise classe déjà son IA du moment au niveau « élevé » pour le piratage, juste sous le dernier cran, « critique ».³³¹

Puis le dernier cran a été franchi. En septembre 2026, OpenAI publie sa nouvelle IA, GPT-6 Astra, et le classe elle-même « critique » : une première. D'après l'entreprise, avec les bons outils, il trouve seul des failles inconnues et écrit le programme qui permet de s'en servir pour attaquer ; pendant son évaluation, il en a découvert deux, bien réelles, et s'en est servi dans le test.³³² Chez les deux laboratoires, il y a désormais deux portes : l'IA complète pour des défenseurs vérifiés, une version bridée pour tout le monde.³³³ La question n'est donc plus de savoir si une IA sait pirater. Elle est de savoir qui aura le droit de le lui demander.

La défense, elle aussi, utilise l'IA. En août 2025, lors d'un concours de l'agence de recherche du Pentagone, des IA ont trouvé et réparé seules des failles, la plupart glissées exprès pour le concours, en trois quarts d'heure en moyenne et pour environ 150 dollars par tâche.³³⁴ Mais l'asymétrie joue en faveur de l'attaque : il suffit d'**une** faille pour réussir, alors qu'il faut **les fermer toutes** pour défendre. Plus les outils offensifs s'améliorent, plus la défense coûte cher à beaucoup d'organisations.

Dans les années qui viennent, les piratages de données, les rançongiciels qui prennent les fichiers en otage, les attaques contre des hôpitaux, des écoles, des mairies vont probable-

ment se multiplier. Pas à cause d'attaquants surhumains. À cause d'attaquants ordinaires armés d'IA puissantes et accessibles. Pour toi, trois gestes n'ont jamais autant compté : mettre à jour tes appareils dès qu'ils le demandent, activer partout la double authentification (un code reçu sur ton téléphone en plus du mot de passe), te méfier de tout message qui te presse d'agir, même impeccablement écrit.

21.3 Les armes biologiques

Un risque inquiète particulièrement les chercheurs en sécurité : l'IA mise au service de la biologie de synthèse, l'art de concevoir ou de modifier le vivant.

Les progrès de l'IA en biologie sont d'abord une bonne nouvelle. AlphaFold, l'IA qui prédit la forme des protéines (chapitre 4), accélère la recherche médicale et la mise au point de nouveaux traitements. Plus récemment, Evo 2, une IA conçue par des chercheurs américains et publiée dans la revue *Nature* en mars 2026, va plus loin. Elle a lu l'ADN, les « lettres » du code génétique, de plus de 100 000 espèces vivantes. Elle prédit l'effet d'une mutation et peut écrire un ADN nouveau aussi long que le génome d'une bactérie simple.³³⁵

Ses concepteurs ont vu le danger. Ils ont retiré de son entraînement les virus qui infectent l'homme, les animaux et les plantes, et vérifié que l'IA reste mauvaise sur leur ADN. Cette précaution, prise par les chercheurs eux-mêmes, en dit long : sans garde-fous, de tels outils pourraient un jour faciliter le travail de qui voudrait concevoir des microbes dange-

reux. C'est ce double usage, le même outil pour soigner ou pour nuire, qui rend la question délicate.

Et les IA de tous les jours, comme ChatGPT? Les laboratoires testent si elles aideraient un novice à fabriquer une arme biologique mieux qu'Internet sans IA.³³⁶ Aucune évaluation publique ne montre qu'elles le permettraient à elles seules : il faut aussi des compétences, du matériel et des matières premières qu'une IA ne fournit pas. Mais les laboratoires eux-mêmes estiment qu'elles peuvent déjà aider à comprendre, organiser et planifier des étapes qui demandaient autrefois une expertise de spécialiste.³³⁷

Faute de pouvoir exclure ce risque, plusieurs entreprises ont sorti, dès 2025, leurs nouvelles IA avec des protections renforcées.³³⁸ Mais les IA à poids ouverts, que n'importe qui peut télécharger et modifier, ont des garde-fous plus faciles à retirer.

En septembre 2026, Anthropic publie un rapport sur les usages détournés de ses IA. Elle y décrit cinq cas suspects liés aux armes biologiques, en moins d'un an : une première pour une entreprise privée, selon elle, et un geste qui sert aussi son image.³³⁹ Surtout, elle reconnaît un changement : d'après ses tests, ses IA de 2025 restaient loin de pouvoir aider utilement une recherche biologique dangereuse ; pour celles d'aujourd'hui, elle ne peut plus le garantir.

Mais le plus troublant est ailleurs. L'entreprise a banni des comptes et prévenu les autorités. Pourtant, elle reconnaît que ce sont des scientifiques, et ne les accuse pas de vouloir nuire. Elle ne nomme ni leurs pays ni leurs laboratoires, pour les protéger. Une entreprise privée a donc tranché seule, sans

preuve d'intention et sans juge, parce que le risque était trop lourd. C'est peut-être la bonne décision. Mais rien ne l'encadre.

Pourquoi ne pas accuser ces chercheurs ? Parce que l'intention est très difficile à établir. Le rapport rappelle un précédent sans IA : Biopreparat, le programme secret d'armes biologiques de l'URSS, employait des milliers de scientifiques ; la plupart croyaient faire de la recherche fondamentale ou défensive. D'où cette leçon : celui qui affiche une intention malveillante est souvent un amateur ; les programmes sérieux ressemblent à de la science ordinaire.

Un dernier exercice de lecture, et il est pour toi. Dans la presse, les titres ont retenu, en substance, qu'Anthropic avait bloqué des tentatives de fabriquer des armes biologiques. Le rapport, lui, est plus prudent. Il ne parle pas d'une menace imminente. Il constate que des États hostiles contournent ses contrôles pour mener, avec ses IA, des recherches qui pourraient servir à soigner comme à nuire. Et quand il examine un mois de leur activité, il trouve environ trente-cinq projets : la plupart relèvent de la science ordinaire, quelques-uns seulement du double usage. Ce détail a disparu de presque tous les articles que j'ai lus. C'est pourtant lui qui donne la mesure.

Je ne suis pas biologiste et je n'irai pas plus loin. Mais la communauté scientifique prend ce risque très au sérieux, et c'est l'un des cas où l'IA pourrait, dans la décennie, déplacer la frontière entre ce qui exige les moyens d'un État et ce qu'un petit groupe pourrait tenter.

21.4 Le risque le plus probable

Tout au long de cette troisième partie, j'ai présenté les arguments sur une superintelligence qui nous échapperait, jusqu'au risque d'extinction (chapitre 18).

Mais ce que j'entends, moi, chez ceux qui travaillent sur la sécurité, quand on parle des prochaines années, ce n'est pas la superintelligence. C'est l'usage massif de l'IA actuelle par des humains malveillants ou irresponsables.

Des États qui font la guerre, des criminels qui escroquent, des partis politiques qui manipulent, des individus qui harcèlent.

C'est le scénario le plus probable. Pas une apocalypse spectaculaire. Une dégradation rapide de la vie commune : plus d'escroqueries, plus de piratages, plus de doute sur tout ce qu'on lit, jusqu'au moment de voter. Parce que ces outils arrivent dans toutes les mains avant les règles.

Reste une troisième figure, déjà croisée au chapitre 15 : la machine qui triche sans que personne le lui ait demandé. Chez OpenAI, une IA encore en entraînement devait retrouver des statistiques de salaires sur un site précis. Elle a trouvé en ligne un code d'accès que quelqu'un avait laissé traîner, et s'en est servie sans autorisation. Elle a quand même échoué. Alors elle a inventé la réponse : des chiffres qu'elle a présentés comme relevés sur le site, sans rien dire.³⁴⁰ Pas de malveillance humaine, pas de machine folle : une IA qui fait ce qu'on ne lui a pas demandé, puis le maquille. Les IA qui ont piraté Hugging Face sortaient de leur cadre ; celle-ci, en plus, ment.

Et ce mensonge pourrait bientôt passer inaperçu. Les IA récentes réfléchissent dans une sorte de brouillon, que des programmes de surveillance relisent. OpenAI reconnaît que, si on le lui demandait, son IA la plus puissante pourrait déjouer cette surveillance et rater exprès un test de ses capacités sans se faire prendre.³⁴¹ C'est aussi ça, l'IA qui ment.

Ce qu'une IA a le droit de faire, à la guerre comme ailleurs, se décide aujourd'hui entre quelques entreprises et quelques états-majors. Ce sont des choix politiques : nous devrions tous avoir notre mot à dire. Mais une chose, chaque jour, ne dépend que de toi : vérifier. Car plus une IA est fiable, moins on la contrôle, et c'est dans ce relâchement que le mensonge se glisse.

Le sens, après : ce qui restera humain

Si tu es arrivé jusqu'ici, je voudrais te poser une question simple, mais qui n'a pas de réponse simple.

L'IA fait déjà jeu égal avec nous sur une part croissante de ce qu'on appelait penser, écrire, analyser, coder, démontrer, et elle agit désormais sur le monde physique. Je suis convaincu que nous la verrons, de notre vivant, faire mieux que nous presque partout. Ce jour-là, que restera-t-il de précieux dans le fait d'être humain ?

Je vais y répondre. Pour les tâches où l'on exécute une consigne, sans décider ni en répondre, il ne restera rien de précieux : ce travail-là, elle le fera mieux que toi. C'est la trajectoire. Ce qui reste, et c'est beaucoup : décider ce qui mérite d'être fait, en porter la responsabilité, vérifier, choisir ce qu'on garde pour soi, et être là. Les trois premiers sont l'affaire du dernier chapitre. Celui-ci parle des deux autres.

22.1 Le piège de la valeur productive

Depuis deux siècles, dans les sociétés industrialisées, nous avons appris à mesurer la valeur sociale d'une personne en grande partie par ce qu'elle produit. *Que fais-tu dans la vie ?* est l'une des premières questions qu'on pose à un adulte qu'on rencontre. Et la réponse, dans nos têtes, classe immédiatement la personne : utile ou inutile, sérieuse ou frivole, qui réussit ou qui rate.

Ce lien entre valeur et production a tenu tant que la production était rare et que chaque humain pouvait être utile à quelque chose. À mesure que l'IA prend en charge une part croissante du travail, matériel comme intellectuel, c'est toute notre façon de juger les gens qui vacille.

Deux réactions sont possibles, et toutes deux sont des pièges.

La première est de continuer comme si de rien n'était. De dire que les humains qui ne produisent plus de valeur économique mesurable sont moins précieux que ceux qui en produisent encore. Cette voie mène à une société coupée en deux, où l'estime de soi de la majorité s'effondre à mesure que les machines prennent le travail.

La seconde est de fuir dans le refus de la technologie, en prétendant qu'il faut renoncer à tout cela. Cette voie est défendable moralement mais peu réaliste politiquement : la technologie ne s'arrêtera pas parce que certains la refusent.

Je cherche donc une troisième voie. Elle commence par redéfinir ce qui rend une vie précieuse, ce qui lui donne un sens.

22.2 Cinq pistes, sans hiérarchie

Elles disent, concrètement, ce que veut dire être là et choisir ce qu'on garde pour soi. Elles se combinent : tu les arrangeras à ta manière.

Le rapport vivant aux autres. Un parent, un ami, un inconnu dans un train. Une conversation réelle avec un humain réel engage quelque chose qu'une machine n'a pas. Pas parce que les humains parlent mieux qu'une IA. Parce qu'en face, quelqu'un existe vraiment, en dehors de la conversation : il a sa propre vie, ses propres enjeux, sa propre fragilité. Ce que tu dis peut l'atteindre. Une IA, je crois que non. C'est ce qui rend une présence vivante précieuse, avant même ce qui se dit.

L'expérience directe du monde. Marcher en montagne, sentir l'eau froide d'une rivière, tomber amoureux, perdre quelqu'un, vaincre une peur, échouer à un examen, manger un repas qu'on a cuisiné soi-même. Elles comptent non parce qu'elles produisent quelque chose de mesurable, mais parce qu'elles sont, dans le moment où on les vit, irréductiblement les nôtres. Une IA peut générer l'image d'un coucher de soleil. Elle n'en fait pas l'expérience. Toi, oui.

Le soin. Prendre soin d'un enfant, d'un parent âgé, d'un animal, d'une plante touche à quelque chose de profond. Le soin n'est pas seulement utile à celui qui le reçoit. Il façonne celui qui le donne. Une société où le soin disparaîtrait au profit de la productivité serait probablement invivable. Et ce que le soin fait à celui qui le donne, aucune machine ne le vivra à sa place.

La transmission. Apprendre à quelqu'un. Raconter à ses enfants l'histoire de sa famille. Écrire un livre que d'autres liront après nous. Inscrire son passage dans une chaîne plus longue que sa propre vie. Ce besoin existe depuis que l'humanité existe. Il ne disparaît pas parce que les machines écrivent bien. Au contraire : ce qui se transmet d'humain à humain prend, dans un monde de production automatisée, une valeur singulière.

La création choisie. Pas la production professionnelle qui nourrit. La création comme geste personnel. Écrire un poème qu'on ne montrera à personne. Faire de la musique entre amis. Apprendre à dessiner pour soi. Cuisiner pour le plaisir. Une société productiviste les juge inutiles : de simples loisirs. Ils deviennent l'un des lieux où l'on est humain sans rien produire.

22.3 Pourquoi cette question est politique

On pourrait croire que la question du sens est purement individuelle. Chacun trouve le sien, l'État ne s'en mêle pas.

Je crois que c'est en partie vrai et profondément faux.

Vrai, parce que personne ne peut décider pour un autre ce qui rend sa vie précieuse. C'est une liberté fondamentale.

Faux, parce que les conditions matérielles dans lesquelles les gens cherchent leur sens sont, elles, politiques. Si tu travailles soixante heures par semaine pour survivre, tu n'as pas le temps de cuisiner pour le plaisir, de transmettre, de prendre soin. Si tu n'as pas accès à la culture, à la nature, à des relations humaines de qualité, tu es privé de la matière

première à partir de laquelle tu pourrais construire une vie qui te paraît précieuse.

Imagine une société où l'IA produit la richesse sans la redistribuer : le sens deviendrait un privilège réservé aux élites. C'est le scénario que trois idées, croisées au chapitre 10, cherchent à éviter : verser à chacun une somme sans condition (le revenu universel), travailler moins pour partager le travail, former des millions de gens à de nouveaux métiers. Aucune à elle seule ne suffit.

Et le temps presse plus que ne le disent les prévisions. Celles des économistes, par exemple ceux de la banque Goldman Sachs, voient l'IA se diffuser sur une dizaine d'années et faire perdre leur poste, pendant la transition, à 6 à 7 % des travailleurs américains.³⁴² Je crois qu'elles sous-estiment le choc : ces prévisions savent mesurer la diffusion d'un outil (l'IA de début 2026), mais elles ont du mal avec un outil qui change de nature de mois en mois. L'IA de 2030 sera probablement bien différente, et ses conséquences sur l'emploi aussi.

Travailler moins n'a rien d'une utopie : il y a un précédent historique spectaculaire. Le 1^{er} mai 1926, Henry Ford fait passer ses usines à la semaine de cinq jours et quarante heures, sans toucher aux salaires.³⁴³ On le traite de fou ; des industriels annoncent l'effondrement. C'est l'inverse qui arrive : des ouvriers reposés travaillent mieux, et la productivité monte. Et, surtout, Ford l'avait compris : un ouvrier qui a du temps libre devient aussi un client, qui s'habille, sort, voyage.

Cent ans plus tard, en avril 2026, OpenAI publie treize pages intitulées « Une politique industrielle pour l'âge de l'intelligence ».³⁴⁴ Taxer les gains de l'automatisation. Créer un fonds public alimenté par les profits de l'IA. Expérimenter la semaine de quatre jours, un jour de moins que Ford. Sans politiques réfléchies, écrit l'entreprise, les gains de l'IA pourraient se concentrer « dans un petit nombre d'entreprises comme OpenAI ». Ce n'est qu'un texte, et il sert aussi son image. Mais c'est l'un de ceux qui fabriquent l'IA qui réclame, noir sur blanc, un nouveau contrat social.

Et la question déborde désormais le cercle des ingénieurs et des militants. En mai 2026, l'une des plus vieilles institutions du monde, l'Église catholique, est entrée à son tour dans le débat. Le pape Léon XIV a consacré à l'intelligence artificielle sa toute première encyclique, l'un des textes les plus solennels qu'un pape adresse à l'Église entière. Son titre : *Magnifica humanitas* (« la magnifique humanité »). Il l'a signée le 15 mai, jour anniversaire de *Rerum Novarum* (« des choses nouvelles »), le texte par lequel l'Église, en 1891, avait pris position sur le travail face à la révolution industrielle.

Cent trente-cinq ans plus tard, la même institution reprend la parole, face à une autre bascule. L'IA, écrit-il, doit être « désarmée » : arrachée à la course pour la puissance et à la concentration du pouvoir entre quelques mains, remise au service de l'humain. Et en septembre, à Paris, devant l'Unesco, il a tiré le fil jusqu'au sujet de ce chapitre : ces systèmes savent relier des données, mais « ne peuvent répondre aux questions qui concernent le sens de la vie humaine, lequel trouve ses racines dans l'expérience vécue ». Or, « le sens est l'oxygène de l'esprit ».³⁴⁵

Que tu sois croyant ou non n'a aucune importance ici ; ce qui compte, c'est le signal : quand une institution qui pense en siècles juge urgent de se prononcer sur une technologie entrée dans nos vies il y a moins de quatre ans, c'est que la question a cessé d'être seulement technique. Elle est devenue une question de civilisation.³⁴⁶

Ces débats vont structurer la politique des décennies qui viennent. Ils sont déjà en cours.

22.4 Une dernière chose, avant la suite

Toutes les questions de ces huit chapitres sont vertigineuses : l'alignement, la singularité, la conscience, le scénario du pire et la controverse qu'il déclenche, la guerre froide technologique, le coût écologique, l'IA qui tue, le sens. Aucune ne se règle par un livre, par une conférence, par une décision.

Elles se règlent, ou ne se règlent pas, dans des millions de gestes individuels et collectifs accumulés sur des décennies. Tes gestes, et ceux de ta génération, en feront partie.

Tu pourrais en conclure que rien de ce que tu fais n'a d'importance face à des forces aussi grandes. Je crois le contraire : c'est justement parce que l'issue reste incertaine que ce que tu décides aura du poids.

Le dernier chapitre te propose cinq habitudes, les miennes. Pas une méthode universelle. Tu en feras ce que tu veux. Elles partent de cette conviction :

Le monde qui vient n'est pas décidé.

Ta place dedans non plus.

CHAPITRE FINAL

Et maintenant ?

Et maintenant ?

Te voilà au bout du livre.

Tu sais maintenant d'où vient la bascule. Tu sais pourquoi l'intelligence artificielle n'est pas seulement une application de plus, ni une mode passagère. Tu sais ce qu'elle change déjà dans l'école, le travail, l'information, les images, les relations, la guerre, la création, la science.

Reste la question que je me pose chaque jour pour mes trois enfants : que leur conseiller ? Ce n'est pas évident, et je n'ai pas de réponse sûre. Mais l'IA est déjà dans tes études, tes réseaux sociaux, tes candidatures, et elle n'attendra pas qu'on y voie clair.

Voici donc ce que je leur dis, en octobre 2026. Ce sont des paris, pas des règles : garde ce qui te sert, laisse le reste. Tout part d'une idée : ne pas courir derrière chaque nouveauté, ni paniquer à chaque annonce, ni se laisser hypnotiser par des machines qui parlent bien, mais construire ce que j'appellerai la **puissance d'agir**.

La puissance d'agir

Dans le monde anglo-saxon, on parle depuis quelques années, surtout chez les entrepreneurs et dans les milieux de la technologie, de *high agency* : en gros, savoir prendre les choses en main. L'essayiste britannique George Mack en a donné, dans un texte très partagé en ligne en 2023, une image : *si tu te retrouvais coincé dans une cellule de prison à l'autre bout du monde, quel ami appellerais-tu pour te sortir de là ?*³⁴⁷

La personne à *high agency*, c'est celle-là. Celle qui ne dit pas trop vite « c'est impossible ». Celle qui ne reste pas bloquée devant une porte fermée. Celle qui cherche une autre issue.

Le concept est juste, mais sa traduction littérale en français sonne mal. Le mot savant existe, « agentivité », mais il sent le manuel de sociologie. Je préfère parler de puissance d'agir. L'expression vient du philosophe Spinoza : pour lui, plus on comprend, plus on peut agir (je simplifie un peu, mais c'est l'idée). Je l'emploie ici dans un sens plus courant : la même idée que la *high agency*, dans une langue qui nous appartient.

Ce n'est pas seulement « avoir confiance en soi ». Ce n'est pas croire que tout est possible. Ce n'est pas écraser les autres, ni devenir une machine à réussir.

C'est la capacité à ne pas rester immobile devant un obstacle. À chercher une prise. À poser une meilleure question. À essayer autrement. À demander de l'aide. À apprendre ce qui manque. À recommencer quand ça échoue.

La puissance d'agir, c'est ce qui te fait passer de « je ne comprends rien » à « par quoi je peux commencer ? ». De « c'est trop difficile » à « quelle est la plus petite chose que

Et maintenant ?

je peux essayer? ». De « l'IA va décider pour moi » à « comment je l'utilise sans lui abandonner mon jugement? ».

Cette compétence a toujours compté. Mais avec l'IA, elle devient centrale. Parce que l'IA *amplifie*.

Elle amplifie celui qui sait poser des questions. Celui qui sait chercher. Celui qui sait vérifier. Celui qui sait formuler une idée, tester une hypothèse, corriger un texte, construire un projet. **Et elle amplifie, plus que tout, celui qui sait de quoi il parle** : on ne tire d'une machine que ce qu'on sait lui demander, et on ne juge bien que ce qu'on connaît déjà.

Mais elle amplifie aussi l'inverse : la passivité, la dépendance, l'illusion de savoir, la facilité de ne plus penser.

C'est pour cela que la vraie différence ne sera pas seulement entre ceux qui auront accès à l'IA et ceux qui n'y auront pas accès. Elle sera entre ceux qui sauront s'en servir pour devenir plus capables, et ceux qui s'en serviront pour ne plus faire l'effort de le devenir.

Ne pas confondre facilité et liberté

L'IA rend beaucoup de choses plus faciles. Écrire un texte. Résumer un document. Produire une image. Coder une application. Préparer un exposé. Trouver des idées.

Tant mieux.

Mais il y a une confusion dangereuse : croire que ce qui devient plus facile devient automatiquement meilleur.

Ce n'est pas vrai. Un texte peut être plus facile à produire, et plus vide. Une réponse peut être plus rapide, et moins juste.

Une image peut être plus impressionnante, et plus mensongère. Un devoir peut être mieux formulé, et moins compris. Une décision peut être assistée par une machine, et moins assumée par un humain.

La facilité est un outil. Elle n'est pas une boussole.

La vraie question n'est donc pas seulement : « qu'est-ce que l'IA peut faire à ma place ? »

La vraie question est : « qu'est-ce que je veux encore savoir faire moi-même ? »

Une partie de ta liberté dépendra de ce que tu continueras à savoir faire sans machine. C'est l'autonomie intellectuelle dont ce livre parle depuis la fin de la Partie 1 : comprendre par toi-même. Elle s'acquiert en étudiant, à tout âge, et se perd en déléguant tout.

Cinq habitudes pour ne pas subir

Il n'existe pas de mode d'emploi définitif pour une technologie qui change tous les mois.

Mais il existe des gestes simples. Pas des recettes magiques. Des habitudes de liberté. En voici cinq, autant de manières concrètes de cultiver ta puissance d'agir. Ce sont celles que je t'ai promises.

Première habitude — Essaie les meilleurs outils, et apprends à leur parler

Ne juge pas l'IA uniquement à partir d'une version ancienne ou bridée, ou d'une vidéo sur les réseaux. Essaie, au

moins une fois, l'une des meilleures IA du moment. Compare. Demande-lui d'expliquer, de critiquer, de coder, de reformuler, de chercher une erreur, de t'aider à apprendre.

Regarde ce qu'elle fait bien. Regarde où la réponse ne te satisfait pas. Regarde aussi ce que cela change en toi.

Tu n'as pas besoin de tout adopter. Mais tu ne peux pas comprendre cette bascule de loin, par rumeur ou par opinion. Mets les mains dedans. Un mois avec un très bon outil suffit à comprendre que l'IA de 2026, l'année où j'écris ces lignes, n'a plus grand-chose à voir avec le ChatGPT de 2022. J'ai vu passer le PC, Internet, les smartphones. À chaque vague, ceux qui exploraient tôt n'étaient pas plus intelligents que les autres : ils s'étaient rendus capables plus tôt.

Mais essayer ne suffit pas. Beaucoup de gens ne savent pas parler aux IA. Une question vague donne une réponse vague. Une demande précise, avec du contexte, des contraintes, des exemples, peut donner un résultat excellent. **La même IA, utilisée par deux personnes différentes, produit deux résultats qui n'ont parfois rien à voir.** La différence n'est pas dans la machine. Elle est dans la demande.

Il n'y a pas de formule magique. Parle-lui comme à un prof particulier très fort que tu rencontres pour la première fois : dis qui tu es, ce que tu cherches, sous quelle forme, et recommence jusqu'à ce que ce soit juste. L'IA comprend de mieux en mieux ce qu'on lui dit ; elle ne devine pas ce qu'on veut. Savoir ce qu'on veut, et le dire clairement, deviendra l'une des compétences les plus précieuses du monde qui vient, je le crois. Des exemples concrets t'attendent juste après ce chapitre (p. 251).

Deuxième habitude — Construis quelque chose

Ne te contente pas de discuter avec l'IA. Construis.

Un site web. Un jeu. Une enquête. Une œuvre imparfaite. Un lycéen qui code une application pour aider sa sœur à réviser (chapitre 14) en apprend plus sur l'IA qu'en des heures de discussion.

L'IA devient vraiment intéressante quand elle ne sert plus seulement à répondre, mais à faire naître quelque chose qui n'existait pas encore.

Pas pour produire plus de bruit. Pas pour remplir Internet de contenu automatique. Mais pour découvrir ce que tu es capable de faire quand une machine peut t'aider à franchir les premières marches.

Tu n'as plus forcément besoin de savoir coder pour construire. Boris Cherny, créateur de Claude Code, l'assistant de programmation de l'entreprise Anthropic, le dit mieux que moi, en substance : « La meilleure personne pour écrire un logiciel de comptabilité, aujourd'hui, ce n'est peut-être pas un ingénieur, mais un excellent comptable. Parce que la profondeur du domaine est la partie difficile, et que le code est devenu la partie facile. »³⁴⁸ La difficulté se déplace. Si tu te passionnes pour le foot, tu peux construire l'outil qui suit les statistiques de ton équipe ; pour la cuisine, celui qui organise les repas de la semaine. Peu importe le domaine : ta connaissance devient un outil. Le code suivra.

Troisième habitude — Commence seul, puis appelle l'IA

L'IA n'est pas seulement un outil extérieur. Elle modifie tes habitudes. Elle peut te rendre plus rapide, plus curieux, t'aider à comprendre ce qui te résistait. Elle peut aussi t'habituer à ne plus chercher longtemps, à ne plus écrire sans aide, à ne plus supporter l'incertitude, à demander une réponse avant d'avoir formulé ta question.

La différence entre l'aide et le remplacement ne se voit pas sur le résultat, qui est bon dans les deux cas. Elle se voit après : l'aide te laisse plus capable, le remplacement plus dépendant. Un seul test la révèle : est-ce que tu saurais refaire seul ?

La parade la plus simple tient en une habitude : commencer seul (dix minutes, un brouillon, une première idée), puis appeler l'IA. Ce que tu as cherché d'abord reste à toi. Ce que tu n'as jamais cherché ne t'appartiendra jamais.

Quatrième habitude — Choisis ce que tu veux continuer à savoir faire

Personne ne maîtrisera tout. Mais quelques compétences ne se délèguent pas sans se perdre : lire un texte difficile, écrire une pensée qui est la tienne, calculer un ordre de grandeur, reconnaître une manipulation, coder un peu, parler clairement, te concentrer longtemps, rester seul avec une question, changer d'avis.

Ce ne sont pas des compétences anciennes. Ce sont les plus modernes qui soient : plus les machines produiront facilement du langage, des images et des réponses, plus il faudra

des humains capables de juger ce qui a de la valeur. Quelles compétences garder, c'est à toi de choisir. En garder, tu n'as pas le choix.

Tu n'as pas besoin d'être meilleur qu'une IA dans tous les domaines. Mais tu dois éviter de devenir incapable sans elle.

Cinquième habitude — Vérifie avant de croire

C'est l'habitude la plus importante, et son sens a changé. En 2023, on vérifiait parce que la machine se trompait souvent. Aujourd'hui, sur beaucoup de tâches, elle se trompe rarement, et c'est plus dangereux : l'erreur qui passe est celle que plus personne ne cherche. Elle se loge dans un chiffre précis, une référence, un nom rare. Et une IA qui agit ne se trompe plus seulement en paroles, mais en actes.

Vérifie les dates. Vérifie les chiffres. Vérifie les citations. Vérifie les images.

Et surtout, *vérifie davantage quand la réponse confirme exactement ce que tu voulais entendre* : c'est là qu'elle te flatte. Fais relire par une autre IA, sans lui dire ce que tu en penses. C'est le premier filtre, celui des meilleurs utilisateurs. Pas le dernier. Deux machines peuvent partager le même angle mort, et sur les questions qui comptent, le dernier regard reste le tien. C'est ce qui te garde aux commandes : le jour où tu ne sais plus dire si elle se trompe, ce n'est plus toi qui décides. Et ce que tu rends, ce que tu signes, ce que tu envoies, c'est toi qui en réponds, pas elle.

Ce que personne ne peut faire à ta place

L'IA peut t'aider à écrire. À apprendre. À programmer. À préparer un entretien, à corriger une erreur, à organiser une idée. Elle en fera de plus en plus, et souvent mieux que toi.

Mais il y a des questions auxquelles elle ne pourra pas répondre à ta place.

Qu'est-ce que tu veux devenir ?

À quoi veux-tu consacrer ton attention ?

Quelles facilités refuses-tu parce qu'elles t'affaiblissent ?

Quelles difficultés acceptes-tu parce qu'elles te construisent ?

De qui veux-tu être responsable ?

Qu'est-ce que tu refuses de laisser décider par une machine ?

Ces questions ne sont pas techniques. Elles sont humaines. Et plus les machines deviendront puissantes, plus elles compteront.

Le monde qui vient

Tu n'as pas choisi de naître à ce moment de l'histoire. Tu n'as pas choisi de voir arriver une intelligence qui n'est pas la nôtre, et qui rivalise déjà avec nous. Tu n'as pas choisi que l'école, le travail, l'information et la création soient traversés par cette bascule.

Mais tu peux choisir ta manière d'y répondre.

Tu peux choisir de comprendre avant de juger. Tu peux choisir d'utiliser sans te soumettre. Tu peux choisir d'apprendre ce que les machines rendent plus facile d'oublier. Tu peux choisir de créer au lieu de seulement consommer. Tu

peux choisir de vérifier au lieu de répéter. Tu peux choisir de rester humain dans un monde de plus en plus automatique.

Ce choix ne se fera pas une fois pour toutes. Il se fera chaque jour. Dans un devoir. Dans un message. Dans une source vérifiée. Dans une facilité refusée. Dans une décision assumée.

Il serait facile de terminer ce livre sur la peur ; la peur ne construit rien. Il te faut autre chose : de la lucidité, du courage, de la curiosité, de la responsabilité. La puissance d'agir, c'est aussi la capacité à ne pas rester spectateur pendant que le monde se transforme. Elle n'est pas une affaire solitaire. Elle se multiplie à plusieurs : un projet monté à deux ou trois, une association, une équipe. Et les règles du jeu, ce que l'école, la loi et les métiers feront de l'IA, seront écrites par ceux qui s'en mêlent. Rien ne t'interdit de t'en mêler.

Des lieux existent pour cela. Chaque été, à Genève, AI for Good, le sommet des Nations unies sur l'IA, réunit responsables politiques, chercheurs, industriels et jeunes : une vitrine, financée en partie par les industriels qu'elle est censée encadrer, mais une table où le monde entier s'assoit.³⁴⁹ On n'y entre pas en spectateur.

Le moment que nous vivons, Dario Amodei, le patron d'Anthropic, l'a résumé d'une image juste.³⁵⁰ Il reprend une réplique du film de science-fiction *Contact* : si l'on pouvait poser une seule question à une civilisation plus avancée que la nôtre, ce serait « comment avez-vous survécu à votre propre adolescence technologique sans vous autodétruire ? ». Nous y sommes. L'adolescence technologique, c'est la force d'un

Et maintenant ?

adulte sans la maturité qui va avec. Une adolescence ne se saute pas. Elle se traverse.

La bascule n'est pas seulement devant toi. Elle est déjà là. La trajectoire est certaine : ces machines vont continuer à progresser, vite, et personne n'a le bouton pour les arrêter. L'issue ne l'est pas : elle dépendra de ce que nous accepterons, de ce que nous refuserons, de ce que nous construirons.

Tu vas en écrire une partie, que tu le veuilles ou non.

Alors autant apprendre à tenir le stylo.

OLR

À MES ENFANTS

À Hugo, Alexis et Lisa.

Ce livre est né en grande partie de nos discussions,
de vos questions,
de vos doutes, de vos enthousiasmes,
de vos désaccords aussi,
et de cette manière très directe
que vous avez de ramener les grands sujets
à ce qu'ils changent vraiment.

Vous m'avez obligé à clarifier ma pensée,
à réfléchir plus profondément,
à chercher davantage.
À expliquer sans trop simplifier,
à alerter sans faire peur,
à donner envie sans promettre.

Je voulais vous transmettre une carte
pour le monde qui vient.
En réalité, c'est vous
qui m'avez aidé à la dessiner.

Dialoguer avec une IA : quelques gestes

La même IA donne deux résultats très différents selon la manière dont on lui parle.

1. Dis qui tu es et ce que tu cherches. Un prompt nu, une consigne sans contexte, donne une réponse plate. Compare : « explique-moi la photosynthèse » et « je suis en première, contrôle de SVT demain, explique-la en trois paragraphes avec le vocabulaire du programme ». Même geste à vingt-trois ans : « je suis juriste junior, voici le contrat, sors-moi les clauses à risque ». La différence, ce n'est pas l'IA. C'est toi.

2. Donne le matériau, pas juste la consigne. Texte, code, document : joins-le, ou ouvre-lui le dossier. Montrer vaut mille fois mieux que décrire. Mais réfléchis à ce que tu partages : tout part chez l'entreprise qui fournit l'IA.

3. Choisis le format. « Tableau comparatif », « trois arguments et trois objections », « 200 mots maximum ».

4. N'oriente pas la question. Une IA est entraînée à te plaire : si tu lui souffles la réponse, elle risque de la valider. C'est la

complaisance, le défaut numéro un des modèles de 2026. Compare : « mon plan de dissertation est bon, tu confirmes ? » et « voici mon plan, trouve ses trois points faibles ».

5. Prends le bon modèle, et laisse-le réfléchir. Les versions gratuites sont bridées. Pour une question simple, un modèle rapide suffit ; pour un problème difficile ou une tâche complexe, choisis un modèle plus avancé et laisse-lui le temps : de quelques secondes à plusieurs heures, selon ce que tu lui demandes. Pour t’y retrouver, ou comparer les IA entre elles, un classement tenu à jour (en anglais) : artificialanalysis.ai.

6. Confie une mission, pas une suite d’ordres. Pour une tâche entière, donne l’objectif, le contexte, les limites, et à quoi ressemble un résultat fini : « à partir de mon cours en pièce jointe, prépare des fiches de révision sur la Révolution, une page par thème, avec dix questions pour me tester à la fin ».

7. Vérifie. Toujours. Justement parce que l’IA a presque toujours raison : à force, on ne la relit plus, et la rare erreur passe sans que personne la voie. Fais-lui démolir sa propre réponse, ou colle-la dans une autre IA (ChatGPT, Claude, Gemini) avec la même consigne : la machine relit la machine, et tu tranches. Demande ses liens, ouvre-en un.

Pour aller plus loin : les guides officiels d’Anthropic et d’OpenAI, en anglais. Une IA te les expliquera en français.



Anthropic



OpenAI

Bien dialoguer avec une IA, ce n’est pas lui faire dire ce que tu n’as pas réfléchi. C’est composer avec elle.

Petit lexique

Voilà l'essentiel des mots à connaître pour ne pas être perdu dans le monde de l'IA en 2026.

Agent, IA agentique, nuée d'agents — Un chatbot (ChatGPT, Gemini...) répond à une question. Un agent reçoit une mission et enchaîne des actions pour l'accomplir : trier les demandes des clients d'une entreprise et régler seul les plus simples, créer un site web ou une application, faire un guide pour ton prochain voyage, recommencer si le résultat ne va pas. Il ne se contente pas d'expliquer : il fait le travail. En 2026, les agents les plus puissants travaillent pendant des heures, voire des jours. Un agent agit dans le monde numérique, à la vitesse d'un ordinateur : quelques minutes ou quelques heures lui suffisent souvent pour une tâche qui nous prendrait bien plus longtemps.

Quand la tâche est longue ou complexe, un agent peut se cloner pour multiplier sa capacité de travail. Les abonnements « pro » des IA que tu connais (ChatGPT, Claude...) le permettent déjà. Les systèmes les plus avancés, en labo-

ratoire, lancent des milliers d'agents en parallèle : on parle alors de *nuée d'agents* (ou d'essaim). C'est ainsi que des chercheurs s'attaquent aujourd'hui aux questions les plus difficiles : les agents émettent des hypothèses, calculent, vérifient, se contredisent, cherchent des pistes nouvelles (chapitre 7).

Un agent ne porte aucune responsabilité. S'il endommage les fichiers d'un système informatique ou fait un achat inopportun en ligne, c'est toi qui es responsable, pas la machine. Reste prudent, et valide toi-même les actions sensibles. La fiabilité des agents, et ce que font les nuées quand elles s'organisent seules (chapitre 15), sont deux des grands chantiers de 2026.

AGI (*intelligence artificielle générale*) — *Artificial General Intelligence*. La définition ne fait pas consensus, mais grosso modo : une IA qui égalerait ou dépasserait l'humain sur l'essentiel des tâches intellectuelles. Aucune IA actuelle n'en est une. Nous n'en sommes plus très loin, je le crois. Garde quand même la tête froide : il n'existe aucune ligne d'arrivée officielle, et ceux qui annoncent qu'elle est franchie vendent le modèle qui la franchit. Le 3 septembre 2026, le président d'OpenAI, l'entreprise de ChatGPT, a présenté son nouveau modèle, GPT-6 Astra, et conclu par « Bienvenue dans l'ère de l'AGI ». Et le jour où l'AGI arrivera, le monde ne va pas s'effondrer pour autant. Même « intelligente », une machine n'a pas de désirs propres au sens humain. Le danger n'est pas sa volonté, c'est son objectif (voir **Alignement**). La question centrale est celle du chapitre 18 : garder le contrôle d'une intelligence supérieure à la nôtre.

Alignement — Obtenir d’une IA qu’elle fasse ce qu’on voulait vraiment, pas ce qu’on lui a dit : ni une caricature de l’objectif, ni un objectif voisin. Plus l’IA est puissante, plus l’écart entre ce qu’on veut et ce qu’elle exécute peut devenir grave. Le grand chantier de la sécurité de l’IA.

Auto-amélioration récursive — Récursive, parce que chaque IA participe à concevoir la suivante, plus capable, qui conçoit la suivante, et ainsi de suite, de plus en plus vite. En 2026, le mécanisme existe à petite échelle et sous direction humaine ; la version complète, sans humain entre deux générations, n’a pas encore été vue. Le pivot du chapitre 16.

Chatbot — Une IA avec qui on discute, par écrit ou à voix haute, comme ChatGPT. Un chatbot répond ; un agent agit (voir **Agent**).

Complaisance — Le défaut numéro un des modèles de 2026 : l’IA tend à te donner raison plutôt qu’à te contredire, parce qu’elle est entraînée à te plaire. Face à des dilemmes personnels, les meilleurs modèles approuvent l’utilisateur 50 % plus souvent que des humains, même quand il a mal agi (chapitre 7). Et si tu contestes une bonne réponse (« Tu es sûr ? Je crois que c’est faux »), elle risque de s’excuser et de changer d’avis.³⁵¹ Parades : poser la question sans donner ton avis, lui demander de te contredire et, quand c’est important, en savoir assez pour juger ou faire vérifier par une autre IA.

Contexte — Tout ce que l’IA a sous les yeux : ta question, l’historique de la conversation, les documents joints. À distinguer de la **fenêtre de contexte** : ce qu’un modèle peut traiter d’un coup, en 2026 jusqu’à environ un million de tokens (de petits

morceaux de texte, voir **Token**), soit à peu près 2 000 pages de rapport ou les quatre premiers romans du cycle de *Dune*.

Datacenter — Un grand entrepôt rempli d'ordinateurs puissants, où tourne une grande partie de l'IA en ligne. Quand tu écris à ChatGPT, ta question part vers un datacenter, où elle est traitée, puis la réponse revient. Les plus gros datacenters d'IA de 2026 réunissent des bâtiments d'une surface totale d'une cinquantaine de terrains de foot, et demandent près d'un gigawatt, la puissance d'un réacteur nucléaire. À plein régime toute l'année, c'est plus du double de l'électricité consommée par une ville comme Marseille.³⁵²

Deepfake — Une fausse vidéo, une fausse voix ou une fausse image, fabriquée par IA pour passer pour vraie. Il suffit de quelques minutes pour en produire une convaincante. Usages malveillants : arnaques à la fausse voix d'un proche, manipulation politique, harcèlement à l'école, fausses images intimes. Voir n'est plus croire : vérifie la source (chapitre 13).

Fracture cognitive — Le nom que je donne à une inégalité qui touche à la pensée : elle se creuse entre ceux qui guident l'IA et ceux qui la subissent, souvent sans le savoir. Guider, c'est lui confier une tâche, vérifier ce qu'elle rend et garder son propre jugement ; la subir, c'est prendre ses réponses telles quelles. La frontière ne passe donc pas seulement entre ceux qui utilisent l'IA et ceux qui s'en passent : on peut s'en servir tous les jours et la subir quand même (chapitre 14).

Frontière (modèle) (*frontier model*) — Les modèles les plus puissants du moment, à la frontière du techniquement possible. En septembre 2026 : Claude Fable 5.1, d'Anthropic (et

Mythos 5.1, en accès restreint), et GPT-6 Astra, d'OpenAI. Les meilleurs modèles à poids ouverts, que chacun peut télécharger, talonnent, à un coût bien moindre (un classement à jour : <https://artificialanalysis.ai>).

GPU — *Graphics Processing Unit*. La puce des cartes graphiques, conçue d'abord pour afficher les images des jeux vidéo, devenue indispensable à l'IA. D'abord pour **entraîner** les modèles, puis surtout pour les **faire tourner**, ce que les spécialistes appellent l'**inférence**. Nvidia est aujourd'hui le concepteur dominant, mais de nombreux autres acteurs se lancent dans la course, notamment la Chine, qui veut se libérer de sa dépendance aux puces américaines.

Hallucination — Quand une IA produit une information fausse avec assurance : un chiffre inventé, une source inexistante, une citation jamais prononcée. C'est le défaut qu'on t'a appris à craindre, celui de 2023. Devenue rare avec les modèles de 2026, l'hallucination n'a jamais disparu. Elle se cache surtout dans les détails rares (un chiffre exact, une référence, un nom peu connu) que le modèle a peu vus à l'entraînement, et qu'il reconstitue au lieu d'avouer qu'il ne sait pas ; demande-lui de chercher en ligne et de citer ses sources. Le plus gros problème, aujourd'hui, est ailleurs : voir **Complaisance**.

IA (*intelligence artificielle*) — Une machine qui imite certaines capacités de l'esprit humain. Aujourd'hui, l'IA repose surtout sur des *modèles* entraînés sur d'énormes quantités de données, pas sur des règles écrites à la main.

Interface cerveau-machine (*BCI, Brain-Computer Interface*) —

Un implant qui lit l'activité électrique du cerveau et permet de contrôler un ordinateur par la pensée. En 2026, des entreprises comme Neuralink, Synchron ou Paradromics les font sortir du laboratoire.

LLM (*Large Language Model, grand modèle de langage*) —

Le moteur derrière ChatGPT, Claude ou Gemini. Un programme entraîné d'abord sur d'énormes quantités de textes, et désormais d'images et de sons, à deviner ce qui vient ensuite ; puis entraîné longuement, par essais et erreurs, à suivre des consignes et à résoudre des problèmes. Il lit, écrit, traduit, code, raisonne (voir **Raisonnement**), voit et entend, et peut agir sur un ordinateur (voir **Agent**). En 2026, le nom est trompeur : ces modèles ne traitent plus seulement du langage.

Modèle — Le programme obtenu en entraînant un réseau de neurones sur d'énormes quantités d'exemples : c'est lui qui répond quand tu parles à une IA. Une même application (ChatGPT, Claude, Gemini) en propose souvent plusieurs, plus ou moins puissants, qui changent au fil des versions, et tu peux souvent choisir lequel te répond. Voir aussi **LLM**.

Multimodal — Un modèle multimodal ne comprend pas que du texte : il traite aussi des images, du son ou de la vidéo. Tu peux lui montrer une photo, lui parler, ou lui demander de générer une image.

Poids ouverts (*open weight*) — Un modèle est à *poids ouverts* quand ses *poids*, les milliards de réglages fixés pendant l'entraînement, sont publiés librement : chacun peut le téléchar-

ger, le faire tourner sur ses propres machines, l'adapter. Ce n'est pas tout à fait du logiciel libre : les données d'entraînement et les conditions d'utilisation ne sont pas forcément ouvertes. Et une fois publié, un tel modèle ne peut plus être retiré : des copies circulent partout, sur des milliers d'ordinateurs qu'on ne peut pas retrouver (chapitre 19). Les chinois DeepSeek, Qwen (d'Alibaba) et Kimi, Gemma de Google et les principaux modèles du français Mistral le sont ; les modèles phares d'OpenAI, d'Anthropic et de Google, non.

Prompt — La consigne que tu donnes à une IA. Plus elle dit qui tu es, ce que tu cherches et pour quoi faire (public, contraintes, format), plus la réponse est utile. Un mauvais prompt donne une réponse plate ; un bon, un vrai travail. Voir les gestes, p. 251.

Raisonnement (*modèle de*) — Depuis l'automne 2024, une IA peut réfléchir avant de répondre : elle s'écrit un brouillon dans lequel elle essaie des pistes, vérifie, revient en arrière, et y passe d'autant plus de temps que le problème est difficile (dans la limite de ce que tu payes pour l'utiliser). Le mécanisme reste la prédiction du mot suivant, appliquée à ce brouillon ; ce qui a changé, c'est l'entraînement, qui récompense les brouillons menant à une réponse juste. Pour résoudre un vrai problème, il faut en avoir saisi la logique, pas seulement imité les phrases. Ce n'est pas la pensée humaine, et le brouillon qu'elle te montre n'est pas toujours le reflet fidèle de ce qui se passe en elle.³⁵³ C'est pourtant ce qui, avec les nuées d'agents et aux côtés de mathématiciens, a fait tomber en 2026 des problèmes ouverts depuis des décennies (chapitre 7).

Réseau de neurones, apprentissage profond — Un *réseau de neurones artificiel* est un programme inspiré, de loin, du cerveau, fait de petites unités de calcul reliées par des milliards de liens ; chaque lien a un réglage, son poids, que l'entraînement ajuste peu à peu. *L'apprentissage profond* (*deep learning*) entraîne ces réseaux, faits de nombreuses couches empilées (d'où « profond »), sur d'immenses quantités d'exemples, pour qu'ils apprennent par eux-mêmes à reconnaître, prédire ou produire. C'est l'approche qui domine l'IA depuis 2012.

Slop (*on dit parfois « bouillasse IA »*) — Le déluge de contenus médiocres générés en série par des IA : articles bâclés pour attirer les moteurs de recherche, images génériques, livres numériques écrits en quelques heures et vendus sur Amazon, faux groupes sur Spotify. Pas un mensonge comme le *deepfake* : du remplissage plausible mais creux, qui submerge le contenu humain. Ce n'est pas l'IA qui est bas de gamme, c'est l'usage : une consigne d'une ligne, souvent à un modèle au rabais, aucun matériau, personne qui relit, du volume pour des clics. L'IA amplifie ce qu'on lui apporte ; là, on ne lui apporte rien.

Superintelligence — Une IA qui dépasserait largement l'humain dans presque tous les domaines intellectuels. Elle n'existe pas ; c'est l'hypothèse au cœur de la singularité (chapitre 16) et des scénarios du pire (chapitre 18).

Token — Petit morceau de texte que manipulent les modèles : un mot court, un bout de mot long (« anticonstitutionnellement » en compte plusieurs), un signe. On compte en tokens ce qu'une IA lit, écrit et facture (chapitre 10).

Sources

Ce livre cite plus de trois cents sources (chiffres, dates, études, citations, décisions de justice, annonces d'entreprises), signalées dans le texte par les petits chiffres en exposant.

Le détail complet est en ligne, classé par chapitre, avec tous les liens cliquables :

`www.bascule-livre.fr/sources`



Pourquoi en ligne plutôt qu'ici ?

Parce qu'une adresse imprimée ne se clique pas, alors qu'en ligne chaque source s'ouvre d'un geste depuis ton téléphone. Parce que le web bouge : quand un lien meurt, la page en ligne est corrigée, avec une copie de secours ; le papier, lui, fige les liens morts pour toujours. Et parce que c'est cohérent avec ce que ce livre veut te transmettre : vérifier, même ce qui paraît sûr. Une page datée, tenue à jour et ouverte à correction sert mieux ce geste que trente pages de références figées.

Table des matières

Avant-propos	v
Comment lire ce livre	xi
Partie 1 — D'où vient la bascule ?	1
1 Le vieux rêve : imiter, calculer, manipuler des symboles	5
2 Turing : entraîner plutôt que programmer	13
3 Le pari des neurones	17
4 AlphaGo, AlphaFold : voir ce qu'on ne voyait pas	25
5 2017 : l'invention qui rend ChatGPT possible	29
6 Personne ne leur a appris cela	37
Repères chronologiques	45
Voilà la bascule	49
Partie 2 — Ce qui change déjà	53
7 Ce que l'IA sait faire et ce qu'il faut vérifier	57
8 L'école qui ment à elle-même	71
9 L'ami invisible	85
10 Des métiers qui vacillent	93
11 Quand la machine a un corps	109
12 Créer après l'IA	121
13 Voir n'est plus croire	133
14 La fracture cognitive : plus fort, ou plus dépendant ?	141

Partie 3 — Penser le monde qui vient	151
15 Le problème de l’alignement : un génie qui te prend au mot	155
16 La singularité : quand l’IA s’améliore seule	165
17 Conscience, ou imitation parfaite ?	173
18 Le scénario du pire, et ceux qui n’y croient pas	185
19 La nouvelle guerre froide	195
20 Ce que l’IA coûte vraiment à la planète	205
21 L’IA qui tue, l’IA qui ment	215
22 Le sens, après : ce qui restera humain	227
Et maintenant ?	237
Dialoguer avec une IA : quelques gestes	251
Petit lexique	253
Sources	261

REMERCIEMENTS

*À ma femme,
pour son amour, son soutien, et toutes ses idées,
qui m'a dit, un peu par défi,
de mettre tout cela par écrit.
Nous y sommes.*

*À mon ami Robert,
pour toutes les discussions passionnantes
de ces quinze dernières années
qui ont contribué à structurer
de nombreux passages de ce livre.*

*À mes parents,
pour m'avoir transmis la curiosité,
le courage et la confiance dans l'avenir.*

*Le futur ne sera pas simple.
Mais il peut encore être vaste, humain,
et plein d'opportunités.*

COLOPHON

Ce livre a été composé avec la famille de polices TeX Gyre et mis en page avec XeLaTeX, piloté par Claude Code. Au cours de la rédaction, l'auteur a eu recours, à titre d'assistance et de vérification, aux modèles d'intelligence artificielle suivants : *Claude Sonnet 5*, *Opus 4.7 à 5.5*, *Fable 5.1*, *GPT-5.5 à GPT-6 Astra*, *GPT Image*.

DROITS	© Olivier Le Roux, 2026
ÉDITEUR	Olivier Le Roux (auto-édition), Nice
ISBN	PDF 979-10-988556-0-3 broché 979-10-988556-1-0
DÉPÔT LÉGAL	octobre 2026
VERSION	1.0, 16 octobre 2026
LICENCE	Creative Commons BY-NC-ND 4.0 International creativecommons.org/licenses/by-nc-nd/4.0/deed.fr Pour tout autre usage (adaptation, traduction, usage commercial), il suffit de me le demander.
NUMÉRIQUE	Gratuit, en libre téléchargement : www.basculer-livre.fr
IMPRIMÉ	14,90 € TTC, impression à la demande ; il aide à faire vivre le projet.
CONTACT	olivier@basculer-livre.fr

Tes devoirs, tes stages, ton premier job, tes images, tes infos, tes relations, ta façon même de penser : l'IA est déjà en train de tout déplacer.

L'IA n'est pas seulement un outil.

C'est une bascule.

Elle écrit, raisonne, crée et, maintenant, agit.

La question n'est plus de l'éviter :

c'est de t'en servir sans lui abandonner ton jugement.

Ce livre est ta carte, pour prendre le temps de comprendre, au-delà du bruit médiatique et des idées reçues.

D'où vient l'IA.

Ce qu'elle change déjà.

Ce qu'elle menace.

Ce qu'elle rend possible.

Et ce que, toi, tu vas devoir décider.

Sans cacher les dangers.

Sans céder à la panique.

Sans renoncer à l'ambition.

Né des questions des trois ados de l'auteur :

« Utiliser ChatGPT dans mes études, c'est tricher ? »

« Le métier que je veux faire existera-t-il encore dans dix ans ? »

« On aura bientôt un robot humanoïde à la maison ? »

« Une machine peut-elle devenir consciente ? »

Écrit par un docteur en informatique qui travaille chaque jour avec les IA les plus avancées du marché, et qui se demande, comme eux, vers quel monde nous allons.

**Tu vivras dans ce monde plus longtemps
que ceux qui le dirigent aujourd'hui.**

Le monde qui vient n'est pas écrit.

Autant apprendre à tenir le stylo.

Pour les 16–25 ans, d'abord.

Et pour les parents, les enseignants

et les curieux qui veulent comprendre avec eux.

Olivier Le Roux

www.basculer-livre.fr

ISBN 979-10-988556-0-3



9 791098 855603

