

L'IA et Nous — Le guide du cyber citoyen

Ivan Berlocher

Décembre 2025

Table des matières

L'IA et Nous — Le guide du cyber citoyen	5
Table des matières	5
Le ton de ce livre	7
Structure du dépôt	7
Licence	7
Chapitre 1	8
L'IA n'est pas ce que vous croyez	8
Chapitre 2	46
Forces et faiblesses : ce que l'IA fait vraiment bien (et vraiment mal)	46
PARTIE A : Les super-pouvoirs (réels)	47
PARTIE B : Les faiblesses cachées	51
Chapitre 3	55
Les acteurs : qui contrôle l'IA ?	55
Chapitre 4	60
L'économie de l'IA : suivez l'argent	60
Chapitre 5	64
Vos données, leur trésor	64
Fin de la Partie I — COMPRENDRE	67
PARTIE II : SE PROTÉGER	68
Chapitre 6	69
L'Europe et vos droits : la troisième voie	69
PARTIE A : L'Europe dans la guerre de l'IA	70
PARTIE B : Vos droits concrets (RGPD + AI Act)	73
Chapitre 7	77
Questions à poser — et réponses à exiger	77
PARTIE A : Questions aux entreprises tech	78
PARTIE B : Questions à votre employeur	80

Chapitre 8	83
Évaluer et se protéger — guide pratique	83
PARTIE A : Évaluer un outil IA avant de l'utiliser	84
PARTIE B : Protéger ses données au quotidien	86
Chapitre 9	89
Quand utiliser — et quand ne pas utiliser — l'IA	89
PARTIE A : Les pièges à éviter	90
PARTIE B : Les domaines où ne PAS utiliser l'IA	92
PARTIE C : Quand et comment bien utiliser l'IA	94
Fin de la Partie II — SE PROTÉGER	95
PARTIE III : IMAGINER	96
Chapitre 10	97
L'IA dans 5 ans — prédictions honnêtes	97
Chapitre 11	109
Les emplois transformés	109
Chapitre 12	115
L'éducation repensée	115
Chapitre 13	121
L'avenir de la créativité humaine	121
Chapitre 14	127
Questions aux journalistes — l'esprit critique	127
Chapitre 15	132
Conclusion — Le citoyen augmenté	132
À propos de l'auteur	137
Contact	137
Licence	137
Épilogue	138
L'IA et Nous — Le choix qui reste	138
Annexe A	141
Glossaire — Le lexique de l'IA démystifié	141
Annexe B	148
Checklist : Avant d'utiliser un nouvel outil IA	148
[OK] La Checklist	148
[Feu] Système de scoring	149
[Liste] Template à remplir	149
[Lien] Version imprimable	150
Annexe C	151
Vos droits en Europe (RGPD + AI Act)	151

Partie 1 : Le RGPD — Vos droits sur vos données	151
Partie 2 : L'AI Act — Vos droits face aux systèmes IA	153
Partie 3 : Tableau récapitulatif	154
Partie 4 : Modèles de lettres	155
Ressources utiles	156
Annexe D	157
Ressources pour continuer	157
[Guide] Pour tout le monde — Les incontournables	157
[H][Edu] Pour les curieux — Aller plus loin	158
[F][Ecole] Pour les enseignants — Intégrer l'IA en classe	158
[Pro] Pour les professionnels — IA au travail	159
[H][F][E] Pour les parents — Accompagner ses enfants	160
[Labo] Pour aller encore plus loin — Ressources techniques	161
[Date] Événements à ne pas manquer	161
[!] À éviter — Sources non fiables	162
Annexe E	163
L'architecture Harmonia — Pour aller plus loin	163
Pourquoi cette annexe ?	163
Du fichier au concept : quand l'informatique devient invisible	163
Le problème que Harmonia cherche à résoudre	165
Les trois couches d'Harmonia	165
Le document .life — Reprendre le contrôle	166
Le Gouverneur — Des règles vérifiables	167
L'Alter Ego — Une IA qui vous connaît (avec permission)	168
Cas d'usage illustratifs	168
État d'avancement	169
Pour aller plus loin	169
Conclusion	170
Annexe F	171
Bibliographie	171
Livres — Accessibles au grand public	171
Livres — Techniques (pour aller plus loin)	172
Articles académiques fondateurs	172
Rapports institutionnels	173
Médias et newsletters	173
Podcasts	174
Sites et ressources en ligne	174
Références citées par chapitre	175
Note sur les sources	175

L'IA et Nous

Le guide du cyber citoyen

Comprendre vraiment l'intelligence artificielle
pour rester maître de vos choix

Ivan Berlocher

Version 1.0 — Décembre 2025

Ce livre est diffusé sous licence CC BY-NC-SA 4.0
Attribution — Pas d'utilisation commerciale — Partage dans les mêmes conditions

L'IA et Nous — Le guide du cyber citoyen

Comprendre vraiment l'intelligence artificielle pour rester maître de vos choix

Ce livre peut se lire linéairement ou par chapitres indépendants.

Table des matières

PARTIE 1 : COMPRENDRE

Ce que nous savons avec certitude

Chapitre	Titre	Contenu
1	Comprendre vraiment l'IA	Le manifeste : vocabulaire honnête, MUSIC, ontologies, Europe
2	Forces et faiblesses	Ce que l'IA fait vraiment bien — et vraiment mal
3	Les acteurs	OpenAI, Google, Meta, Anthropic : qui contrôle l'IA ?
4	L'économie de l'IA	Suivez l'argent : freemium, API, cloud, dépendance
5	Vos données, leur trésor	Ce qui se passe quand vous tapez dans ChatGPT

PARTIE 2 : SE PROTÉGER

Ce que vous pouvez faire concrètement

Chapitre	Titre	Contenu
6	L'Europe et vos droits	RGPD, AI Act, la troisième voie européenne

Chapitre	Titre	Contenu
7	Questions à poser	Aux entreprises tech, à votre employeur
8	Évaluer et se protéger	Guide pratique : grilles, outils, alternatives
9	Quand utiliser l'IA	Les bons usages — et les pièges à éviter

PARTIE 3 : IMAGINER

Ce que nous ne savons pas encore — mais pouvons anticiper

Chapitre	Titre	Contenu
10	L'IA dans 5 ans	Prédictions honnêtes : le probable, le possible, le fantasmé
11	Les emplois transformés	Métiers qui changent, compétences qui comptent
12	L'éducation repensée	Que doit-on encore apprendre aux enfants ?
13	L'avenir de la créativité	Art, création, authenticité à l'ère de l'IA
14	Questions aux journalistes	L'esprit critique face au buzz médiatique

CONCLUSION

15	Conclusion — Le citoyen augmenté Épilogue	Ce que vous avez appris, ce que vous pouvez faire L'IA et Nous — Le choix qui reste
-----------	---	--

ANNEXES

Annexe	Titre	Contenu
A	Glossaire	Le lexique de l'IA démystifié
B	Ressources	Pour aller plus loin
C	Droits Europe	Vos droits RGPD en détail
D	Ressources supplémentaires	Liens et références
E	Harmonia	Histoire illustrative
F	Bibliographie	Sources et références

Le ton de ce livre

Ce livre adopte une posture **sage et confiante** :

[OK] “Voici ce que nous SAVONS avec certitude...” (*références, preuves, consensus scientifique*)

[!] “Voici ce que nous SUPPOSONS raisonnablement...” (*tendances observées, extrapolations prudentes*)

[?] “Voici ce que nous NE SAVONS PAS encore...” (*honnêteté épistémique, humilité*)

[>] “Voici ce que vous POUVEZ FAIRE...” (*action concrète, empowerment*)

Structure du dépôt

```
livre-ia-et-nous/  
├── README.md           # Ce fichier  
├── chapitre-01.md → 15.md # Les 15 chapitres  
├── epilogue.md         # Épilogue  
├── annexe-a-glossaire.md # Glossaire  
├── annexe-b-ressources.md # Ressources  
├── annexe-c-droits-europe.md # Droits RGPD  
├── annexe-d-ressources.md # Ressources supplémentaires  
├── annexe-e-harmonia.md  # Histoire Harmonia  
├── annexe-f-bibliographie.md # Bibliographie  
└── old/                 # Anciens fichiers (backup)
```

Licence

Ce livre est en cours de rédaction. Tous droits réservés.

Dernière mise à jour : 17 décembre 2025

Chapitre 1

L'IA n'est pas ce que vous croyez

1.1 Ni robot, ni cerveau, ni oracle

Fermez les yeux. Pensez “intelligence artificielle”.

Qu'est-ce qui apparaît ?

Pour certains, c'est **Terminator** — un robot métallique aux yeux rouges, programmé pour détruire l'humanité. Pour d'autres, c'est **HAL 9000**, l'ordinateur de *2001, l'Odyssée de l'espace*, qui refuse calmement d'ouvrir les portes. D'autres encore voient un cerveau électronique, des fils et des circuits qui pensent comme nous, peut-être mieux que nous.

Oubliez tout ça.

L'IA qui est entrée dans votre vie — celle avec laquelle vous parlez peut-être tous les jours — n'est **rien de tout cela**.

- Elle n'a pas de corps.
- Elle n'a pas d'objectifs propres.
- Elle ne “veut” rien.
- Elle ne vous “comprend” pas au sens où vous comprenez votre ami.
- Elle ne “sait” pas vraiment ce qu'elle dit.

Alors qu'est-ce que c'est ?

C'est un programme qui prédit des mots.

C'est tout. Vraiment. Quand vous posez une question à ChatGPT, le programme calcule, statistiquement, quel est le mot le plus probable après votre question. Puis le mot suivant. Puis le suivant. Jusqu'à former une phrase, un paragraphe, une réponse.

C'est extraordinairement simple comme principe. C'est extraordinairement puissant comme résultat. Et c'est extraordinairement trompeur comme expérience.

Ce livre ne décrit pas le langage comme le ferait un linguiste, mais comme un citoyen doit le comprendre pour ne pas confondre maîtrise du texte et compréhension du monde.

Pourquoi “générationnelle” ?

On parle d'**IA générative** pour une raison précise : elle *génère* du contenu nouveau.

L'IA "d'avant" — celle qui existe depuis des décennies — était essentiellement **analytique** ou **prédictive** : - Détecter des spams dans vos emails - Recommander des films sur Netflix - Reconnaître des visages sur vos photos - Prédire si vous allez rembourser un crédit - Diagnostiquer une maladie à partir d'une radio

Ces IA classent, trient, prédisent. Elles analysent des données existantes pour en extraire des patterns.

L'IA générative fait autre chose : elle **crée**. - Du texte (ChatGPT, Claude) - Des images (DALL-E, Midjourney, Stable Diffusion) - De la musique (Suno, Udio) - Du code (GitHub Copilot) - De la vidéo (Sora, Runway)

C'est cette capacité de création qui a tout changé en 2022. Soudain, l'IA n'était plus seulement un outil d'analyse invisible. Elle était un *interlocuteur*, un *créateur*, un *assistant* capable de produire ce qui n'existait pas avant.

1.2 Une brève histoire : de Turing à ChatGPT

L'intelligence artificielle n'est pas née en 2022. Elle a 70 ans.

1950 — Le test de Turing

Alan Turing, le mathématicien qui a cassé le code Enigma pendant la Seconde Guerre mondiale, pose une question simple : "Les machines peuvent-elles penser?"

Sa réponse est pragmatique : si une machine peut converser avec un humain sans que celui-ci puisse distinguer si c'est une machine ou un autre humain, alors on peut dire qu'elle "pense" — au moins du point de vue de l'observateur.

Ce test a obsédé les chercheurs pendant des décennies. Mais il pose un problème fondamental qui persiste aujourd'hui : **on mesure l'intelligence sans la définir**.

Le test de Turing ne dit pas ce qu'est l'intelligence. Il dit seulement comment la *détecter* de l'extérieur. C'est comme définir la santé par "si le médecin ne trouve rien, vous êtes en bonne santé". Pratique, mais insuffisant.

Cette confusion — mesurer plutôt que définir — hante encore le domaine de l'IA. Quand on dit qu'un modèle "performe" sur un benchmark, on ne dit pas qu'il est intelligent. On dit qu'il réussit un test. Ce n'est pas la même chose.

Et c'est crucial à comprendre : **même si une machine vous bat aux échecs, au Go, ou à n'importe quel jeu, elle n'est pas pour autant "plus intelligente" que vous**. Elle est meilleure à *ce jeu précis*. Votre calculatrice aussi est meilleure que vous en calcul. Ça ne la rend pas intelligente.

1956 — Naissance officielle

Le terme "intelligence artificielle" est inventé lors d'une conférence à Dartmouth. Les chercheurs sont optimistes : ils pensent qu'en 20 ans, les machines seront aussi intelligentes que les humains.

Ils se trompent de 50 ans. Au moins.

1960-1990 — Les montagnes russes

L'IA traverse des cycles d'euphorie et de désillusion. On appelle les périodes de déprime les "hivers de l'IA". Les promesses ne sont pas tenues. Les financements se tarissent. Mais la recherche continue.

1997 — Deep Blue bat Kasparov

Un ordinateur d'IBM bat le champion du monde d'échecs. C'est un choc médiatique. Mais Deep Blue ne "réfléchit" pas — il calcule des millions de positions par seconde. Force brute, pas intelligence.

2011 — Watson gagne à Jeopardy !

Un autre système IBM bat les champions du jeu télévisé américain. Il comprend les questions en langage naturel, avec leurs jeux de mots et leurs subtilités. Impressionnant, mais encore très limité.

2012 — L'ère du Deep Learning commence

Un réseau de neurones "profond" (avec plusieurs couches) écrase la compétition ImageNet — un concours de reconnaissance d'images. L'écart est stupéfiant.

C'est quoi le Deep Learning ?

Le "Deep Learning" (apprentissage profond) est une technique où l'on empile plusieurs couches de "neurones" artificiels. Chaque couche apprend à reconnaître des patterns de plus en plus abstraits : - Couche 1 : détecte des bords et des contours - Couche 2 : combine les bords en formes simples - Couche 3 : combine les formes en objets - Et ainsi de suite...

Avant 2012, on pensait que les réseaux profonds étaient trop difficiles à entraîner. Trois ingrédients ont changé la donne : 1. **Beaucoup plus de données** (Internet) 2. **Beaucoup plus de puissance de calcul** (GPU) 3. **Quelques astuces mathématiques** (dropout, ReLU, etc.)

Le Deep Learning va révolutionner la reconnaissance d'images, la reconnaissance vocale, la traduction automatique. C'est l'IA "analytique" — celle qui classe, détecte, prédit.

Mais pour le langage, ça coince. Les réseaux de l'époque (RNN, LSTM) ont du mal avec les textes longs. Ils "oublient" le début de la phrase quand ils arrivent à la fin.

2016 — AlphaGo bat Lee Sedol

L'événement qui change tout. Le jeu de Go est trop complexe pour la force brute — il y a plus de positions possibles que d'atomes dans l'univers. AlphaGo, de DeepMind (Google), utilise des réseaux de neurones et apprend en jouant contre lui-même.

Il fait des coups que les experts humains ne comprennent pas. Et il gagne.

2017 — L'attention est tout ce dont vous avez besoin

Un article de Google introduit l'architecture "Transformer". Titre original : *Attention Is All You Need*. C'est la base technique de tout ce qui va suivre.

Presque personne ne le remarque en dehors des cercles de recherche.

2018 — GPT : le modèle pré-entraîné

OpenAI publie GPT (Generative Pre-trained Transformer). Décortiquons ce nom :

— **Generative** : Le modèle *génère* du texte (il crée, il ne classe pas)

- **Pre-trained** : Il est *pré-entraîné* sur d'énormes quantités de texte avant toute utilisation
- **Transformer** : Il utilise l'architecture Transformer (le mécanisme d'attention)

Le "Pre-trained" est l'innovation clé. Avant, il fallait entraîner un modèle spécifiquement pour chaque tâche. Avec GPT, on entraîne d'abord un modèle généraliste sur tout Internet, puis on peut l'utiliser directement ou l'affiner (fine-tuning) pour des tâches spécifiques.

C'est comme la différence entre former quelqu'un de zéro pour chaque job, et embaucher quelqu'un qui a déjà une culture générale étendue.

2020 — GPT-3

OpenAI publie GPT-3, un modèle de langage avec 175 milliards de paramètres. Il peut écrire des articles, du code, des poèmes. Certains sont émerveillés. D'autres sont terrifiés. La plupart n'en ont jamais entendu parler.

30 novembre 2022 — ChatGPT

OpenAI rend GPT-3.5 accessible au grand public via une interface simple : un chat. N'importe qui peut essayer. Gratuitement.

En 5 jours, un million d'utilisateurs. En 2 mois, 100 millions.

Le monde découvre l'IA.

2023-2024 — La course aux modèles

Google répond avec Gemini. Anthropic lance Claude. Meta publie Llama en open source. Mistral émerge en France. Les modèles deviennent multimodaux (texte + image + audio). La puissance augmente. Les prix baissent.

2024 — Le "Thinking" : quand l'IA réfléchit avant de répondre

OpenAI lance o1, puis o3. Nouveauté : le modèle "réfléchit" avant de répondre.

Concrètement, au lieu de générer directement la réponse mot après mot, le modèle produit d'abord une longue "chaîne de pensée" interne (chain of thought) où il décompose le problème, explore des pistes, vérifie sa logique.

Vous voyez "Thinking..." pendant quelques secondes (parfois minutes), puis la réponse arrive.

Pourquoi c'est important ?

Les LLM classiques sont des "répondeurs réflexes" — ils génèrent immédiatement. Le mode Thinking ajoute une étape de raisonnement délibéré. Les résultats sur les problèmes complexes (maths, logique, code) s'améliorent significativement.

Ce que ça ne change pas :

- Le modèle n'a toujours pas de vraie compréhension
- Il peut toujours halluciner (même après réflexion)
- La "pensée" reste une simulation statistique, pas un raisonnement conscient

C'est comme la différence entre répondre du tac au tac et prendre le temps de réfléchir. Mieux, mais pas fondamentalement différent.

2024-2025 — L'ère des Agents

Nouveau buzzword, nouvelle promesse : les “Agents” (*définition complète au chapitre 10*).

Un LLM seul répond à des questions. Un **Agent logiciel** peut *agir* : naviguer sur le web, envoyer des emails, réserver des billets, modifier des fichiers, exécuter du code.

L'idée : au lieu de vous donner une réponse, l'IA accomplit la tâche pour vous.

Où en est-on vraiment ?

Les agents actuels sont impressionnants en démo, fragiles en pratique : - Ils peuvent se perdre dans des boucles - Ils font des erreurs difficiles à rattraper - Ils n'ont pas de vraie “compréhension” de ce qu'ils font - Ils manquent de garde-fous fiables

C'est le “state of the art” fin 2025 : des assistants qui peuvent agir, mais qu'il faut surveiller de près. La promesse d'une IA vraiment autonome reste... une promesse.

Le paradoxe de la performance sans intelligence

Résumons la course actuelle : - L'IA bat les champions du monde aux échecs ✓ - L'IA bat les champions du monde au Go ✓ - L'IA réussit les examens de médecine ✓ - L'IA réussit les examens du barreau ✓ - L'IA bat les humains aux olympiades de mathématiques ✓ - L'IA code mieux que la plupart des développeurs ✓

Et pourtant... elle n'est pas “intelligente”.

Pourquoi ?

Parce que la seule métrique qu'on sait mesurer, c'est la **performance sur des tâches**. On crée des benchmarks, des concours, des examens. L'IA les réussit de mieux en mieux. Donc on en crée de plus difficiles. Elle les réussit aussi.

Mais réussir un test n'est pas comprendre.

Un étudiant qui obtient 20/20 en apprenant par cœur les réponses n'a pas compris le cours. Il a optimisé pour le test. C'est exactement ce que fait l'IA — à une échelle astronomique.

Ce qui manque :

L'IA peut...	L'IA ne peut pas...
Résoudre des problèmes de maths olympiques	Comprendre <i>pourquoi</i> une solution est élégante
Passer un examen de médecine	Ressentir l'inquiétude d'un patient
Écrire du code fonctionnel	Savoir <i>si</i> ce code devrait exister
Battre tout le monde aux échecs	Se demander si le jeu en vaut la chandelle
Générer des réponses parfaites	Savoir quand il vaut mieux se taire

Le problème fondamental :

On mesure ce qu'on sait mesurer. Et on sait mesurer la performance, pas l'intelligence.

L'intelligence humaine inclut : - La conscience de soi - La capacité de douter - Le jugement moral - La sagesse de s'abstenir - L'expérience vécue du monde

Aucun benchmark ne mesure ça. Donc l'IA ne l'optimise pas. Donc l'IA ne l'a pas.

La course à la performance est la seule course qu'on sache organiser. Ce n'est pas la course à l'intelligence.

1.3 Pourquoi novembre 2022 a tout changé

ChatGPT n'était pas une rupture technologique. GPT-3 existait depuis deux ans. Les chercheurs savaient ce que ces modèles pouvaient faire.

Ce qui a changé, c'est **l'accessibilité**.

Avant novembre 2022, pour utiliser un LLM (Large Language Model — grand modèle de langage), il fallait : - Savoir programmer - Avoir accès à une API - Payer pour chaque requête - Comprendre comment formuler des "prompts"

Après novembre 2022, il suffisait de : - Taper une question - Appuyer sur Entrée

C'est la différence entre l'électricité dans les laboratoires et l'électricité dans les maisons. La technologie existait. Ce qui manquait, c'était l'interrupteur.

Et soudain, tout le monde a pu allumer la lumière.

Les étudiants ont découvert qu'ils pouvaient faire rédiger leurs dissertations. Les développeurs ont découvert qu'ils pouvaient faire écrire leur code. Les écrivains ont découvert un partenaire de brainstorming infatigable. Les entreprises ont découvert un assistant qui ne dort jamais.

Et tout le monde s'est posé la même question : **"Qu'est-ce que c'est que ce truc ?"**

1.4 Les mots qui embrouillent : intelligence, apprentissage, neurone

Le problème avec l'IA, c'est le vocabulaire. On utilise des mots humains pour décrire des processus qui n'ont rien d'humain.

"Intelligence"

Quand on dit "intelligence artificielle", on pense à notre intelligence. La capacité de comprendre, de raisonner, de ressentir, de créer du sens.

L'IA ne fait rien de tout ça.

Elle manipule des symboles selon des règles statistiques. C'est incroyablement utile. Ce n'est pas de l'intelligence au sens où vous êtes intelligent.

Un meilleur terme serait "simulation statistique de comportement intelligent". Mais c'est moins accrocheur.

Mais a-t-on seulement défini l'intelligence ?

C'est là que le bât blesse. Même pour l'intelligence humaine, on n'a pas de définition consensuelle.

L'intelligence individuelle — qu'est-ce que c'est ?

On a multiplié les "quotients" pour tenter de la mesurer :

Quotient	Mesure	Ce que l'IA en a
QI (Intellectuel)	Logique, abstraction, raisonnement	Simule bien — réussit les tests
QE (Émotionnel)	Empathie, gestion des émotions, relations	Simule les mots — pas les émotions
QS (Social)	Lire les situations, s'adapter aux autres	Pas de "situation" réelle
QC (Créatif)	Innovation, pensée divergente	Recombine — ne crée pas vraiment
QF (Financier)	Gestion de l'argent, risque, opportunité	Pas de notion de valeur propre
QA (Adversité)	Résilience face aux difficultés	Pas de difficulté vécue
QSp (Spirituel)	Sens, transcendance, connexion au tout	Néant

L'IA actuelle excelle sur le QI (tests de logique, maths). Elle *simule* du QE (répond avec "empathie"). Mais elle n'a aucune des autres formes d'intelligence — parce qu'elle n'a pas d'expérience vécue, pas de corps, pas de relations, pas de risque, pas de sens.

Le problème : on a optimisé l'IA pour le QI, parce que c'est ce qu'on sait mesurer. Et on s'étonne qu'elle soit "intelligente" sur les tests mais "stupide" dans la vie réelle.

L'intelligence collective — encore plus floue : - Une fourmilière est-elle "intelligente"? Aucune fourmi ne l'est individuellement - Un marché financier? Il produit des prix "optimaux"... et des crises absurdes - Wikipédia? Des millions de contributions, un résultat remarquable - Internet lui-même?

L'IA actuelle n'est ni intelligente individuellement (pas de "soi" qui comprend), ni collectivement (c'est un modèle unique, pas une émergence de multiples agents). Elle *simule* des comportements qui *ressemblent* à de l'intelligence.

"Apprentissage" — Learning

Quand on dit qu'une IA "apprend", on pense à un enfant qui apprend à faire du vélo. Essais, erreurs, compréhension, maîtrise.

L'IA n'apprend pas comme ça.

Elle ajuste des millions de paramètres numériques pour minimiser une fonction d'erreur. C'est de l'optimisation mathématique. Pas de l'apprentissage au sens humain.

Attention : deux sens complètement différents

Le “Learning” de **Machine Learning** n’a rien à voir avec le “Learning” de l’école.

Learning (école)	Learning (Machine Learning)
Un élève, un professeur	Des données, un algorithme
Comprendre le <i>pourquoi</i>	Optimiser une fonction
Poser des questions	Pas de questions
Faire des erreurs et en tirer des leçons	Minimiser une métrique d’erreur
Motivation, curiosité, ennui	Aucun état interne
Progression consciente	Ajustement de paramètres
Savoir qu’on a appris	Ne sait pas qu’il a “appris”
Pouvoir enseigner à son tour	Ne peut pas expliquer comment il fait

À l’école, apprendre *c’est devenir quelqu’un qui sait*. En Machine Learning, “apprendre” *c’est ajuster des poids jusqu’à ce que la sortie soit correcte*.

Un enfant qui apprend à lire *comprend* que les lettres représentent des sons. Un modèle qui “apprend” à lire *corrèle* des patterns de pixels avec des labels.

Le même mot. Deux réalités incomparables.

Ce que “apprendre” implique vraiment :

L’apprentissage humain	Le “learning” de l’IA
Savoir qu’on ne sait pas	Ne sait pas qu’elle ne sait pas
Poser des questions	Attend qu’on lui pose des questions
Reformuler un problème mal posé	Répond même aux questions absurdes
Avoir un but (goal)	Pas de but propre — optimise une fonction
Avoir une méta-cognition (réfléchir sur sa réflexion)	Pas de niveau “méta” authentique
Mémoriser sélectivement	Tout est dans les poids, rien n’est “choisi”
Transférer à des domaines nouveaux	Limité au domaine d’entraînement
Désapprendre des erreurs	Très difficile à corriger

Un enfant qui apprend à faire du vélo *sait* qu’il apprend. Il *sent* le progrès. Il *veut* réussir. Il *demande* de l’aide.

L’IA ne fait rien de tout ça. Elle optimise.

“Language” — Langage

Et le “L” de LLM ? Large *Language* Model.

Le mot “langage” suggère communication, interprétation, compréhension mutuelle.

Ce que le langage humain implique : - Un émetteur qui *veut* dire quelque chose
- Un récepteur qui *interprète* le sens - Un contexte partagé - Des intentions, des sous-entendus, de l’ironie - La possibilité de malentendu (et de le résoudre)

Ce que fait un LLM : - Prédire le token suivant - Pas d’intention - Pas de “vouloir dire”
- Pas de compréhension du contexte réel - Simulation de langage, pas langage

Quand vous parlez à ChatGPT, vous communiquez. Lui génère des tokens probables.

“Neurone”

Les “réseaux de neurones” n’ont rien à voir avec votre cerveau.

Le terme vient d’une analogie lointaine avec les neurones biologiques, proposée dans les années 1940. Mais un “neurone artificiel” est juste une fonction mathématique simple. Il n’y a pas de biologie là-dedans.

Votre cerveau a environ 86 milliards de neurones, connectés par des trillions de synapses, qui fonctionnent par signaux chimiques et électriques, formés par des millions d’années d’évolution.

Un réseau de neurones artificiel a des matrices de nombres qui se multiplient entre elles.

Ce n’est pas la même chose.

“Hallucination”

Quand l’IA invente des faits avec assurance, on dit qu’elle “hallucine”. Le terme suggère un dysfonctionnement, comme si l’IA devait normalement dire la vérité et parfois dérapait.

C’est l’inverse.

L’IA génère toujours des séquences de mots statistiquement probables. Parfois, ces séquences correspondent à des faits réels. Parfois non. L’IA ne fait pas la différence. Elle ne “sait” pas ce qui est vrai.

L’hallucination n’est pas un bug. C’est le fonctionnement normal du système.

“Explainable AI” (XAI) — L’IA explicable

Encore un terme trompeur. Comme “intelligence”, “learning”, “hallucination”... le mot “explain” suggère quelque chose qu’il ne désigne pas.

Ce que “expliquer” signifie pour un humain :

Quand un médecin explique son diagnostic, il : - Comprend *pourquoi* il pense ce qu’il pense - Peut justifier chaque étape de son raisonnement - Adapte son explication à son interlocuteur - Peut répondre à “mais pourquoi?” en profondeur - Sait ce qu’il ne sait pas - Peut dire “je me suis trompé parce que...”

Ce que fait l’XAI en réalité :

L’XAI ne fait pas “expliquer” l’IA. Elle produit des **indicateurs techniques** sur le fonctionnement du modèle.

Technique XAI	Ce qu’elle fait	Ce que ça n’est PAS
LIME	Montre quels inputs ont influencé la sortie localement	Une explication du “pourquoi”
SHAP	Attribue une contribution à chaque variable	Une compréhension du raisonnement
Attention maps	Montre où le modèle “regarde” dans l’image	Une intention ou un but

Technique XAI	Ce qu'elle fait	Ce que ça n'est PAS
Saliency maps	Colore les pixels importants pour la décision	Une justification
Counterfactuals	Dit "si X avait été Y, le résultat aurait changé"	Une explication causale

Le problème fondamental :

Explication humaine :

"Je pense que c'est un chat parce qu'il a des oreilles pointues, des moustaches, et qu'il miaule. Les chiens n'ont pas ces caractéristiques."

"Explication" XAI :

"Les pixels 234-267 et 890-912 ont contribué positivement à la classe 'chat' avec un score de saillance de 0.73"

Ce n'est pas la même chose.

L'XAI répond à "quels inputs ont influencé la sortie ?" — pas à "pourquoi cette décision est-elle correcte ?".

La confusion des niveaux :

Niveau	Question	Ce qu'il faudrait	Ce que l'XAI donne
Technique	Comment le modèle calcule-t-il ?	Mécanisme interne	[OK] Partiellement
Sémantique	Qu'est-ce que le modèle "comprend" ?	Représentation du sens	[X] Non
Intentionnelle	Pourquoi le modèle fait-il ce choix ?	But, objectif	[X] Non (pas de but)
Justificatif	Cette décision est-elle bonne/juste ?	Jugement de valeur	[X] Non

L'XAI opère au niveau technique. On l'appelle "explicable" comme si elle opérait au niveau justificatif.

Pourquoi c'est dangereux ?

1. **Fausse impression de compréhension** : "L'IA a expliqué sa décision" → on fait confiance
2. **Illusion de contrôle** : "On peut vérifier l'IA" → on déploie dans des contextes critiques
3. **Déresponsabilisation** : "L'IA a ses raisons" → on ne cherche plus les vraies causes

4. **Validation circulaire** : L'XAI dit que le modèle regarde les "bonnes" features
→ on conclut qu'il "comprend"

Un meilleur vocabulaire :

Terme marketing	Terme honnête
Explainable AI "L'IA explique sa décision"	Inspection de modèle "On visualise quels inputs ont influencé la sortie"
"IA transparente"	"IA dont certains mécanismes sont observables"
"L'IA justifie son choix"	"On génère une rationalisation post-hoc"

Ce que l'XAI peut vraiment faire (et c'est utile) :

- Détecter si un modèle utilise des features problématiques (biais)
- Vérifier que le modèle "regarde" les bonnes régions d'une image
- Identifier des comportements inattendus
- Aider au debugging technique
- Satisfaire des exigences réglementaires (RGPD, AI Act)

Mais ce n'est pas de l'explication. C'est de l'**inspection**.

Le parallèle avec les autres termes trompeurs :

Mot	Sens humain	Sens IA	Écart
Intelligence	Comprendre, ressentir, juger	Prédire des patterns	Abyssal
Learning	Comprendre pourquoi, progresser consciemment	Ajuster des poids	Abyssal
Explain	Justifier, comprendre ses propres raisons	Montrer des corrélations	Abyssal
Hallucination	Perception sans objet réel	Génération normale	Inversé

Le vocabulaire de l'IA est un champ de mines sémantique. Chaque mot familier cache une réalité technique très différente.

Et les ontologies ? Le web sémantique ?

Avant les LLM, il existait une autre vision de l'IA : celle de la **connaissance structurée**.

L'idée : au lieu de laisser une machine "deviner" à partir de statistiques, on lui donne une *représentation explicite du monde*. C'est le domaine des **ontologies** et du **Web sémantique**.

C'est quoi une ontologie (en informatique) ?

Une ontologie définit : - Des **classes** (catégories d'objets) : Personne, Ville, Livre - Des **instances** (objets spécifiques) : "Marie Curie", "Paris", "Le Petit Prince" - Des **relations** (liens entre objets) : "est née à", "a écrit", "est capitale de" - Des **propriétés** (attributs) : date de naissance, population, nombre de pages - Des **axiomes** (règles logiques) : "Tout mammifère est un animal"

Exemple concret :

Classe: Scientifique (sous-classe de: Personne)

Instance: Marie_Curie (type: Scientifique)

Relation: Marie_Curie → a_découvert → Polonium

Relation: Marie_Curie → est_née_à → Varsovie

Axiome: Si X a_découvert Y, alors X est un Scientifique

La puissance des règles : déduction de faits

Ce qui distingue vraiment une ontologie d'une simple base de données, c'est la capacité de **raisonnement automatique**. On définit des règles, et la machine *déduit* de nouveaux faits.

Les règles IF-THEN (ou règles d'inférence)

Structure de base :

SI [condition] ALORS [conclusion]

Exemples :

RÈGLE 1: SI X est_parent_de Y ET Y est_parent_de Z
ALORS X est_grand_parent_de Z

RÈGLE 2: SI X est Mammifère
ALORS X est Animal

RÈGLE 3: SI X habite_dans Y ET Y est_dans Pays Z
ALORS X habite_dans Z

RÈGLE 4: SI X a_épousé Y
ALORS Y a_épousé X (symétrie)

Comment ça marche? Le chaînage

Chaînage avant (forward chaining) — partir des faits connus :

Faits connus:

- Pierre est_parent_de Marie
- Marie est_parent_de Lucas

Application de RÈGLE 1:

→ NOUVEAU FAIT DÉDUIT: Pierre est_grand_parent_de Lucas

On n'a jamais écrit explicitement que Pierre est grand-parent de Lucas. On a défini **une seule règle**, et la machine l'applique **automatiquement à toutes les instances**.

S'il y a 10 000 personnes dans la base, la machine déduit tous les grands-parents sans qu'on ait à écrire 10 000 faits à la main. C'est la puissance de la déduction : **définir une fois, appliquer partout**.

Chaînage arrière (backward chaining) — partir d'une question :

Question: Est-ce que Pierre est_grand_parent_de Lucas ?

Le système cherche une règle qui pourrait conclure "X est_grand_parent_de Z"
→ Trouve RÈGLE 1: SI X est_parent_de Y ET Y est_parent_de Z...

→ Vérifie: Pierre est_parent_de qui? → Marie
 → Vérifie: Marie est_parent_de Lucas? → OUI
 → RÉPONSE: OUI (avec explication du raisonnement)

Exemple médical complet

ONTOLOGIE:

- Classe: Maladie
- Classe: Symptôme
- Classe: Médicament
- Relation: présente_symptôme
- Relation: contre_indiqué_avec

RÈGLES:

R1: SI patient présente_symptôme Fièvre
 ET patient présente_symptôme Toux
 ET patient présente_symptôme Fatigue
 ALORS patient possiblement_atteint_de Grippe

R2: SI patient possiblement_atteint_de Grippe
 ET patient prend Anticoagulant
 ALORS ALERTE contre_indication Aspirine

FAITS:

- Jean présente_symptôme Fièvre
- Jean présente_symptôme Toux
- Jean présente_symptôme Fatigue
- Jean prend Warfarine (qui est un Anticoagulant)

DÉDUCTIONS AUTOMATIQUES:

1. Jean possiblement_atteint_de Grippe (par R1)
2. ALERTE: Ne pas prescrire Aspirine à Jean (par R2)

La différence fondamentale avec un LLM

Aspect	Raisonnement ontologique	LLM
Mécanisme	Déduction logique (modus ponens)	Prédiction statistique
Traçabilité	Chaque conclusion a une preuve	"Boîte noire"
Garantie	Si les règles sont vraies, les conclusions sont vraies	Aucune garantie
Explicabilité	"Parce que règle R1 + faits F1, F2"	"Parce que c'est probable"
Nouveauté	Ne peut déduire QUE ce qui découle des règles	Peut "inventer" (halluciner)

Pourquoi c'est crucial ?

Dans certains domaines, on ne veut PAS de créativité statistique : - **Médecine** : Une contre-indication médicamenteuse doit être certaine - **Droit** : Un raisonnement juridique doit être traçable - **Finance** : Une règle de conformité doit être vérifiable - **Aviation** : Un système de sécurité ne doit pas “improviser”

C’est là que les ontologies et les règles restent irremplaçables : elles offrent des **garanties logiques** qu’un LLM ne peut pas fournir.

Les standards du Web sémantique

Acronyme	Signification	Rôle
RDF	Resource Description Framework	Format de base pour décrire des faits (sujet-prédicat-objet)
RDFS	RDF Schema	Vocabulaire pour définir des classes et hiérarchies
OWL	Web Ontology Language	Langage riche pour des ontologies complexes avec raisonnement
SPARQL	SPARQL Protocol and RDF Query Language	Langage de requête pour interroger des données RDF

La vision originelle (2001-2010)

Tim Berners-Lee, inventeur du Web, a rêvé d’un “Web sémantique” où : - Les machines comprendraient *le sens* des pages, pas juste leur texte - Les données seraient liées et interrogeables - Un agent intelligent pourrait raisonner sur ces données

C’était l’IA “symbolique” — fondée sur la logique, la représentation explicite, le raisonnement formel.

Pourquoi ça n’a pas conquis le monde ?

Problème	Explication
Coût de création	Construire une ontologie complète demande des experts, beaucoup de temps
Rigidité	Une ontologie est figée — le monde réel change constamment
Passage à l’échelle	Difficile de couvrir tout le savoir humain en triplets RDF
Ambiguïté du langage	Le langage naturel est flou — les ontologies sont précises

Problème	Explication
Effet “bootstrap”	Sans données, pas d’usage — sans usage, personne ne crée les données

Alors, c’est mort ?

Non ! Les ontologies sont partout — mais invisibles.

Où les ontologies sont encore essentielles :

Domaine	Utilisation
Médecine	SNOMED CT (terminologie médicale), ICD (classification des maladies)
Sciences	Gene Ontology (biologie), ChEBI (chimie)
Google	Knowledge Graph (le panneau à droite des résultats)
Wikidata	La base de connaissances structurée derrière Wikipédia
Entreprises	Graphes de connaissances internes, catalogues produits
IA juridique	Ontologies du droit, taxonomies réglementaires

LLM vs Ontologies : deux philosophies

Aspect	Ontologie / Web sémantique	LLM (ChatGPT, Claude...)
Connaissances	Explicite, structurée, vérifiable	Implicite, dans les poids, statistique
Vérité	Définie par des faits et des règles	Pas de notion de vérité — probabilités
Raisonnement	Déductif, logique, traçable	Prédiction de tokens, opaque
Mise à jour	Ajouter/modifier des faits	Réentraîner le modèle (coûteux)
Ambiguïté	Mal gérée	Bien gérée (statistiquement)
Création	Manuelle, experte	Automatique à partir de textes
Explicabilité	Totale	Très limitée

La convergence actuelle : le meilleur des deux mondes ?

La tendance “state of the art” en 2025 :

1. **RAG + Knowledge Graphs** : On utilise un graphe de connaissances (ontologie) comme source de vérité, et le LLM pour formuler les réponses en langage naturel.
2. **Extraction d’ontologies par LLM** : Le LLM aide à *construire* des ontologies à partir de textes non structurés.
3. **Grounding / Ancrage** : On “ancrer” les réponses du LLM dans des faits vérifiables issus d’une base de connaissances.

4. **Neuro-symbolique** : Combinaison de réseaux de neurones (statistiques) et de raisonnement symbolique (logique).

Ce qu'il faut retenir :

- Les LLM sont impressionnants mais opaques et sujets aux hallucinations
- Les ontologies sont précises mais rigides et coûteuses à créer
- L'avenir est probablement dans leur **complémentarité**
- Quand quelqu'un vous dit que "l'IA sait tout", demandez : *sait* au sens ontologique (fait vérifié) ou *prédit* au sens statistique (probable) ?

L'aspect logique de l'intelligence

L'intelligence humaine utilise plusieurs formes de raisonnement. Les comprendre permet de mieux situer ce que l'IA fait — et ne fait pas.

Les trois formes de raisonnement

Type	Mouvement	Exemple	Garantie
Déduction	Du général au particulier	"Tous les hommes sont mortels. Socrate est un homme. → Socrate est mortel."	Certaine (si les prémisses sont vraies)
Induction	Du particulier au général	"Ce cygne est blanc. Ce cygne aussi. Et celui-là. → Tous les cygnes sont blancs."	Probable (mais peut être fausse — cygnes noirs)
Abduction	De l'observation à l'explication	"La pelouse est mouillée. → Il a probablement plu."	Plausible (mais d'autres explications existent — arrosage)

La déduction : la logique formelle

Prémisse majeure : Tous les A sont B

Prémisse mineure : X est un A

Conclusion : X est B

C'est le raisonnement des mathématiques, du droit, des ontologies. Si les prémisses sont vraies et la forme est correcte, la conclusion est **garantie**.

Les systèmes à règles (IF-THEN) utilisent la déduction. C'est pourquoi leurs conclusions sont traçables et vérifiables.

L'induction : généraliser à partir d'exemples

Observation 1 : Le soleil s'est levé aujourd'hui

Observation 2 : Le soleil s'est levé hier

Observation n : Le soleil s'est levé tous les jours observés

Généralisation : Le soleil se lèvera demain

C'est le raisonnement de la science empirique, des statistiques... et du Machine Learning.

Un LLM "apprend" par induction : il observe des millions d'exemples et en extrait des patterns. Mais l'induction n'offre aucune garantie — elle peut toujours être réfutée par un contre-exemple.

L'abduction : trouver la meilleure explication

Observation : Le patient a de la fièvre et tousse

Hypothèse : Il a probablement la grippe
(mais aussi possible : rhume, COVID, pneumonie...)

C'est le raisonnement du médecin, du détective, du scientifique qui formule des hypothèses. On cherche l'explication la plus *plausible*, pas la seule possible.

Les LLM font quelque chose qui *ressemble* à de l'abduction quand ils "raisonnent" — mais c'est en fait une prédiction statistique de ce qui serait une réponse plausible.

Où se situe chaque système ?

Système	Déduction	Induction	Abduction
Ontologies / Règles	[OK] Excellent	[X] Non	[X] Non
Machine Learning classique	[X] Non	[OK] Excellent	[X] Non
LLM	[!] Simule	[OK] Base de l'entraînement	[!] Simule
Humain	[OK] Peut faire	[OK] Fait naturellement	[OK] Fait naturellement

Le problème fondamental des LLM :

Les LLM *simulent* tous les types de raisonnement — mais n'en *pratiquent* vraiment qu'un seul : la prédiction statistique (une forme d'induction).

Quand ChatGPT semble "déduire" une conclusion logique, il ne fait pas de déduction formelle. Il prédit ce qu'un texte de déduction *ressemblerait*. La plupart du temps, ça coïncide. Parfois, ça dérape.

Exemple de dérapage logique :

Humain : "Tous les chats sont des animaux. Félix est un animal.
Donc Félix est un chat ?"

LLM (parfois) : "Oui, Félix est un chat."

Erreur ! C'est un sophisme (affirmation du conséquent).
Félix pourrait être un chien.

Un système déductif formel ne ferait *jamais* cette erreur. Un LLM peut la faire — parce qu'il ne raisonne pas, il prédit.

Pourquoi c'est important ?

Contexte	Type de raisonnement nécessaire	LLM adapté ?
Preuve mathématique	Déduction stricte	[!] Risqué
Diagnostic médical	Abduction + déduction	[!] À vérifier
Analyse de données	Induction	[OK] Bon
Rédaction créative	Aucun formel	[OK] Excellent
Raisonnement juridique	Déduction à partir de règles	[!] Dangereux seul
Conversation générale	Mélange informel	[OK] Excellent

Le vrai test de l'intelligence logique :

L'intelligence logique humaine inclut : - Savoir *quel type* de raisonnement utiliser - Reconnaître quand une conclusion est *certaine* vs *probable* vs *plausible* - Détecter les sophismes et les erreurs de raisonnement - Savoir quand on ne peut pas conclure

Un LLM ne distingue pas ces nuances. Pour lui, tout est prédiction. Une conclusion "certaine" et une conclusion "plausible" sont générées de la même façon — avec la même assurance trompeuse.

1.5 Intelligence : une hiérarchie de niveaux

Pour comprendre ce que l'IA fait — et surtout ce qu'elle ne fait pas — il faut d'abord clarifier ce qu'on entend par "intelligence".

L'intelligence n'est pas un bloc unique. C'est une montée en profondeur, avec des niveaux distincts.

Niveau 1 — Cognition : l'intelligence opérationnelle

Capacité à : - percevoir des informations - former des représentations - raisonner, planifier, optimiser - produire des actions cohérentes

→ C'est l'intelligence fonctionnelle — le "savoir faire". → Les systèmes IA actuels sont principalement ici.

Niveau 2 — Métacognition : l'intelligence réflexive

Capacité à : - évaluer sa propre performance - détecter l'incertitude et l'erreur - expliquer ses décisions - ajuster ses stratégies

→ C'est le "savoir que l'on fait". → Sans ce niveau, il n'y a ni confiance, ni apprentissage responsable. → C'est ici que commence la différence entre **outil** et **agent**.

Niveau 3 — Conscience : l'intelligence phénoménale

Capacité à : - avoir une expérience subjective - maintenir une continuité vécue du "soi" - être affecté par ses propres états

→ C'est "être celui qui fait". → Ce niveau n'est pas démontré chez les systèmes artificiels. → Il marque la frontière entre traitement et existence vécue.

Niveau 4 — Morale : l'intelligence axiologique

Capacité à : - attribuer une valeur intrinsèque aux actes - juger le juste et l'injuste indépendamment de l'efficacité - assumer la responsabilité de ses décisions

→ C'est "répondre de ce que l'on fait". → Une IA peut simuler des règles morales. Elle ne peut pas être moralement responsable.

Niveau 5 — Sagesse : l'intelligence existentielle

Capacité à : - intégrer cognition, émotion, valeurs et finitude - reconnaître les limites de l'action et du savoir - renoncer à optimiser quand il le faut - agir avec discernement plutôt qu'avec puissance

→ C'est "savoir jusqu'où ne pas faire". → La sagesse est une forme de retenue, pas de performance.

Où se situe un LLM ?

Un LLM seul (ChatGPT, Claude, Gemini...) se situe au **niveau 0** — ce qu'on pourrait appeler la "compétence sans compréhension".

Un LLM est : - un modèle statistique de langage - optimisé pour prédire le prochain token - sans but propre - sans représentation du monde persistante - sans continuité de soi - sans responsabilité

Il ne pense pas. Il **performe**.

Même s'il décrit un raisonnement, le raisonnement n'est pas causalement actif. Il peut dire "je peux me tromper" — mais il n'évalue pas réellement sa fiabilité. Ce sont des phrases apprises, pas des jugements internes.

Quand un LLM participe-t-il à l'intelligence ?

Un LLM peut devenir un *composant* d'un système intelligent uniquement si on ajoute : - des objectifs explicites → début Niveau 1 - une mémoire persistante structurée - une boucle de contrôle / échec / correction - une traçabilité des décisions → Niveau 2

Le LLM reste alors un moteur de langage, pas le siège de l'intelligence.

Configuration	Niveau
LLM seul	Niveau 0 (compétence aveugle)
LLM + Agent + Objectifs	Niveau 1 (cognition)
LLM + Mémoire + Auto-critique	Niveau 2 (métacognition)
Conscience / Morale / Sagesse	Non démontré

La clarification cruciale

[X] Dire "l'IA est intelligente" sans préciser le niveau est une erreur conceptuelle.

[OK] Il faut toujours préciser : intelligence *cognitive*, *réflexive*, *morale*...

C'est cette distinction qui permet de comprendre pourquoi l'IA peut être extraordinairement utile tout en étant fondamentalement limitée.

1.6 Ce que l'IA fait vraiment : prédire le mot suivant

Voici la vérité, aussi simple que surprenante :

Tout ce que fait un LLM comme ChatGPT, c'est prédire le prochain mot.

Quand vous tapez : "La capitale de la France est"

Le modèle calcule : quel mot a la plus forte probabilité de venir ensuite ?

Réponse : "Paris" (avec une probabilité très élevée)

Le modèle ajoute "Paris" à la séquence, puis recommence : après "La capitale de la France est Paris", quel mot vient ensuite ?

Peut-être un point. Peut-être une virgule et une précision. Le modèle choisit, ajoute, recommence.

C'est tout. Il n'y a pas de base de données de faits. Il n'y a pas de raisonnement logique. Il n'y a pas de compréhension du monde.

Il y a un système qui a vu tellement de texte pendant son entraînement qu'il peut produire des séquences de mots qui *ressemblent* à ce qu'un humain informé écrirait.

Comment ça marche en pratique ?

Imaginez que vous avez lu tous les livres du monde, tous les articles de Wikipédia, tous les forums, tous les sites web. Pas pour les comprendre — juste pour mémoriser les patterns. Quels mots viennent souvent après quels autres mots.

Vous remarqueriez que : - "La capitale de la France" est très souvent suivi de "est Paris" - "Il était une fois" est souvent suivi de mots féériques - "Le théorème de Pythagore" est souvent suivi d'explications mathématiques

Maintenant, imaginez qu'on vous demande de compléter des phrases. Vous ne "savez" rien vraiment — vous reproduisez des patterns que vous avez observés.

C'est ce que fait l'IA.

Pourquoi ça marche si bien ?

Parce que le langage humain est incroyablement structuré. Il y a des règles (grammaire), des conventions (style), des patterns (rhétorique). Un système qui capture ces patterns peut produire du texte qui semble sensé.

Et parce que la quantité de données est astronomique. GPT-4 a été entraîné sur des centaines de milliards de mots. À cette échelle, les patterns deviennent très fins, très subtils.

Pourquoi ça échoue parfois ?

Parce que le système ne comprend pas. Il ne vérifie pas. Il ne raisonne pas vraiment.

Si vous lui demandez quelque chose qui n'était pas dans ses données d'entraînement, il va quand même produire une réponse — en combinant des patterns qui lui semblent pertinents. Parfois, ça marche. Parfois, c'est du non-sens.

Et vous ne pouvez pas savoir à l'avance lequel des deux.

Pourquoi chaque réponse est différente ?

Posez la même question deux fois à ChatGPT. Vous n'obtiendrez pas la même réponse. C'est dû à un paramètre appelé **température** qui introduit du hasard :

- **Température = 0** : le modèle choisit toujours le mot le plus probable. Réponses prévisibles.
- **Température élevée** : le modèle explore des mots moins probables. Plus créatif, mais plus risqué.

ChatGPT utilise une température intermédiaire. Il n'y a donc pas UNE réponse à une question — mais une distribution de réponses possibles. Si vous n'aimez pas une réponse, régénérez-la.

La fenêtre de contexte : la limite de la mémoire

Le modèle ne peut "voir" qu'un nombre limité de mots à la fois — sa **fenêtre de contexte** : - GPT-3.5 : ~4 000 tokens - GPT-4 : 32 000 à 128 000 tokens - Claude : jusqu'à 200 000 tokens

Au-delà ? Le modèle "oublie" le début de la conversation. C'est pourquoi les échanges longs deviennent parfois incohérents : le modèle ne se souvient plus de ce que vous avez dit au début.

Ce que vous pouvez faire

1. **Changez votre image mentale.** L'IA n'est pas un robot, pas un cerveau, pas un oracle. C'est un prédicteur de mots très sophistiqué.
2. **Méfiez-vous du vocabulaire.** Quand quelqu'un dit que l'IA "comprend", "pense", ou "sait", demandez-vous ce que ça signifie vraiment.
3. **Gardez votre esprit critique.** Ce n'est pas parce qu'une réponse semble intelligente qu'elle est vraie.

1.7 Le grand glossaire redéfini : les mots de l'IA traduits en honnête

Récapitulons tous les termes qui nous trompent — et ce qu'ils signifient vraiment.

Terme anglais	Ce qu'on imagine	Ce que ça fait vraiment	Traduction honnête
Intelligence	Comprendre, ressentir, juger, avoir conscience	Prédire des patterns statistiques	Simulation de comportement intelligent
Learning	Comprendre pourquoi, progresser consciemment, vouloir apprendre	Ajuster des paramètres pour minimiser une erreur	Optimisation mathématique

Terme anglais	Ce qu'on imagine	Ce que ça fait vraiment	Traduction honnête
Understanding	Saisir le sens profond, pouvoir reformuler, enseigner	Corréler des patterns de tokens	Correspondance statistique
Reasoning	Déduire logiquement, justifier, argumenter	Prédire ce qu'un raisonnement "ressemblerait"	Simulation de raisonnement
Memory	Souvenirs vécus, rappel conscient, oubli sélectif	Poids figés + contexte temporaire	Paramètres statiques + fenêtre de contexte
Knowledge	Savoir vérifié, compréhension des sources	Patterns extraits de données d'entraînement	Corrélations compressées
Interpreting	Comprendre l'intention, le contexte, le non-dit	Calculer la suite probable	Prédiction conditionnelle
Explaining	Justifier ses raisons, comprendre son propre raisonnement	Montrer quels inputs ont influencé la sortie	Inspection technique post-hoc
Creativity	Imagination, originalité, vision nouvelle	Recombinaison de patterns existants	Interpolation statistique
Language	Communication intentionnelle, sens partagé	Prédiction de tokens	Génération de séquences probables
Hallucination	Dysfonctionnement, erreur anormale	Fonctionnement normal sans ancrage factuel	Génération non contrainte
Thinking (o1, o3)	Réflexion consciente, délibération	Génération de chaîne de tokens "raisonnement"	Raisonnement simulé allongé
Agent	Entité autonome avec buts propres	LLM + outils + boucle de contrôle externe	Orchestration automatisée

Ce que ces redéfinitions révèlent :

Tous ces termes ont un point commun : ils empruntent au vocabulaire de l'**expérience consciente** pour décrire des **processus mécaniques**.

Ce n'est pas un hasard. C'est du **marketing conceptuel** — conscient ou non. En utilisant des mots qui évoquent l'humain, on crée une impression de familiarité, de proximité, presque de parenté.

Mais l'écart est abyssal.

Vers un modèle alternatif : MUSIC

Et si, au lieu d'accepter le vocabulaire trompeur de l'industrie, on proposait un **cadre conceptuel** qui reflète notre intuition de ce qu'est vraiment l'intelligence ?

Un modèle simple, lisible, qui permette d'évaluer n'importe quel système — humain ou artificiel :

M.U.S.I.C.

Lettre	Dimension	Question clé
M	Memory	Se souvient-il ? Apprend-il de l'expérience ?
U	Understanding	Comprend-il le sens, le contexte, le pourquoi ?
S	Self-awareness	Sait-il qu'il sait ? Doute-t-il ? A-t-il un "je" ?
I	Intentionality	A-t-il des buts propres ? Veut-il quelque chose ?
C	Conscience	Y a-t-il quelqu'un "à l'intérieur" ?

Et une dimension bonus, peut-être la plus importante :

| **No** | Sait-il dire **Non** ? Refuser ? S'abstenir ? |

Un système vraiment intelligent ne répond pas toujours. Il sait quand : - La question n'a pas de sens - Il ne sait pas (et l'admet) - Répondre serait nuisible - Le silence est préférable

L'IA actuelle répond toujours. C'est son talon d'Achille.

Score rapide :

Système	M	U	S	I	C	Total
Humain	[OK]	[OK]	[OK]	[OK]	[OK]	5/5
LLM (ChatGPT, Claude)	[!]	[X]	[X]	[X]	[X]	0.5/5
LLM + Agent	[!]	[X]	[!]	[!]	[X]	1-2/5
Système rêvé	[OK]	[OK]	[OK]	?	?	?/5

Le modèle américain optimise la **performance**. Le modèle chinois optimise le **contrôle**.

L'Europe pourrait proposer autre chose : **une IA qui vise le MUSIC** — pas pour l'atteindre (peut-être impossible), mais pour savoir **où on en est vraiment**.

Un cadre d'évaluation honnête. Un vocabulaire qui ne ment pas. Une boussole pour le citoyen.

*Qui sera le chef d'orchestre de cette **MUSIC** ?*

1.8 Vers une définition plus complète : une grille de l'intelligence

Si l'IA actuelle ne "coche" que quelques cases de l'intelligence, quelles sont les cases *qu'on sait identifier* ?

(Cette liste n'est pas exhaustive — elle ne peut pas l'être. L'intelligence humaine reste en partie mystérieuse, même pour nous.)

Quelques dimensions importantes :

Dimension	Description	Humain	LLM seul	LLM + Agent	Système idéal
Perception	Recevoir des informations du monde	[OK]	[!] (texte/image)	[!]	[OK]
Représentation	Former un modèle interne du monde	[OK]	[X] (pas de modèle persistant)	[!]	[OK]
Raisonnement déductif	Conclure avec certitude à partir de règles	[OK]	[!] (simule)	[!]	[OK]
Raisonnement inductif	Généraliser à partir d'exemples	[OK]	[OK] (c'est sa base)	[OK]	[OK]
Raisonnement abductif	Trouver la meilleure explication	[OK]	[!] (simule)	[!]	[OK]
Mémoire épisodique	Souvenirs d'événements vécus	[OK]	[X]	[!] (externe)	[OK]
Mémoire sémantique	Connaissances générales	[OK]	[!] (dans les poids)	[!]	[OK]
Mémoire de travail	Manipulation active d'informations	[OK]	[!] (contexte limité)	[!]	[OK]
Apprentissage continu	Apprendre de nouvelles choses en permanence	[OK]	[X] (figé après entraînement)	[!]	[OK]
Métacognition	Savoir ce qu'on sait et ne sait pas	[OK]	[X]	[!]	[OK]
Intentionnalité	Avoir des buts propres	[OK]	[X]	[!] (buts externes)	[OK]
Conscience de soi	Expérience subjective du "je"	[OK]	[X]	[X]	?
Émotions	États affectifs qui guident l'action	[OK]	[X] (simule les mots)	[X]	?

Dimension	Description	Humain	LLM seul	LLM + Agent	Système idéal
Embodiment	Corps physique dans le monde	[OK]	[X]	[!] (robots)	[OK]
Interaction sociale	Comprendre les autres esprits	[OK]	[!] (simule)	[!]	[OK]
Jugement moral	Distinguer le bien du mal, assumer	[OK]	[X] (pas de responsabilité)	[X]	?
Créativité authentique	Créer du vraiment nouveau	[OK]	[!] (re-combine)	[!]	?
Sagesse	Savoir quand ne pas agir	[OK]	[X]	[X]	?

Score approximatif : - Humain : 18/18 - LLM seul : 3-4/18 - LLM + Agent : 6-8/18 - **Système "idéal"** : ? (certaines cases sont peut-être impossibles pour une machine)

1.9 Roadmap : vers une IA plus complète

Que faudrait-il pour construire une IA qui "coche plus de cases" ?

Niveau 1 — Ce qui est en cours (2024-2026)

Capacité manquante	Solution en développement	État
Mémoire persistante	RAG, vector stores, knowledge graphs	[JAUNE] En progrès
Raisonnement formel	Intégration de moteurs logiques (neuro-symbolique)	[JAUNE] Recherche active
Apprentissage continu	Fine-tuning incrémental, RLHF continu	[JAUNE] Expérimental
Métacognition	Calibration d'incertitude, refus de répondre	[JAUNE] Début
Embodiment	Robots + LLM (Figure, Tesla Bot)	[JAUNE] Proto-types

Niveau 2 — Ce qui est difficile (2026-2030 ?)

Capacité manquante	Défi	Obstacle
Modèle du monde persistant	Représentation interne cohérente	Pas de consensus sur l'architecture
Buts propres	Motivation intrinsèque	Questions de sécurité (alignement)
Raisonnement causal	Comprendre cause \neq corrélation	Fondamentalement différent du pattern matching
Transfert généralisé	Appliquer à des domaines inconnus	Généralisation hors distribution

Niveau 3 — Ce qui est peut-être impossible (ou indésirable)

Capacité	Pourquoi c'est problématique
Conscience	On ne sait même pas ce que c'est ni comment la détecter
Émotions authentiques	Nécessite peut-être un corps, une mortalité, des enjeux
Jugement moral autonome	Voulons-nous vraiment des machines qui décident du bien et du mal ?
Sagesse	La sagesse vient de l'expérience de la finitude

La question fondamentale :

Faut-il même essayer de cocher toutes les cases ?

Une IA "complète" — si elle était possible — poserait des questions vertigineuses :
- Aurait-elle des droits ? - Serait-elle responsable de ses actes ? - Pourrions-nous l'éteindre ? - Serions-nous ses créateurs... ou ses geôliers ?

1.10 La "mort" du développeur : mutation d'un métier

Le développeur logiciel est le premier métier intellectuel massivement impacté par les LLM. Ce qui lui arrive préfigure ce qui arrivera à d'autres.

Ce que l'IA sait déjà faire (2025)

Tâche	Niveau de l'IA	Impact
Écrire du code basique	[VERT] Excellent	Fonctions simples, CRUD, boilerplate
Compléter du code	[VERT] Excellent	GitHub Copilot, Cursor, etc.
Traduire entre langages	[VERT] Très bon	Python → JavaScript, etc.

Tâche	Niveau de l'IA	Impact
Débugger des erreurs simples	[VERT] Très bon	Stack traces, erreurs courantes
Écrire des tests unitaires	[JAUNE] Bon	Tests basiques, couverture
Refactoriser du code	[JAUNE] Bon	Patterns connus
Architecturer un système	[ORANGE] Limité	Suggestions génériques
Comprendre le contexte métier	[ROUGE] Faible	Pas de compréhension réelle
Prendre des décisions de design	[ROUGE] Faible	Pas de jugement
Négocier avec les stakeholders	[ROUGE] Nul	Pas d'interaction humaine

Ce qui va disparaître

Le “développeur-traducteur” — celui qui traduit des spécifications en code ligne par ligne — est une espèce en voie d'extinction.

Avant (2020):

Chef de projet → Spécifications → Développeur → Code → Tests → Déploiement

Bientôt (2026-2028):

Chef de projet → Prompt → IA → Code → Vérification humaine → Déploiement

Les tâches condamnées : - Écrire du code “commodity” — le code “basique” qu'on retrouve dans toutes les applications : - **CRUD** : Create, Read, Update, Delete — les 4 opérations de base sur une base de données (ajouter, lire, modifier, supprimer) - **Formulaires** : pages web où l'utilisateur remplit des champs - **API standards** : interfaces pour que des programmes communiquent entre eux

(Notez comment les développeurs adorent transformer des choses simples en acronymes intimidants. “J'ai codé le CRUD de l'API REST avec du GraphQL” = “J'ai fait un formulaire qui enregistre des données”. Le jargon, c'est aussi une barrière à l'entrée. Et c'est précisément là que les LLM excellent : traduire du jargon en action. Pas étonnant que ce soit les premières tâches automatisées.)

- Faire du copier-coller-adapter depuis Stack Overflow (le forum où les développeurs trouvent des solutions toutes faites — le secret le mieux gardé de l'industrie)
- Maintenir du code ancien sans valeur stratégique
- Écrire de la documentation technique basique
- Faire des revues de code superficielles

Ce qui va rester (et prendre de la valeur)

Compétence	Pourquoi l'IA ne peut pas la remplacer
Comprendre le problème métier	L'IA ne sait pas ce que le client veut vraiment
Questionner les spécifications	L'IA ne dit pas "cette demande n'a pas de sens"
Architecture de systèmes complexes	Décisions avec compromis, contexte, histoire
Sécurité et fiabilité	Responsabilité, conséquences, risques
Dettes techniques	Jugement sur quand rembourser, quand ignorer
Leadership technique	Faire grandir une équipe, trancher, assumer
Éthique du code	"On peut le faire, mais doit-on le faire?"

Le nouveau métier : "Ingénieur en systèmes augmentés"

Le développeur de demain ne sera plus celui qui écrit du code. Ce sera celui qui :

1. Orchestre les IA

Ancien workflow:

Développeur écrit 500 lignes de code

Nouveau workflow:

Développeur → prompt l'IA → revoit le code généré →
corrige les erreurs → assemble les morceaux →
valide la cohérence → teste les edge cases

2. Garantit ce que l'IA ne peut pas garantir

L'IA génère...	L'humain garantit...
Du code qui compile	Du code qui fait ce qu'on veut
Du code qui passe les tests	Du code qui gère les cas non testés
Du code "propre" en apparence	Du code maintenable à long terme
Une solution technique	La bonne solution au bon problème

3. Prend les décisions que l'IA ne peut pas prendre

- "On utilise cette dépendance ou on code nous-mêmes?"
- "On optimise maintenant ou on livre d'abord?"
- "Ce bug est-il critique ou acceptable?"
- "Cette feature vaut-elle le coût de maintenance?"

Les nouvelles compétences essentielles

Compétence	Description
Prompt engineering avancé	Savoir formuler des demandes précises, itérer

Compétence	Description
Revue de code IA	Détecter les erreurs subtiles, les hallucinations de code
Architecture système	Concevoir des systèmes où l'IA est un composant
Spécification formelle	Décrire précisément ce qu'on veut (l'IA exécute)
Pensée critique	Ne pas faire confiance aveuglément
Compréhension métier	Le "pourquoi" que l'IA ne peut pas deviner
Communication	Traduire entre humains et systèmes
Éthique technique	Savoir dire non

Le paradoxe de la productivité

Avec l'IA, un développeur peut produire 10x plus de code. Mais :

Plus de code \neq Plus de valeur

Le risque : **générer massivement du code médiocre.**

Les faux dieux : Productivité, Performance, Compétitivité

Ces trois mots dominent le discours sur l'IA. Mais posons-nous la question :

Mot-clé	Ce qu'on célèbre	Ce qu'on sacrifie
Productivité	Faire plus, plus vite	Faire <i>bien</i> , faire <i>juste</i>
Performance	Battre les benchmarks	Comprendre, expliquer, douter
Compétitivité	Gagner contre les autres	Coopérer, partager, ralentir

Ces valeurs ne sont pas mauvaises en soi. Mais quand elles deviennent les *seules* métriques, on perd de vue : - La **qualité** (vs la quantité) - Le **sens** (vs l'optimisation) - Le **bien commun** (vs l'avantage concurrentiel) - La **durabilité** (vs la croissance) - L'**humain** (vs le système)

L'IA amplifie ce qu'on optimise. Si on optimise uniquement productivité/performance/compétitivité on aura des systèmes très efficaces à produire... quoi exactement ? Pour qui ? Pourquoi ?

La question européenne :

Les États-Unis optimisent la *performance*. La Chine optimise le *contrôle*.

Et si l'Europe proposait d'optimiser autre chose ? - La **sobriété** plutôt que la croissance infinie - L'**explicabilité** plutôt que la performance opaque - Le **bien commun** plutôt que l'avantage concurrentiel - La **dignité humaine** plutôt que la productivité

Ce n'est pas naïf. C'est un **choix de civilisation**.

L'opportunité : **se concentrer sur ce qui compte vraiment.**

Ce que cela signifie pour les développeurs actuels

Si vous êtes...	Ce qu'il faut faire
Junior (0-3 ans)	Apprenez les fondamentaux ET l'IA. Ne soyez pas "celui qui sait prompter" — soyez "celui qui comprend ce que l'IA génère"
Confirmé (3-8 ans)	Montez en abstraction. Architecture, systèmes distribués, sécurité. L'IA ne remplace pas l'expérience
Senior (8+ ans)	Votre valeur est dans le jugement, pas le code. Mentorat, décisions techniques, vision produit
Lead/Architecte	Vous êtes irremplaçables... si vous comprenez l'IA. Sinon, vous serez contournés

Le vrai danger

Ce n'est pas que l'IA remplace les développeurs.

C'est que les développeurs **qui utilisent l'IA** remplacent ceux **qui ne l'utilisent pas**.

Et ensuite, les **non-développeurs qui utilisent l'IA** remplacent les développeurs basiques.

La question à se poser :

"Qu'est-ce que je fais que l'IA ne peut pas faire?"

Si la réponse est "écrire du code"... il est temps d'évoluer.

Si la réponse est "comprendre le problème, prendre des décisions, assumer la responsabilité"... vous avez de l'avenir.

Le parallèle historique

Révolution	Métier "menacé"	Ce qui s'est passé
Métier à tisser (1800)	Tisserand	Devenu opérateur de machine, puis designer
Calculatrice (1970)	Comptable	Devenu analyste financier
Excel (1985)	Employé de bureau	Devenu analyste, gestionnaire
Internet (1995)	Agent de voyage	Devenu conseiller spécialisé (ou disparu)
IA générative (2023)	Développeur	Deviendra... ?

L'histoire montre que les métiers ne disparaissent pas — ils **mutent**. Mais la mutation est douloureuse pour ceux qui ne s'adaptent pas.

Pourquoi apprendre ? C'est une question de survie.

Ce livre n'est pas un luxe intellectuel. C'est un **kit de survie**.

Comprendre l'IA en 2025, c'est comme : - Savoir lire en 1900 - Savoir utiliser un ordinateur en 1995 - Savoir naviguer sur Internet en 2005

Ceux qui ne l'ont pas fait ont été marginalisés. Pas immédiatement. Progressivement. Puis définitivement.

La fenêtre de temps est courte.

Année	Ce qui se passe
2023	Les early adopters expérimentent
2024	Les entreprises commencent à intégrer
2025	Les métiers se reconfigurent
2026-2027	Les non-adaptés deviennent "legacy"
2028+	Nouvelle normalité — trop tard pour rattraper

Ce n'est pas de l'alarmisme. C'est de l'observation.

Les trois postures possibles :

Posture	Attitude	Résultat probable
Déni	"C'est une mode, ça passera"	Obsolescence progressive
Panique	"L'IA va tout détruire"	Paralysie, opportunités manquées
Adaptation	"Je comprends et j'évolue"	Survie, puis avantage

Ce que "apprendre l'IA" veut dire :

Ce n'est PAS : - Devenir data scientist - Savoir coder des réseaux de neurones - Comprendre les mathématiques des transformers

C'est : - Savoir ce que l'IA peut et ne peut pas faire (ce chapitre) - Savoir l'utiliser efficacement (chapitres suivants) - Savoir quand ne pas lui faire confiance - Savoir comment elle va changer votre métier - Savoir quelles questions poser

Le vrai privilège aujourd'hui :

Ce n'est plus l'accès à l'information — tout le monde l'a. Ce n'est plus l'accès aux outils — ils sont gratuits ou peu chers.

C'est **la capacité de comprendre** ce qu'on utilise.

Les analphabètes du XXI^e siècle ne seront pas ceux qui ne savent pas utiliser l'IA. Ce seront ceux qui **l'utilisent sans comprendre ce qu'elle fait**.

1.11 Le rôle de l'Europe : une troisième voie

Dans la course mondiale à l'IA, trois visions s'affrontent.

Les trois modèles

Modèle	Philosophie	Priorité	Risque
États-Unis	Innovation maximale, marché libre	Performance, croissance, domination	Course incontrôlée, inégalités, monopoles
Chine	Contrôle étatique, souveraineté	Surveillance, puissance nationale	Autoritarisme, pas de contre-pouvoir
Europe	Régulation, droits fondamentaux	Protection des citoyens, éthique	Décrochage technologique, dépendance

Ce que l'Europe a fait

- **RGPD (2018)** : Premier cadre mondial sur les données personnelles
- **AI Act (2024)** : Première législation mondiale sur l'IA
- **Classification par risque** : Interdit / Haut risque / Limité / Minimal
- **Droits des citoyens** : Transparence, recours, non-discrimination

Ce que l'Europe pourrait faire

L'Europe n'a pas les géants (OpenAI, Google, Meta, Anthropic). Mais elle a quelque chose qu'ils n'ont pas : **la légitimité démocratique pour définir les règles**.

Une vision européenne de l'IA :

Principe	Implication concrète
IA comme outil, pas comme entité	Refuser le vocabulaire anthropomorphique trompeur
Transparence obligatoire	Savoir quand on parle à une IA, comment elle fonctionne
Explicabilité réelle	Pas juste XAI technique — justification compréhensible
Responsabilité humaine	Toujours un humain responsable des décisions
Souveraineté des données	Données européennes traitées selon les règles européennes
IA au service du bien commun	Santé, éducation, environnement — pas juste profit

Le défi européen : ne pas être naïf

Réguler sans innover = dépendance technologique Innover sans réguler = course au fond **Réguler ET innover = influence mondiale**

L'Europe doit : 1. **Investir massivement** dans la recherche IA (Mistral, Aleph Alpha, etc.) 2. **Former ses citoyens** à comprendre l'IA (ce livre !) 3. **Exporter ses standards** comme elle l'a fait avec le RGPD 4. **Définir une vision** qui ne soit ni le Far West américain ni le contrôle chinois 5. **Créer des consortiums mondiaux de standardisation**

Le modèle W3C : une leçon d'histoire

En 1994, Tim Berners-Lee a fondé le **W3C** (World Wide Web Consortium) pour standardiser le Web.

Résultat : - HTML, CSS, XML sont des standards ouverts - Aucune entreprise ne “possède” le Web - L’interopérabilité est garantie - L’innovation reste possible dans un cadre commun

Pourquoi ça a marché :

Facteur	W3C
Neutralité	Hébergé par MIT, ERCIM (Europe), Keio (Japon)
Inclusivité	Entreprises + universités + gouvernements
Processus ouvert	Drafts publics, commentaires, consensus
Standards libres	Pas de brevets bloquants, royalty-free

Ce qui manque pour l’IA : un “W3C de l’IA”

Aujourd’hui, l’IA est dominée par des standards de fait (OpenAI API, formats propriétaires) et des initiatives fragmentées.

Initiative actuelle	Limite
Partnership on AI	Club de géants tech — pas de pouvoir normatif
IEEE AI standards	Trop technique, pas assez politique
OECD AI Principles	Principes vagues, pas contraignants
AI Act européen	Régional, pas mondial

Un exemple prometteur : MCP (Model Context Protocol)

En novembre 2024, Anthropic (créateur de Claude) a publié **MCP** — un protocole ouvert pour standardiser la communication entre les LLM et les outils externes.

C’est quoi MCP ?

MCP définit comment un LLM peut : - Se connecter à des sources de données (fichiers, bases de données, APIs) - Utiliser des outils externes (navigateur, terminal, services web) - Maintenir un contexte cohérent à travers les interactions

Avant MCP :

Chaque intégration = code custom pour chaque LLM + chaque outil

Avec MCP :

LLM \leftrightarrow [Protocole standard MCP] \leftrightarrow Outil
N'importe quel LLM N'importe quel outil

Pourquoi c’est important ?

Aspect	Sans standard	Avec MCP
Développement	Refaire pour chaque LLM	Écrire une fois, fonctionne partout
Vendor lock-in	Prisonnier d’un écosystème	Interopérabilité
Innovation	Fragmentée	Composable

Aspect	Sans standard	Avec MCP
Sécurité	Chacun réinvente	Bonnes pratiques partagées

Le paradoxe Anthropic :

Une entreprise privée américaine qui publie un standard ouvert. C'est à la fois : - **Encourageant** : Preuve que l'industrie peut produire des standards - **Inquiétant** : Le standard vient d'un acteur commercial, pas d'un consortium neutre

Si MCP devient dominant, Anthropic aura une influence disproportionnée sur son évolution — même si le protocole est "ouvert".

Ce qu'il faudrait :

Que des initiatives comme MCP soient : - Gouvernées par un consortium multipartite (pas une seule entreprise) - Adoptées et enrichies par la communauté mondiale - Intégrées dans un cadre réglementaire (l'Europe pourrait l'exiger)

Le cas Hugging Face : Model Cards et Open Weights

Un autre exemple de standardisation "par le bas" : **Hugging Face**, startup franco-américaine devenue la plateforme de référence pour les modèles d'IA.

Qu'est-ce qu'une Model Card ?

Une "carte d'identité" du modèle, standardisée par Hugging Face :

MODEL CARD - Exemple

=====

Nom : Llama-3-8B

Créateur : Meta

Licence : Llama 3 Community License

Langues : Anglais (principalement)

Taille : 8 milliards de paramètres

Données d'entraînement :

- Sources : Web crawl, livres, code
- Période : Jusqu'à 2023
- Filtrage : Contenu adulte retiré

Usages prévus :

- Assistants conversationnels
- Génération de texte
- NON recommandé pour : décisions médicales, juridiques

Limites connues :

- Biais culturels occidentaux
- Peut halluciner des faits
- Performance réduite en langues non-anglaises

Évaluation :

- MMLU : 68.4%
- HellaSwag : 82.1%

- ...

Open Weights vs Open Source : la nuance cruciale

Terme	Signification	Ce que vous pouvez faire
Closed source	Rien n'est public	Utiliser via API uniquement
Open weights	Les poids du modèle sont publics	Télécharger, exécuter localement, fine-tuner
Open source	Poids + données + code d'entraînement	Reproduire entièrement, auditer

Attention au "open washing" :

Beaucoup de modèles se disent "open source" alors qu'ils sont seulement "open weights" : - **Llama (Meta)** : Open weights, mais licence restrictive et données non publiques - **Mistral** : Open weights, licence permissive - **GPT-4** : Closed source total - **OLMo (AI2)** : Vraiment open source (poids + données + code)

Pourquoi c'est important ?

Niveau d'ouverture	Reproductibilité	Audit	Souveraineté
Closed source	[X]	[X]	[X] Dépendance totale
Open weights	[!] Partielle	[!] Limité	[OK] Exécution locale possible
Open source	[OK] Totale	[OK] Total	[OK] Indépendance complète

Le rôle de Hugging Face :

Contribution	Impact
Hub de modèles	>500 000 modèles disponibles
Format standardisé	SafeTensors, formats communs
Model Cards	Documentation standardisée
Transformers library	Code unifié pour tous les modèles
Spaces	Démo et déploiement facile
Communauté	Recherche collaborative

Un acteur européen (presque) :

Hugging Face a été fondée à Paris en 2016 par des Français (Clément Delangue, Julien Chaumond). Le siège est maintenant à New York, mais une grande partie de l'équipe reste en France.

C'est un exemple de ce que l'Europe peut produire — et de ce qu'elle perd quand elle ne finance pas assez ses startups.

Les standards dont l'IA a besoin :

Couche	Ce qui existe	Ce qui manque
Protocole d'outils	MCP (Anthropic)	Gouvernance neutre
Format de modèles	ONNX, SafeTensors (HF)	Standard universel obligatoire
APIs	OpenAI API (de facto)	Standard ouvert officiel
Model cards	Hugging Face (de facto)	Format légalement obligatoire
Watermarking	Recherche fragmentée	Standard technique
Évaluation	Benchmarks ad hoc (MMLU, etc.)	Framework d'audit certifié
Sécurité	Pratiques variées	Protocole commun
Licences	Jungle (Apache, Llama, etc.)	Clarification juridique

Ce que l'Europe pourrait créer : le "AIC" — AI Consortium

Un consortium mondial pour standardiser :

Domaine	Standard à créer
Interopérabilité	Format commun pour les modèles, APIs standardisées
Transparence	"Model cards" obligatoires (origine des données, limites, biais connus)
Traçabilité	Watermarking des contenus générés par IA
Évaluation	Benchmarks standardisés, audits indépendants
Sécurité	Protocoles de test, red teaming, signalement de vulnérabilités
Éthique	Lignes directrices pour l'entraînement, le déploiement

Pourquoi l'Europe est légitime :

Atout	Explication
Crédibilité réglementaire	RGPD adopté mondialement de facto
Neutralité perçue	Pas de géant IA européen dominant
Tradition de standardisation	ISO, CEN, ETSI, W3C (ERCIM)
Valeurs démocratiques	Légitimité pour parler de droits et éthique
Marché de 450 millions	Poids économique pour imposer des standards

Le précédent RGPD : "l'effet Bruxelles"

Le RGPD n'était "que" européen. Mais : - Les entreprises mondiales l'ont adopté (trop coûteux d'avoir deux systèmes) - D'autres pays s'en sont inspirés (Brésil, Japon, Californie...) - Il est devenu le standard de facto mondial

L'Europe peut refaire la même chose avec l'IA — mais cette fois, en créant un **consortium** plutôt qu'une simple réglementation.

La différence cruciale :

Approche	Mécanisme	Adoption
Réglementation seule	Contrainte légale	Résistance, contournement
Consortium + Réglementation	Co-construction + contrainte	Adhésion + obligation

Un consortium permet aux acteurs (y compris américains et asiatiques) de **participer** à la définition des standards, plutôt que de les **subir**.

Les alliés potentiels :

- **Canada** : Tradition de multilatéralisme, recherche IA forte (Yoshua Bengio)
- **Japon** : Culture du consensus, G7 AI principes
- **Corée du Sud** : Puissance tech, démocratie
- **Australie, Nouvelle-Zélande** : Alliés naturels sur les valeurs
- **Brésil, Inde** : Poids démographique, intérêt pour des standards non-américains
- **Certaines entreprises US** : Anthropic, certains chercheurs Google/Meta qui veulent des règles

L'enjeu géopolitique :

Sans consortium mondial : - Standards américains (de fait, par domination commerciale) - OU standards chinois (par la route de la soie numérique) - L'Europe subit

Avec consortium mondial initié par l'Europe : - Standards co-construits - Valeurs européennes intégrées - L'Europe influence

Ce que vous pouvez faire :

En tant que citoyen européen : - Soutenir les initiatives de standardisation - Exiger que vos représentants portent ce projet - Comprendre que c'est un enjeu aussi important que le climat

Une IA "à l'européenne" :

- **Sobre** : Efficace énergétiquement, pas de course à la taille
- **Explicable** : On comprend pourquoi elle décide
- **Responsable** : Quelqu'un répond des erreurs
- **Inclusive** : Accessible à tous, pas réservée aux élites
- **Gouvernable** : Les citoyens ont leur mot à dire

Ce n'est pas une IA "moins puissante". C'est une IA **différemment puissante** — au service des humains plutôt que des actionnaires.

1.12 Ce qu'il faut retenir de ce chapitre

Les faits : 1. L'IA actuelle (LLM) prédit des mots. C'est tout. 2. Elle ne comprend pas, n'apprend pas, ne raisonne pas au sens humain. 3. Le vocabulaire de l'IA est systématiquement trompeur. 4. Les ontologies et la logique formelle offrent des garanties que les LLM n'ont pas. 5. L'intelligence a plusieurs dimensions — les LLM n'en couvrent que quelques-unes.

Les questions : 1. Voulons-nous des IA qui cochent *toutes* les cases de l'intelligence? 2. Qui décide de ce que l'IA doit faire et ne pas faire? 3. Quel modèle de société voulons-nous construire avec ces outils?

L'enjeu :

L'IA n'est pas une fatalité technologique. C'est un **choix de société**.

Les Américains ont choisi la croissance maximale. Les Chinois ont choisi le contrôle étatique. Les Européens peuvent choisir autre chose : **une IA au service des citoyens**.

Mais pour faire ce choix, il faut comprendre ce qu'est vraiment l'IA.

C'est l'objet de ce livre.

[Livres] Liens et ressources

Les principaux LLM et leurs créateurs : - [OpenAI](#) — ChatGPT, GPT-4, DALL-E - [Anthropic](#) — Claude - [Google DeepMind](#) — Gemini - [Meta AI](#) — Llama - [Mistral AI](#) — Mistral, Mixtral (startup française)

Plateformes et outils : - [Hugging Face](#) — Hub de modèles, Model Cards, Transformers library - [GitHub Copilot](#) — Assistant de code IA

Standards et protocoles : - [MCP - Model Context Protocol](#) — Protocole de connexion LLM ↔ outils (Anthropic) - [W3C](#) — World Wide Web Consortium (modèle de standardisation) - [ONNX](#) — Format ouvert d'échange de modèles - [SafeTensors](#) — Format sécurisé de stockage de modèles

Réglementation européenne : - [RGPD](#) — Règlement Général sur la Protection des Données - [AI Act](#) — Règlement européen sur l'intelligence artificielle - [CNIL](#) — Guide IA de la CNIL

Initiatives mondiales : - [Partnership on AI](#) — Coalition d'entreprises et ONG - [OECD AI Policy Observatory](#) — Principes et recommandations OCDE - [IEEE AI Standards](#) — Standards techniques

Recherche et éducation : - [arXiv](#) — Prépublications en IA - [Papers With Code](#) — Articles + implémentations - [AI2 - Allen Institute for AI](#) — Recherche ouverte (OLMo)

Web sémantique et ontologies : - [W3C Semantic Web](#) — Standards RDF, OWL, SPARQL - [Wikidata](#) — Base de connaissances ouverte - [Schema.org](#) — Vocabulaire structuré pour le web

Pour aller plus loin : - [AI Index Report](#) — État de l'art annuel (Stanford) - [State of AI Report](#) — Rapport annuel sur l'industrie IA - [Our World in Data - AI](#) — Données et visualisations

Chapitre suivant : Comment fonctionne ChatGPT (version honnête)

Chapitre 2

Forces et faiblesses : ce que l'IA fait vraiment bien (et vraiment mal)

L'IA est extraordinairement utile. Des millions de personnes l'utilisent chaque jour pour de bonnes raisons.

L'IA est aussi profondément limitée. Ces limites ne sont pas des bugs à corriger — ce sont des **caractéristiques fondamentales** de la technologie actuelle.

Ce chapitre vous donne les deux faces de la pièce. Parce qu'un cyber-citoyen informé sait quand faire confiance — et quand se méfier.

PARTIE A : Les super-pouvoirs (réels)

2.1 Rédiger, résumer, reformuler

C'est le cas d'usage numéro un. Et c'est là que les LLM excellent.

Rédiger un premier brouillon

Vous avez un email délicat à écrire ? Un rapport à structurer ? Une lettre de motivation à commencer ?

Le LLM produit un premier jet en quelques secondes. Ce ne sera pas parfait — ce ne sera pas *vous* — mais ce sera un point de départ.

“Écris un email à mon propriétaire pour signaler une fuite d'eau dans la salle de bain. Ton poli mais ferme.”

Résultat : un email structuré, avec les bonnes formules de politesse, que vous pouvez ajuster en 2 minutes au lieu de fixer une page blanche pendant 20 minutes.

Résumer un texte long

Vous avez un rapport de 50 pages ? Un article académique dense ? Un contrat interminable ?

“Résume ce texte en 5 points clés”

Le LLM extrait l'essentiel. Ce n'est pas parfait — il peut manquer des nuances — mais pour un premier survol, c'est redoutablement efficace.

Reformuler pour un autre public

“Réécris ce paragraphe pour un enfant de 10 ans”

Le LLM adapte le vocabulaire, simplifie les structures, trouve des analogies.

Pourquoi ça marche si bien ?

Parce que la rédaction, c'est essentiellement du pattern matching linguistique. L'IA a vu des millions d'emails, de résumés, de reformulations. Elle connaît les patterns.

2.2 Expliquer des concepts complexes

Vous voulez comprendre la relativité générale ? La blockchain ? Les mécanismes de l'inflation ?

Le LLM est un professeur patient qui ne juge jamais.

“Explique-moi la mécanique quantique comme si j’avais 15 ans”

“Je ne comprends pas les taux d’intérêt négatifs. Peux-tu m’expliquer avec un exemple concret ?”

Si vous ne comprenez toujours pas, vous pouvez dire :

“Je n’ai pas compris la partie sur X. Peux-tu réexpliquer différemment ?”

Il ne se lassera jamais. Il ne soupirera pas. Il essaiera une autre approche.

[!] Attention cependant :

L’explication sera fluide et convaincante, mais pas nécessairement exacte dans les détails. Pour apprendre les bases d’un sujet, c’est excellent. Pour une compréhension technique approfondie, **vérifiez avec des sources fiables**.

2.3 Traduire avec nuances

Les traducteurs automatiques existent depuis longtemps. Mais les LLM apportent la **traduction contextuelle** :

“Traduis cette phrase en anglais, mais garde un ton professionnel formel”

“Comment dirait-on cette expression française en anglais américain familier ?”

Le LLM comprend que la traduction n’est pas du mot-à-mot. Il adapte le registre, le ton, les expressions idiomatiques.

Cas d’usage puissant : la relecture

“J’ai écrit cet email en anglais. Peux-tu vérifier que ça sonne naturel pour un natif ?”

Le LLM repère les tournures françaises qui “sonnent bizarre” en anglais, les faux-amis, les maladresses culturelles.

2.4 Coder (avec supervision)

C'est peut-être le super-pouvoir le plus transformateur.

Le LLM écrit du code fonctionnel dans presque tous les langages : Python, JavaScript, SQL, C++...

“Écris une fonction Python qui trie une liste de nombres”

“Crée une requête SQL pour trouver tous les clients qui ont acheté plus de 3 fois ce mois”

Pour les développeurs : Le LLM est un copilote. Il génère du code routinier, suggère des implémentations, explique des erreurs.

Pour les non-développeurs : L'IA démocratise la programmation. Vous pouvez créer un petit script pour automatiser une tâche, même sans savoir coder.

[!] Attention critique :

Le code généré peut contenir des bugs, des failles de sécurité, des inefficacités. **Ne mettez jamais en production du code IA sans l'avoir relu et testé.** C'est un assistant, pas un remplaçant.

2.5 Brainstormer et créer

Le LLM ne se fatigue jamais, ne juge jamais, et génère des dizaines d'idées en quelques secondes.

Brainstorming

“Donne-moi 10 idées de cadeaux pour quelqu'un qui aime la cuisine et les voyages”

“Je cherche un nom pour ma startup. Propose 20 options avec leurs connotations”

Création de contenu

“Écris un haïku sur le métro parisien”

“Invente une histoire courte où un chat devient président”

Déblocage créatif

Vous êtes bloqué sur un projet ? Demandez des directions. Même si les suggestions ne sont pas parfaites, elles peuvent débloquent votre propre créativité.

2.6 Le vrai talent : être un assistant polyvalent

Le véritable super-pouvoir des LLM, c'est leur **polyvalence**.

Un expert humain est excellent dans son domaine, mais limité en dehors. Le LLM est "assez bon" dans presque tous les domaines.

Besoin de : - Rédiger un discours de mariage ? - Comprendre un relevé bancaire ? - Planifier un itinéraire de voyage ? - Créer un programme d'entraînement sportif ? - Expliquer un concept juridique ?

Le LLM peut aider sur tout ça. Pas au niveau d'un expert, mais souvent suffisamment pour vos besoins quotidiens.

C'est un assistant généraliste toujours disponible.

2.7 Voir, entendre, comprendre : le multimodal

Depuis 2023-2024, une nouvelle génération est apparue : les modèles **multimodaux**. Ils traitent plusieurs types de données à la fois.

Modalité	Ce que le modèle peut faire	Exemples
Texte	Lire, écrire	Tous les LLM
Image	Analyser des photos	GPT-4V, Claude 3, Gemini
Audio	Écouter et transcrire	GPT-4o, Whisper
Vidéo	Analyser des séquences	Gemini, GPT-4o

Cas d'usage avec les images :

"Qu'est-ce qu'il y a sur cette photo ?"

"Lis le texte sur cette capture d'écran"

"Identifie cette plante / cet insecte"

Vous pouvez envoyer une photo de votre frigo et demander : "Qu'est-ce que je peux cuisiner avec ça ?"

Comment ça marche ?

L'image ou l'audio est converti en représentation numérique (*embeddings*) que le Transformer traite comme des tokens de texte. C'est la même mécanique d'attention qui relie ce qu'il "voit" à ce qu'il sait dire.

Les limites :

- **Hallucinations visuelles** : Le modèle peut "voir" des choses qui n'existent pas
 - **Texte manuscrit** : Lecture difficile
 - **Vidéo longue** : Coûteux et imparfait
-

PARTIE B : Les faiblesses cachées

2.8 L'hallucination : mentir avec aplomb

C'est le problème numéro un.

L'IA invente des faits, des citations, des références — **avec une assurance totale**. Elle ne dit jamais “je ne sais pas” de façon fiable.

Exemples réels :

- Des avocats américains ont soumis des documents juridiques avec des citations de cas **inventés** par ChatGPT. Les juges n'ont pas apprécié.
- Des étudiants ont rendu des dissertations citant des livres qui **n'existent pas**.
- Des journalistes ont publié des articles avec des statistiques **fabriquées**.

Pourquoi ça arrive ?

L'IA prédit le mot suivant. Elle ne vérifie pas la vérité.

Quand vous demandez “Cite-moi une étude sur X”, l'IA génère une séquence qui *ressemble* à une citation. Titre plausible, auteurs plausibles, journal plausible. Mais elle n'a pas de base de données d'études réelles.

Le piège :

Plus la question est spécifique, plus le risque d'hallucination est élevé. Pour les faits très courants, l'IA est généralement correcte. Pour les détails rares, c'est la loterie.

Comment se protéger :

- Ne jamais citer une source IA sans vérifier qu'elle existe
 - Être particulièrement vigilant sur les chiffres, dates, noms propres
-

2.9 L'incohérence : se contredire sans sourciller

Posez la même question deux fois, formulée différemment. Vous pouvez obtenir des réponses contradictoires.

Exemple :

“Est-ce que le café est bon pour la santé?” → “Le café a de nombreux bienfaits...”

“Est-ce que le café est mauvais pour la santé?” → “Le café présente des risques...”

L'IA ne défend pas une position. Elle génère ce qui est statistiquement attendu après votre question. Si vous orientez vers le positif, elle génère du positif.

L'IA n'a pas de “croyances” cohérentes. Elle n'a pas de modèle du monde qu'elle maintient d'une conversation à l'autre.

Comment se protéger :

- Ne pas prendre une réponse pour une vérité définitive
 - Poser des questions sous différents angles
 - Se méfier si l'IA semble toujours d'accord avec vous (biais de confirmation)
-

2.10 La mémoire de poisson rouge

Chaque nouvelle conversation, c'est comme parler à quelqu'un qui ne vous a jamais rencontré.

L'IA n'a pas de mémoire persistante par défaut.

Elle ne se souvient pas : - De ce que vous lui avez dit la semaine dernière - De vos préférences - De votre contexte personnel

Et même dans une conversation en cours...

La “fenêtre de contexte” est limitée. Après quelques milliers de mots, l'IA “oublie” le début.

Solutions partielles :

Certains outils (ChatGPT avec mémoire activée) stockent des informations entre sessions. Mais c'est limité et imparfait.

2.11 L'absence de bon sens

L'IA peut résoudre des équations complexes et échouer sur des questions qu'un enfant de 5 ans réussirait.

Exemples célèbres :

“J'ai mis des pommes dans un frigo, puis je les ai sorties. Où sont les pommes?” → L'IA peut répondre “dans le frigo” — oubliant la deuxième partie.

“Combien de fois la lettre ‘r’ apparaît dans ‘fraise’?” → L'IA se trompe régulièrement.

Pourquoi ?

Le “bon sens” vient d’une compréhension du monde physique, de l’expérience incarnée. L’IA n’a vu que du texte. Elle n’a jamais tenu une pomme, ouvert un frigo, compté sur ses doigts.

Le paradoxe :

L’IA excelle sur ce qui est bien documenté, et échoue sur ce qui est “évident” mais rarement explicite.

2.12 Le perroquet stochastique

Des chercheurs ont qualifié les LLM de “perroquets stochastiques” : ils reproduisent des patterns statistiques sans comprendre ce qu’ils disent.

Est-ce vrai ?

C’est un débat philosophique non résolu. Mais pragmatiquement :

- L’IA génère des réponses sensées sans compréhension au sens humain
- Elle génère aussi du non-sens parfaitement grammatical
- **Elle ne distingue pas les deux**

Conséquence pratique :

Ne supposez jamais que l’IA “comprend” votre problème. Elle génère des séquences qui *ressemblent* à une réponse pertinente.

2.13 Pourquoi l’IA ne peut pas “vérifier” ses réponses

Vous pourriez penser : “Pourquoi ne vérifie-t-elle pas avant de répondre ?”

Parce qu’elle ne peut pas.

L’IA n’a pas accès à Internet en temps réel (sauf plugins). Elle n’a pas de base de données de faits. Elle n’a pas de mécanisme pour distinguer vrai et faux.

L’illusion de l’auto-correction :

“Es-tu sûr de ta réponse précédente ?” → “Vous avez raison, laissez-moi corriger...”

L’IA peut changer sa réponse simplement parce que vous avez exprimé un doute — **même si sa première réponse était correcte.**

Elle n’a pas de confiance calibrée dans ses propres sorties.

Ce que vous pouvez faire

Pour exploiter les forces :

1. **Utilisez l’IA pour les premiers jets.** Ne partez jamais de zéro.

2. **Posez des questions de suivi.** Le LLM ne se lasse pas.
3. **Demandez plusieurs options.** “Donne-moi 5 façons de formuler cette phrase.”
4. **Supervisez le code.** Copilote, pas pilote automatique.
5. **Expérimentez.** Essayez des choses auxquelles vous n’auriez pas pensé.

Pour éviter les pièges :

1. **Partez du principe que l’IA peut se tromper.** Toujours. Sur tout.
2. **Méfiez-vous de l’assurance.** L’IA écrit avec la même confiance, qu’elle ait raison ou tort.
3. **Ne demandez pas à l’IA de “vérifier”.** Elle ne peut pas vraiment. Vérifiez vous-même.
4. **Gardez vos conversations courtes.** Plus c’est long, plus le risque d’incohérence augmente.
5. **Croisez les sources.** Si un fait compte, vérifiez-le ailleurs.

[Stats] Grille récapitulative

Capacité	Force	Faiblesse	Conseil
Rédiger	★★★★★	Peut manquer de personnalité	Affinez le ton
Résumer	★★★★	Peut manquer des nuances	Vérifiez les points clés
Expliquer	★★★★	Peut simplifier à l’excès	Creusez les détails
Traduire	★★★★	Contexte culturel parfois manqué	Relisez avec un natif
Coder	★★★★	Bugs, failles possibles	Testez rigoureusement
Faits	★★	Hallucinations fréquentes	Vérifiez TOUJOURS
Logique	★★	Erreurs sur le simple	Ne faites pas confiance
Mémoire	★	Oublie rapidement	Sessions courtes

Si une IA vous semble comprendre, c’est que le langage vous trompe : elle maîtrise les formes du discours, pas le monde auquel ces formes renvoient.

Chapitre suivant : Les acteurs — Qui contrôle l’IA ?

Chapitre 3

Les acteurs : qui contrôle l'IA ?

Qui contrôle l'IA que vous utilisez chaque jour ?

Cette question paraît simple. La réponse ne l'est pas.

Derrière ChatGPT, Gemini, Claude et les autres se cache un enchevêtrement d'entreprises, d'investisseurs, de partenariats et de dépendances. Comprendre cette carte, c'est comprendre qui décide de ce que l'IA peut faire — et de ce qu'elle fait de vos données.

[!] Note de temporalité : Les acteurs cités ici sont ceux de 2024-2025. Les rapports de force évolueront — des géants peuvent chuter, de nouveaux entrants émerger. Ce qui ne changera pas : les questions à poser sur la propriété, le contrôle et la transparence.

3.1 OpenAI : de l'utopie au business (Microsoft inside)

L'origine (2015)

OpenAI naît comme une organisation à but non lucratif, fondée par Elon Musk, Sam Altman et d'autres. Mission affichée : développer une IA "sûre et bénéfique pour l'humanité", en open source, pour contrebalancer le pouvoir de Google.

Le virage (2019)

Problème : développer une IA de pointe coûte des milliards. OpenAI crée une filiale "capped-profit" (profit plafonné). Elon Musk quitte le conseil d'administration.

L'alliance Microsoft (2019-2023)

Microsoft investit 1 milliard, puis 10 milliards de dollars. En échange : - Microsoft obtient l'exclusivité commerciale sur les modèles GPT - Microsoft intègre GPT dans tous ses produits (Bing, Office, Azure) - Microsoft devient le fournisseur cloud d'OpenAI

Aujourd'hui

OpenAI est une entreprise privée, valorisée à plus de 80 milliards de dollars, largement contrôlée par Microsoft. La mission "bénéfique pour l'humanité" cohabite avec des objectifs commerciaux massifs.

Ce que ça signifie pour vous :

Quand vous utilisez ChatGPT, vos données transitent par l'infrastructure Microsoft Azure. Vos conversations alimentent potentiellement l'amélioration de produits Microsoft. L'utopie initiale a laissé place à un business model classique.

3.2 Google/DeepMind : le géant qui a tout inventé (et raté ChatGPT)

L'ironie de l'histoire

Google a inventé presque toutes les technologies qui rendent ChatGPT possible :
- Le Transformer (2017) — la base de tous les LLM - BERT (2018) — les premiers modèles de langage massifs - Des décennies de recherche en IA chez Google Brain et DeepMind

Mais quand ChatGPT a explosé, Google a été pris de court.

La panique (2022-2023)

En interne, Google a déclaré un "code red". L'entreprise a précipité le lancement de Bard (devenu Gemini), avec des résultats embarrassants au début.

Aujourd'hui

Google rattrape son retard avec Gemini, qui rivalise avec GPT-4 sur de nombreux benchmarks. Mais l'entreprise a perdu sa position dominante perçue.

L'avantage Google :

Google a quelque chose qu'OpenAI n'a pas : l'accès à vos emails (Gmail), votre agenda (Calendar), vos documents (Drive), votre historique de recherche, vos déplacements (Maps). L'intégration de l'IA dans cet écosystème sera redoutable.

Ce que ça signifie pour vous :

Google sait déjà presque tout de vous. L'IA va amplifier cette connaissance. Gemini intégré à Gmail pourra rédiger des emails en connaissant tout votre historique de correspondance.

3.3 Meta : l'open source comme stratégie

Le pari Llama

Meta (Facebook) a fait un choix surprenant : publier ses modèles de langage (Llama) en open source.

Pourquoi ? Parce que Meta n'a pas le cloud à vendre. À la différence de Microsoft, Google et Amazon, Meta ne gagne pas d'argent en louant de la puissance de calcul.

La stratégie :

- Si tout le monde utilise Llama, personne ne paie Microsoft ou Google pour les LLM
- Meta bénéficie des améliorations faites par la communauté
- Meta garde le contrôle de l'architecture fondamentale

Ce que ça signifie pour vous :

Llama permet à des entreprises et des développeurs de faire tourner des IA localement, sans envoyer vos données aux géants américains. C'est une alternative importante pour la vie privée.

3.4 Anthropic : les dissidents de la sécurité

L'origine

Anthropic a été fondée en 2021 par d'anciens dirigeants d'OpenAI, dont Dario Amodei, qui s'inquiétaient de la direction prise par l'entreprise.

Le positionnement

Anthropic se présente comme l'entreprise de l'IA "responsable". Son modèle Claude est conçu avec des garde-fous plus stricts. L'entreprise publie des recherches sur la sécurité de l'IA.

La réalité

Anthropic a levé des milliards auprès de Google et Amazon. L'entreprise est dans la même course commerciale que les autres, malgré son discours différent.

Claude vs ChatGPT

Claude est souvent considéré comme plus "prudent" — il refuse plus de requêtes, est moins susceptible de générer du contenu problématique. Pour certains, c'est un avantage. Pour d'autres, une limitation.

3.5 Mistral : le champion français (et européen ?)

L'espoir européen

Mistral AI, fondée en 2023 par d'anciens de Google et Meta, est la startup IA française la plus valorisée. En un an, elle est devenue une "licorne" (valorisation supérieure à 1 milliard).

Les modèles

Mistral publie des modèles performants, souvent en open source, qui rivalisent avec des modèles beaucoup plus gros. L'efficacité française face aux milliards américains.

Les limites

Mistral reste 100 fois plus petite qu'OpenAI en termes de ressources. L'entreprise a levé des fonds... auprès d'investisseurs américains et de Microsoft.

Ce que ça signifie pour vous :

Mistral est une bonne nouvelle pour la diversité. Mais l'Europe n'a pas encore de champion véritablement indépendant dans cette course.

3.6 La Chine : le rival invisible

Ce que l'Occident ne voit pas

Pendant que ChatGPT faisait les gros titres, la Chine développait ses propres LLM : - **Qwen** (Alibaba) - **Ernie** (Baidu) - **DeepSeek** - Des dizaines d'autres

Ces modèles ne sont pas disponibles en Europe, mais ils sont utilisés par 1,4 milliard de Chinois.

La guerre des standards

Le monde se dirige vers deux écosystèmes IA parallèles : occidental (dominé par les US) et chinois. Avec des valeurs, des censures et des biais différents.

Ce que ça signifie pour vous :

L'IA que vous utilisez n'est pas neutre. Elle porte les valeurs, les biais et les priorités de ceux qui l'ont créée. L'IA américaine et l'IA chinoise ne donneront pas les mêmes réponses sur Taïwan, le Tibet ou Tiananmen.

3.7 L'Europe : arbitre ou colonie ?

Ce que l'Europe n'a pas

- Pas de LLM de classe mondiale (Mistral essaie, mais reste petit)
- Pas de cloud souverain majeur (AWS, Azure et Google dominent)
- Pas de Big Tech européenne dans l'IA

Ce que l'Europe a

- Le plus grand marché unique du monde (450 millions de consommateurs)
- Une tradition réglementaire forte (RGPD)
- La première loi mondiale sur l'IA (AI Act)

Le choix européen

L'Europe a décidé de jouer la carte de la régulation plutôt que de la production. Le pari : même si nous ne créons pas les modèles, nous pouvons définir les règles du jeu pour ceux qui veulent accéder à notre marché.

Est-ce suffisant ? On en parlera au chapitre suivant.

Ce que vous pouvez faire

1. **Sachez qui est derrière l'outil.** ChatGPT = OpenAI + Microsoft. Gemini = Google. Claude = Anthropic (+ Google/Amazon).
2. **Comprenez les incitations.** Ces entreprises veulent vos données, votre engagement, votre argent. Leurs produits sont conçus pour ça.
3. **Diversifiez.** N'utilisez pas qu'un seul outil. Essayez des alternatives, dont des modèles open source.
4. **Privilégiez les options européennes quand possible.** Mistral, mais aussi des startups plus petites qui émergent.

5. **Lisez les conditions d'utilisation.** (Oui, vraiment. Au moins les parties sur les données.)

Chapitre suivant : La guerre de l'IA — L'Europe a-t-elle perdu ?

Chapitre 4

L'économie de l'IA : suivez l'argent

Pour comprendre l'IA, suivez l'argent.

Qui paie ? Qui gagne ? Pourquoi ChatGPT est-il "gratuit" ? Ces questions révèlent les vraies motivations des géants de la tech — et ce que vous leur donnez sans le savoir.

4.1 Pourquoi ChatGPT est "gratuit" (spoiler : vous êtes le produit)

Ouvrez un compte ChatGPT. Posez des questions pendant des heures. Ne payez rien.

Comment est-ce possible ?

Le modèle freemium

ChatGPT offre une version gratuite pour une raison simple : vous attirer, vous habituer, vous rendre dépendant. Ensuite, vous passerez (peut-être) à la version payante (20€/mois) pour avoir accès aux fonctionnalités premium.

Mais ce n'est pas tout.

Même si vous ne payez jamais, vous êtes rentable. Voici comment :

1. Vos données améliorent le produit

Chaque conversation que vous avez avec ChatGPT peut être utilisée pour améliorer les futurs modèles. Vos questions, vos reformulations, vos corrections — tout cela est de l'or pour l'entraînement.

2. Vous générez des effets de réseau

Plus de gens utilisent ChatGPT, plus ça crée du buzz, plus d'autres gens veulent l'essayer. Vous êtes le marketing.

3. Vous validez le produit

Vos usages montrent ce qui marche et ce qui ne marche pas. Vous êtes le département qualité gratuit.

La vraie économie

Si un service est gratuit, vous n'êtes pas le client. Vous êtes le produit.

C'était vrai pour Google, Facebook, et les réseaux sociaux. C'est vrai pour l'IA.

4.2 Le coût réel d'entraîner un LLM

Créer un modèle comme GPT-4 coûte une fortune.

L'entraînement

- Des dizaines de milliers de GPU (cartes graphiques spécialisées à ~30 000€ pièce)
- Des mois de calcul continu
- Une consommation électrique colossale

Estimation pour GPT-4 : **entre 50 et 100 millions de dollars** rien que pour l'entraînement.

L'inférence

Ensuite, chaque fois que vous posez une question, le modèle doit "tourner" sur des serveurs. C'est l'inférence.

À l'échelle de 100 millions d'utilisateurs, ça représente des millions de dollars par jour en coûts de serveurs.

Le paradoxe économique

OpenAI perd de l'argent sur chaque utilisateur gratuit. L'entreprise brûle des milliards en espérant : - Convertir des utilisateurs en payants - Vendre des API aux entreprises - Prendre une position dominante avant que les coûts baissent

C'est un pari massif. Qui marche tant que les investisseurs y croient.

4.3 L'API : comment les entreprises paient

Le vrai business d'OpenAI n'est pas vous. C'est les entreprises.

L'API

Une API (Application Programming Interface) permet à une entreprise d'intégrer ChatGPT dans ses propres produits. Votre banque, votre assurance, votre service client — beaucoup utilisent déjà l'API d'OpenAI en coulisses.

Le prix

Vous payez par "token" (environ 0,75 mot). Quelques centimes pour une réponse simple. Mais à l'échelle, ça chiffre vite :

- Une entreprise qui traite 1 million de requêtes par jour → des dizaines de milliers d'euros par mois
- Un usage intensif → des millions par an

Le verrouillage

Une fois qu'une entreprise a intégré l'API d'OpenAI dans ses systèmes, il est coûteux de changer. C'est le "lock-in" — le verrouillage.

C'est pourquoi la conquête de parts de marché est si agressive. Chaque entreprise convertie est une source de revenus récurrents pour des années.

4.4 La course aux données : votre texte vaut de l'or

Pour entraîner un LLM, il faut du texte. Beaucoup de texte.

D'où viennent les données ?

- Wikipédia (gratuit, mais limité)
- Livres numérisés (parfois sans l'accord des auteurs)
- Le web entier (aspiré par des "crawlers")
- Les réseaux sociaux (Reddit, Twitter...)
- Et... vos conversations avec l'IA

Le problème du consentement

Beaucoup de ces données ont été collectées sans demander l'avis des auteurs originaux. Des procès sont en cours aux États-Unis et en Europe.

La valeur de vos données

Quand vous corrigez ChatGPT ("Non, ce n'est pas ça, voici la bonne réponse"), vous faites du travail d'annotation gratuit. Ce travail vaut normalement plusieurs dollars de l'heure quand il est fait par des professionnels.

Multipliez par 100 millions d'utilisateurs : c'est une main-d'œuvre gratuite colossale.

4.5 Le cloud comme dépendance stratégique

Le dernier élément du puzzle économique : le cloud.

Où tourne l'IA ?

Pas sur votre ordinateur. Sur des serveurs distants, dans des datacenters géants.

Qui possède ces datacenters ?

- Amazon (AWS) : ~32% du marché mondial
- Microsoft (Azure) : ~23%
- Google (GCP) : ~10%
- Alibaba, autres : le reste

Presque aucun datacenter européen de cette échelle.

La dépendance

Utiliser l'IA, c'est envoyer vos données vers des serveurs américains, gérés par des entreprises américaines, soumises au droit américain.

Le Cloud Act de 2018 permet au gouvernement américain d'exiger l'accès aux données stockées par des entreprises américaines, même si les serveurs sont physiquement en Europe.

Le coût de la dépendance

- Pas de contrôle sur vos données
- Pas de levier en cas de conflit
- Des milliards d'euros qui partent aux États-Unis

Ce que vous pouvez faire

1. **Comprenez le deal.** Quand c'est gratuit, vous payez autrement — en données, en attention, en habitudes.
2. **Lisez les conditions.** Oui, c'est long. Cherchez au moins les sections sur les données.
3. **Désactivez le partage quand possible.** ChatGPT a une option pour ne pas utiliser vos conversations pour l'entraînement. Activez-la.
4. **Préférez les modèles locaux.** Des alternatives comme Llama peuvent tourner sur votre ordinateur, sans envoyer de données.
5. **Soutenez les alternatives européennes.** Même si elles sont imparfaites, elles offrent une diversité nécessaire.

Chapitre suivant : Vos données, leur trésor

Chapitre 5

Vos données, leur trésor

Chaque fois que vous tapez quelque chose dans ChatGPT, où va cette information ? C'est une question que presque personne ne pose. Et c'est un problème.

5.1 Ce qui se passe quand vous tapez dans ChatGPT

Vous écrivez : “Mon fils de 8 ans a des difficultés en maths. Comment l'aider ?”

Voici le trajet de cette information :

1. Transmission

Votre question quitte votre ordinateur, traverse Internet, et arrive sur des serveurs d'OpenAI (hébergés chez Microsoft Azure).

2. Traitement

Le modèle génère une réponse. Votre question et la réponse sont stockées.

3. Conservation

Par défaut, cette conversation est conservée. Pendant combien de temps ? Ça dépend des versions et des politiques — qui changent.

4. Utilisation potentielle

Sauf si vous avez désactivé cette option, votre conversation peut être : - Lue par des employés d'OpenAI (pour améliorer le service) - Utilisée pour entraîner les futurs modèles - Analysée pour détecter des abus

5.2 Où vont vos conversations ?

Physiquement

Vos données sont stockées sur des serveurs Microsoft Azure, principalement aux États-Unis. Des copies peuvent exister dans d'autres régions.

Juridiquement

Ces données sont soumises au droit américain. Cela signifie : - Le Cloud Act permet au gouvernement US d'y accéder - Les tribunaux américains peuvent les demander - La FISA (loi de surveillance) peut s'appliquer

Concrètement

Si vous êtes un citoyen européen lambda, personne ne va probablement fouiller vos conversations sur ChatGPT. Mais le *potentiel* existe.

5.3 Qui peut les lire ? (humains et machines)

Les humains

Oui, des employés d'OpenAI peuvent lire vos conversations. C'est dans les conditions d'utilisation.

Pourquoi ? Pour : - Améliorer le modèle - Détecter les abus - Répondre aux demandes légales

Ce n'est pas de la surveillance active de tout le monde. Mais des conversations sont échantillonnées et examinées.

Les machines

Vos données alimentent des systèmes automatiques : - Détection de contenu problématique - Analyse des patterns d'utilisation - (Potentiellement) Entraînement des futurs modèles

5.4 La différence entre “stocké” et “utilisé pour entraîner”

C'est une distinction cruciale que beaucoup de gens ne comprennent pas.

Stocké

Votre conversation est conservée dans une base de données. Elle peut être retrouvée, relue, analysée.

Utilisé pour entraîner

Votre conversation est incorporée dans les données qui servent à créer la prochaine version du modèle. Elle devient, en quelque sorte, une partie permanente de l'IA.

La différence

- Une donnée stockée peut être supprimée sur demande
- Une donnée utilisée pour l'entraînement est “digérée” — elle influence le modèle, mais n'est plus identifiable comme telle

OpenAI permet de refuser l'entraînement

Depuis 2023, vous pouvez aller dans les paramètres et désactiver l'utilisation de vos conversations pour l'entraînement. Mais c'est opt-out (il faut le faire activement), pas opt-in.

5.5 Comment (essayer de) protéger vos données

Règle n°1 : Ne jamais partager d'informations sensibles

Jamais. Vraiment.

[X] Pas de numéros de carte bancaire **[X]** Pas de mots de passe **[X]** Pas de numéros de sécurité sociale **[X]** Pas de données médicales détaillées **[X]** Pas de secrets professionnels **[X]** Pas de documents confidentiels

Règle n°2 : Désactiver l'entraînement

Sur ChatGPT : Settings → Data Controls → Chat History & Training → Désactiver

Attention : cela désactive aussi l'historique. Vous ne pourrez pas retrouver vos anciennes conversations.

Règle n°3 : Utiliser des comptes non identifiables

Si vous voulez plus d'anonymat, utilisez une adresse email qui ne révèle pas votre identité.

Règle n°4 : Envisager des alternatives locales

Des modèles comme Llama peuvent tourner sur votre propre ordinateur. Rien n'est envoyé à l'extérieur.

5.6 Les options "privacy" : ce qu'elles font vraiment

ChatGPT — Mode sans historique

Quand vous désactivez l'historique : - Vos conversations ne sont pas sauvegardées dans votre compte - Elles ne sont pas utilisées pour l'entraînement - MAIS : elles sont quand même conservées 30 jours "pour abus" avant suppression

ChatGPT — Temporary Chat

Un mode encore plus éphémère, où même la mémoire de session est réduite.

ChatGPT Enterprise / API

Les versions payantes pour entreprises ont des garanties plus fortes : données non utilisées pour l'entraînement, suppression sur demande, parfois conformité RGPD explicite.

Le problème du "trust me"

Tout cela repose sur la confiance. Vous devez croire qu'OpenAI fait ce qu'elle dit. Il n'y a pas de vérification indépendante possible.

Cas concrets qui doivent vous alerter

Le médecin qui demande un diagnostic

Un médecin tape les symptômes d'un patient dans ChatGPT pour avoir une idée de diagnostic. Ces données médicales viennent de partir aux États-Unis.

L'avocat qui demande un conseil

Un avocat copie-colle un contrat confidentiel pour demander une analyse. Ce contrat est maintenant sur les serveurs d'OpenAI.

L'employé qui résume un document interne

Un employé utilise ChatGPT pour résumer un rapport stratégique de son entreprise. Ce rapport est potentiellement accessible à OpenAI.

Tous ces cas sont réels. Ils se produisent tous les jours.

Ce que vous pouvez faire

1. **Avant de taper, demandez-vous : “Est-ce que j’accepterais que cette info soit publique ?”** Si non, ne la mettez pas dans l'IA.
 2. **Désactivez l'entraînement** dans les paramètres de ChatGPT.
 3. **Pour les usages sensibles, utilisez des alternatives locales** ou des versions entreprise avec des garanties contractuelles.
 4. **Formez vos proches.** Beaucoup de gens ne réalisent pas les implications.
 5. **Exercez vos droits RGPD.** Vous pouvez demander à OpenAI quelles données ils ont sur vous, et demander leur suppression.
-

Fin de la Partie I — COMPRENDRE

Vous avez maintenant les bases : le vocabulaire, le modèle MUSIC pour évaluer toute IA, les forces et faiblesses réelles, les acteurs et leurs intérêts, les risques concrets.

La suite est différente : nous passons de *comprendre* à *agir*.

Chapitre suivant : L'Europe et vos droits

PARTIE II : SE PROTÉGER

« Savoir, c'est pouvoir — mais seulement si vous savez quoi faire de ce savoir. »

Ce que vous allez apprendre : - Vos droits concrets (et comment les exercer)
- Les questions à poser avant d'utiliser une IA - Comment évaluer la fiabilité d'un système - Quand utiliser l'IA — et quand s'en passer

Vous pouvez lire ces chapitres dans l'ordre ou aller directement à ce qui vous concerne : vos droits (ch.6), questions pratiques (ch.7), évaluation (ch.8), usages (ch.9).

Chapitre 6

L'Europe et vos droits : la troisième voie

L'Europe n'a pas créé ChatGPT. Elle n'a pas les milliards de dollars des Big Tech. Elle n'a pas les fermes de GPU géantes.

Mais elle a quelque chose que ni les États-Unis ni la Chine ne possèdent : **une vision différente de ce que la technologie doit à l'humain.**

Et surtout : **des droits concrets que vous pouvez exercer dès maintenant.**

PARTIE A : L'Europe dans la guerre de l'IA

6.1 Les chiffres qui font peur

Mettons les choses en perspective.

Investissements privés dans l'IA (2023) : - États-Unis : ~67 milliards de dollars
- Chine : ~8 milliards de dollars - **Europe : ~4 milliards de dollars**

Valorisation des leaders : - OpenAI : ~80 milliards \$ - Anthropic : ~18 milliards \$ - Mistral : ~2 milliards €

Puissance de calcul : - Microsoft/OpenAI : des centaines de milliers de GPU dernière génération - Google : infrastructure cloud colossale - Europe : on cherche encore

Ces chiffres sont brutaux. L'Europe n'est pas dans la course au même niveau.

6.2 Pourquoi ce retard ?

1. Le capital-risque Aux États-Unis, investir 10 milliards dans une startup sans revenus est normal. En Europe, c'est impensable.

2. La fragmentation 27 pays, 27 systèmes fiscaux, 24 langues. Une complexité administrative épuisante.

3. L'aversion au risque La culture européenne valorise la stabilité. La culture américaine valorise le "move fast and break things".

4. Le brain drain Les meilleurs talents sont aspirés par les GAFAM. Salaires 3 à 5 fois supérieurs.

6.3 Ce que l'Europe a (et que les autres n'ont pas)

450 millions de consommateurs Aucune entreprise tech ne peut ignorer ce marché.

Une tradition juridique forte Le RGPD a changé les pratiques mondiales. L'AI Act pourrait faire de même.

Des valeurs différentes Vie privée, protection du consommateur, dignité humaine. Ce ne sont pas des “freins au business” — ce sont des choix de société.

Des chercheurs excellents INRIA, Max Planck, ETH Zurich, Oxford, Cambridge... Le problème n’est pas le talent — c’est la commercialisation.

Une industrie réelle Voitures, avions, médicaments, énergie. L’IA appliquée à ces secteurs pourrait être notre force.

6.4 Le RGPD : premier coup gagnant

En 2018, les Big Tech ont hurlé. “Trop contraignant. Impossible à appliquer.”

Et puis... le monde entier a suivi.

L’effet Bruxelles Plutôt que de maintenir deux systèmes, les entreprises ont choisi d’appliquer les standards européens partout.

Le RGPD est devenu un standard mondial de fait. La Californie s’en est inspirée. Le Brésil aussi.

La leçon L’Europe n’a pas besoin de créer la technologie pour influencer son usage. Elle peut définir les règles que tout le monde doit suivre.

6.5 L’AI Act : le pari de la régulation

En 2024, l’Europe a adopté le premier cadre juridique complet sur l’IA au monde.

Le principe — Classification par risque :

Niveau	Ce qui est concerné	Traitement
Inacceptable	Notation sociale, manipulation subliminale	Interdit
Haut risque	IA médicale, recrutement, justice	Réglémenté strictement
Risque limité	Chatbots, génération d’images	Obligations de transparence
Risque minimal	Jeux, filtres	Liberté

Calendrier d’entrée en vigueur :

Date	Ce qui entre en vigueur
Février 2025	Interdiction des pratiques à risque inacceptable
Août 2025	Règles sur les modèles d’IA à usage général (GPT-4, Claude...)
Août 2026	Majorité des obligations (systèmes à haut risque)
Août 2027	Toutes les obligations

6.6 Nos champions : Mistral, Aleph Alpha, et les autres

Mistral AI (France) La star européenne. Modèles performants, open source, croissance fulgurante. Mais investisseurs largement américains.

Aleph Alpha (Allemagne) IA souveraine pour entreprises et gouvernements. Stratégique.

Silo AI (Finlande) Modèles multilingues pour les langues européennes.

Le problème : Ces entreprises restent petites. Quand elles réussissent, elles sont souvent rachetées par des américains.

6.7 La souveraineté numérique : utopie ou nécessité ?

L'état actuel : - Vos données sont sur des serveurs américains (AWS, Azure, Google Cloud) - Vos emails passent par Gmail, Outlook - Tout cela est soumis au Cloud Act américain

Le risque : En cas de conflit géopolitique, les États-Unis pourraient couper l'accès ou exiger l'accès aux données européennes.

Les initiatives : - **Gaia-X** : projet de cloud européen (avancée lente) - **Règles de localisation** : obligation de stocker certaines données en Europe

La réalité : Pour l'instant, la souveraineté numérique reste un slogan. L'infrastructure réelle est américaine.

PARTIE B : Vos droits concrets (RGPD + AI Act)

6.8 Le RGPD : le bouclier qui existe déjà

Le RGPD s'applique à **toute entreprise** qui traite des données de résidents européens — même si elle est américaine.

Vos droits fondamentaux :

Droit	Article	Ce que ça signifie
Accès	Art. 15	"Quelles données avez-vous sur moi ?"
Rectification	Art. 16	"Cette donnée est fausse, corrigez-la"
Effacement	Art. 17	"Supprimez mes données"
Portabilité	Art. 20	"Donnez-moi mes données dans un format utilisable"
Opposition	Art. 21	"Je refuse que vous utilisiez mes données ainsi"

6.9 Ce que l'AI Act ajoute

[!] Cadre juridique évolutif : L'AI Act est entré en vigueur, mais son interprétation précise sera affinée par la jurisprudence dans les années à venir. Les principes ci-dessous sont largement stables ; les détails d'application évolueront.

Transparence obligatoire Quand vous interagissez avec une IA, vous devez en être informé. Pas de chatbot qui se fait passer pour un humain.

Interdiction de certaines pratiques - Notation sociale : interdit - Reconnaissance faciale de masse : très encadrée - Manipulation subliminale par IA : interdit

Supervision humaine Les systèmes IA "à haut risque" doivent maintenir une supervision humaine significative.

Droit à une explication Si une décision IA vous affecte, vous avez le droit de comprendre comment elle a été prise.

6.10 Tableau récapitulatif : vos droits en action

Situation	Votre droit	Base légale	Action
Je veux savoir quelles données OpenAI a sur moi	Accès	RGPD Art. 15	Email au DPO
Je veux que mes conversations ne servent pas à l'entraînement	Opposition	RGPD Art. 21	Paramètres + demande formelle
Je veux supprimer mon historique ChatGPT	Effacement	RGPD Art. 17	Paramètres compte + demande
Un algorithme m'a refusé un crédit	Explication + intervention humaine	RGPD Art. 22 + AI Act	Demande écrite à la banque
Je ne savais pas que je parlais à un bot	Violation transparence	AI Act	Signalement autorité

6.11 Comment exercer vos droits (mode d'emploi)

Étape 1 : Trouver le bon contact

Chaque entreprise doit avoir un DPO (Data Protection Officer).

- OpenAI : privacy@openai.com
- Google : support.google.com/policies/contact/general_privacy_form

Étape 2 : Formuler votre demande

Objet : Demande d'accès aux données personnelles (RGPD Article 15)

Madame, Monsieur,

En application du RGPD, je vous demande de me fournir l'ensemble des données personnelles me concernant que vous détenez.

Merci de me répondre dans le délai légal de 30 jours.

[Votre nom, email utilisé pour le service]

Étape 3 : Attendre (et relancer)

L'entreprise a 30 jours. Si pas de réponse satisfaisante, passez à l'étape 4.

Étape 4 : Porter plainte

Saisissez votre autorité de protection des données.

Pays	Autorité	Site
France	CNIL	cnil.fr

Pays	Autorité	Site
Allemagne	BfDI	bfdi.bund.de
Belgique	APD	autoriteprotectiondonnees.be
Suisse	PFPDT	edoeb.admin.ch

Est-ce efficace ? Oui. La CNIL a infligé des amendes de plusieurs millions à Google, Amazon, Microsoft.

6.12 Cas pratique : exercer vos droits auprès d'OpenAI

Ce qu'OpenAI sait sur vous : - Votre email - Vos conversations (sauf si désactivé)
- Votre historique de paiement

Comment demander vos données : 1. Connectez-vous à ChatGPT 2. Settings → Data Controls → Export data 3. Vous recevrez un fichier avec vos conversations

Comment demander la suppression : 1. Email à privacy@openai.com 2. Ou : Settings → Delete account

6.13 Pourquoi les Américains nous envient

Aux États-Unis, il n'existe pas d'équivalent fédéral du RGPD.

Ce que les Américains n'ont pas : - Pas de droit général d'accès à leurs données
- Pas de droit à l'effacement garanti - Pas de droit d'opposition au profilage - Pas d'autorité unique de protection

L'ironie : Les géants de la tech sont américains, mais leurs utilisateurs les mieux protégés sont européens.

Ce que vous pouvez faire

En tant qu'utilisateur :

1. **Exercez vos droits au moins une fois.** Demandez vos données à un service. C'est éducatif.
2. **Exportez régulièrement.** Vos conversations ChatGPT, par exemple.
3. **Désactivez ce qui peut l'être.** L'entraînement, la personnalisation.
4. **Gardez les preuves.** Conservez vos demandes et les réponses.
5. **N'hésitez pas à vous plaindre.** La CNIL est là pour vous.

En tant que citoyen :

1. **Choisissez en conscience.** Quand une alternative européenne existe, préférez-la.
2. **Votez.** Les élections européennes déterminent qui écrit les règles.
3. **Parlez-en.** Beaucoup de gens ne savent pas que cette guerre existe.

4. **Exigez mieux.** De vos employeurs, écoles, administrations : “Quelle IA utilisez-vous ? Où vont les données ?”
-

L'Europe : chef d'orchestre ou spectateur ?

L'Europe a raté le premier acte de la révolution IA. C'est un fait.

Mais la partie n'est pas finie. L'IA va se déployer pendant des décennies. Les règles ne sont pas écrites.

La question n'est pas “Avons-nous le meilleur LLM ?”

La question est : **“Qui décide de l'avenir de l'IA ?”**

L'Europe peut encore peser — si elle joue intelligemment. Si elle défend ses valeurs : sobriété, explicabilité, bien commun, dignité humaine.

Face aux États-Unis (performance à tout prix) et à la Chine (contrôle absolu), l'Europe peut proposer une **troisième voie**.

Une voie où la technologie sert l'humain — pas l'inverse.

Chapitre suivant : Questions à poser — aux entreprises, à votre employeur

Chapitre 7

Questions à poser — et réponses à exiger

Vous utilisez l'IA tous les jours. Les entreprises, votre employeur, les services publics la déploient autour de vous.

Mais qui vous a expliqué ce qu'elle fait de vos données ?

Ce chapitre est votre guide de questions. Des questions précises, avec les réponses que vous devriez obtenir — et comment réagir si vous ne les obtenez pas.

[Temps] Si vous n'avez que 2 minutes : - Vos conversations IA *peuvent* être lues par des humains - Versions gratuites = vos données servent à l'entraînement (opt-out possible) - Votre employeur *peut* voir ce que vous faites avec Copilot - Vous avez le droit de savoir et de refuser (RGPD) - Le DPO est votre interlocuteur clé

PARTIE A : Questions aux entreprises tech

7.1 “Mes conversations sont-elles lues par des humains ?”

La vraie réponse : OUI, potentiellement.

OpenAI, Google, et les autres emploient des équipes de “reviewers” qui lisent des échantillons de conversations pour : - Améliorer la qualité des réponses - Détecter les abus - Entraîner les systèmes de modération

Ce n’est pas de la surveillance systématique. Mais vos conversations peuvent être sélectionnées aléatoirement.

Ce que ça signifie : - Ne partagez jamais d’informations que vous ne voudriez pas qu’un inconnu lise - Il n’y a pas de confidentialité absolue

7.2 “Mes données servent-elles à entraîner le modèle ?”

Service	Par défaut	Opt-out possible ?
ChatGPT (gratuit)	OUI	Oui (Settings → Data Controls)
Gemini (gratuit)	OUI	Oui
Claude (gratuit)	OUI	Oui
Copilot (gratuit)	OUI	Limité
Versions payantes/entreprise	Généralement NON	Vérifiez les conditions

La règle générale : Version gratuite = vos données peuvent être utilisées. Version payante/entreprise = généralement non.

7.3 “Où sont stockées mes données ?”

OpenAI → Serveurs Microsoft Azure, principalement aux États-Unis. Soumis au Cloud Act.

Google → Datacenters mondiaux. Stockage Europe possible (pas garanti par défaut).

Microsoft → Azure global. Options de résidence pour les clients entreprise.

Pourquoi c’est important : - Le lieu de stockage détermine quel droit s’applique
- Les données aux US sont accessibles au gouvernement américain (Cloud Act) -
Même si stockées en Europe, une entreprise américaine peut être contrainte de les fournir

7.4 “Que se passe-t-il si je supprime mon compte ?”

La réalité : “Supprimer” ne signifie jamais “effacer instantanément et définitivement”.

- Les données traversent caches, sauvegardes, systèmes d’archivage
- OpenAI conserve des données “désidentifiées” pour l’entraînement
- La suppression complète prend des semaines/mois et n’est jamais 100% garantie

7.5 Checklist : 10 questions avant d’utiliser un outil IA

#	Question	Réponse à chercher
1	Qui est derrière ?	Entreprise, pays, investisseurs
2	Où sont stockées mes données ?	USA, Europe, autre
3	Mes données servent-elles à l’entraînement ?	Oui/Non, opt-out possible ?
4	Des humains peuvent-ils lire mes données ?	Pour modération, amélioration ?
5	Puis-je exporter mes données ?	Format, facilité
6	Puis-je supprimer mes données ?	Comment, délai réel
7	Conformité RGPD ?	DPO, procédures
8	Politique de sécurité ?	Chiffrement, audits
9	Si l’entreprise est vendue ?	Transfert des données
10	Alternatives plus respectueuses ?	Options locales, européennes

PARTIE B : Questions à votre employeur

7.6 “Copilot analyse-t-il mes emails ?”

Microsoft 365 Copilot : OUI, c’est exactement ce qu’il fait.

Copilot a accès à : - Vos emails (Outlook) - Vos documents (OneDrive, SharePoint) - Vos conversations (Teams) - Votre calendrier - Vos notes (OneNote)

La question n’est pas si Copilot lit vos données. La question est : **qui d’autre y a accès ?**

Demandez : - Les résumés Copilot peuvent-ils être vus par votre manager ? - Les données sont-elles utilisées pour évaluer votre performance ?

7.7 “Mon manager peut-il voir ce que je demande à l’IA ?”

Réponse : ça dépend de la configuration.

Techniquement possible : - Voir les logs d’utilisation de Copilot - Analyser les prompts envoyés - Mesurer le “temps productif” vs “temps non-productif” - Alertes sur certains mots-clés

Ce que vous pouvez faire : Demandez explicitement à IT ou RH : 1. Quelles données d’utilisation IA sont collectées ? 2. Qui y a accès ? 3. Ces données sont-elles utilisées dans les évaluations ?

7.8 “L’IA évalue-t-elle ma performance ?”

De plus en plus souvent : OUI.

Les outils RH modernes utilisent l’IA pour : - Analyser les emails et détecter le “sentiment” - Mesurer le temps de réponse moyen - Évaluer la “productivité” algorithmique - Prédire le “risque de départ” - Identifier les “hauts potentiels”

Vos droits (RGPD) : - Vous avez le droit de **savoir** si des décisions automatisées vous concernent - Vous avez le droit de **contester** ces décisions - L’entreprise **doit vous informer** de l’utilisation de l’IA dans les processus RH

7.9 “Mes créations appartiennent-elles à l’entreprise ?”

Le droit du travail français est clair : Ce que vous créez dans le cadre de votre travail appartient à votre employeur.

Cela inclut : - Le code que vous écrivez avec Copilot - Les textes que vous rédigez avec ChatGPT - Les images que vous générez avec DALL-E

La zone grise : Vous utilisez le Copilot de l’entreprise pour un projet perso ? Clarifiez **avant** de vous retrouver dans un conflit.

7.10 Vos droits en tant que salarié

Le RGPD vous protège, même au travail.

L’employeur doit : - Informer sur les traitements de données - Justifier la finalité - Respecter la vie privée (même au travail, il y a des limites)

Le Code du travail ajoute : - Information préalable sur tout dispositif de surveillance (Article L1222-4) - Consultation du CSE avant déploiement d’outils de surveillance

En pratique : Si votre entreprise déploie un outil IA sans vous informer, elle est en infraction.

7.11 Checklist pour les salariés

#	Question	À qui la poser
1	Quels outils IA sont déployés ?	IT, Direction
2	Quelles données sont collectées ?	IT, DPO
3	Qui a accès à mes données d’utilisation ?	IT, DPO
4	Ces données servent-elles à m’évaluer ?	RH
5	Ai-je été informé (RGPD) ?	DPO
6	Le CSE a-t-il été consulté ?	Représentants du personnel
7	Puis-je refuser certains outils ?	RH
8	Y a-t-il une charte IA ?	RH, Juridique

7.12 Le DPO : votre interlocuteur clé

Votre entreprise a (probablement) un DPO (Délégué à la Protection des Données).

C’est la personne à contacter pour : - Toute question sur vos données personnelles - Exercer vos droits RGPD - Signaler un problème de confidentialité

Comment le trouver : - Intranet de l’entreprise - Service juridique ou conformité - Demandez à votre manager

Si pas de DPO identifiable : Signal d'alerte. Les entreprises de +250 salariés sont obligées d'en avoir un.

Modèles de lettres

Aux entreprises tech :

Objet : Demande d'information sur le traitement des données

Madame, Monsieur,

J'utilise votre service [nom] et je souhaite comprendre : 1. Si mes conversations sont utilisées pour entraîner vos modèles 2. Où mes données sont stockées 3. Qui peut y avoir accès 4. Comment demander leur suppression

Cette demande est faite en application du RGPD.

Cordialement, [Votre nom]

À votre employeur :

Objet : Demande d'information sur l'utilisation de l'IA

Bonjour,

Suite au déploiement de [nom de l'outil IA], je souhaiterais comprendre : 1. Quelles données professionnelles sont analysées ? 2. Ces données sont-elles partagées avec Microsoft/OpenAI ? 3. Les insights sont-ils accessibles à ma hiérarchie ? 4. Ces données servent-elles à évaluer ma performance ? 5. Le CSE a-t-il été consulté ?

Je fais cette demande (RGPD, Article 13).

Cordialement, [Votre nom]

Ce que vous pouvez faire

Avec les entreprises tech :

1. **Posez au moins UNE question** à un service que vous utilisez
2. **Gardez la checklist** à portée de main avant d'adopter un nouvel outil
3. **Votez avec vos pieds** — si un service ne respecte pas vos droits, changez

Avec votre employeur :

1. **Identifiez le DPO** de votre entreprise
 2. **Posez les questions AVANT** d'utiliser les outils
 3. **Parlez à vos représentants** — le CSE doit être informé
 4. **Soyez prudent** — n'écrivez rien dans Copilot que vous ne voudriez pas voir dans votre dossier RH
-

Chapitre suivant : Évaluer et se protéger — guide pratique

Chapitre 8

Évaluer et se protéger — guide pratique

Un nouvel outil IA sort chaque semaine. Chacun promet de révolutionner votre vie. Comment savoir lesquels valent le coup ? Comment protéger vos données au quotidien ?

Ce chapitre est votre guide pratique.

PARTIE A : Évaluer un outil IA avant de l'utiliser

8.1 La grille d'évaluation en 5 minutes

Avant d'utiliser un nouvel outil IA, passez-le au crible :

Critère	Question	Score /5
Transparence	Qui est derrière ? Où sont mes données ?	
Utilité réelle	Résout-il un vrai problème que j'ai ?	
Alternatives	Existe-t-il des options sans IA ou plus éthiques ?	
Coût caché	Que dois-je donner en échange ?	
Dépendance	Puis-je m'en passer facilement ?	

Interprétation : - **20-25** : À tester avec précaution - **15-19** : Risques significatifs, réfléchissez à deux fois - **10-14** : Probablement pas une bonne idée - **< 10** : Fuyez

8.2 Qui est derrière cet outil ?

Les informations à chercher :

- **L'entreprise** : Nom, pays, taille, historique
- **Les investisseurs** : Qui finance ? Quels intérêts ?
- **Le modèle économique** : Comment gagnent-ils de l'argent ?
- **Le track record** : Historique de scandales, de failles

Où chercher : - Page "About" ou "À propos" - Crunchbase (investisseurs, levées de fonds) - Recherche Google "[nom] controversy"

Signaux d'alerte : - **[X]** Entreprise opaque, pas d'information sur les fondateurs - **[X]** Financée par des fonds liés à certains États - **[X]** Modèle économique flou ("gratuit" sans explication) - **[X]** Historique de failles de sécurité

8.3 Que font-ils de vos données ?

Dans la politique de confidentialité, cherchez :

Question	Ce que ça révèle
Collecte	Quelles données ? Seulement ce que j'entre ?
Utilisation	Améliorer le service ? Entraîner des modèles ? Publicité ?
Partage	Avec qui ? Partenaires ? "Tiers" ?
Conservation	Combien de temps ?
Droits	Comment accéder, modifier, supprimer ?

Les formulations à repérer : - "Partager avec nos partenaires" → **[!]** Qui sont-ils ? - "Pour améliorer nos services" → **[!]** Inclut l'entraînement ? - "Conformément à la loi applicable" → **[!]** Quelle loi ? - "Modifier cette politique à tout moment" → **[!]** Sans prévenir ?

8.4 Le vrai test : données fictives d'abord

Avant d'y mettre vos vraies informations :

1. **Créez un compte avec une adresse email dédiée**
2. **Utilisez des données fictives** (faux nom, fausse adresse)
3. **Testez avec des contenus non sensibles**
4. **Observez** : publicités ciblées ? Emails ? Notifications ?
5. **Essayez de supprimer votre compte** : facile ou compliqué ?

8.5 Les alternatives à considérer

Avant d'adopter un outil IA commercial :

Question	Exemples d'alternatives
Puis-je faire sans IA ?	Méthodes traditionnelles
Version open source ?	Ollama, LocalAI, Whisper local
Options européennes ?	Mistral, OVHcloud, Scaleway
Auto-hébergement ?	Solutions on-premise
Coopératives ?	Framasoft, Chatons

Le principe de subsidiarité numérique : Utilisez la solution la plus locale, la plus simple, la moins invasive.

PARTIE B : Protéger ses données au quotidien

8.6 L'hygiène numérique de base

Geste	Impact	Difficulté
Gestionnaire de mots de passe	Élevé	Facile
Authentification à deux facteurs	Élevé	Facile
Mettre à jour ses appareils	Élevé	Facile
Emails différents (pro/perso/poubelle)	Moyen	Facile
Vérifier les autorisations des apps	Moyen	Facile
Désactiver la géolocalisation	Moyen	Facile

L'erreur classique : Un excellent mot de passe... utilisé partout. Un seul site piraté = tous vos comptes compromis.

8.7 Paramétrer ses outils IA

ChatGPT (OpenAI) Settings → Data Controls : - **[X]** Désactivez "Improve the model for everyone" - **[OK]** Activez "Temporary Chat" pour les conversations sensibles

Google (Gemini) Mon compte Google → Données et confidentialité : - **[X]** Désactivez "Activité sur le Web et les applications" - **[X]** Désactivez "Personnalisation des annonces"

Microsoft (Copilot) - Vérifiez les autorisations Copilot - Pour entreprise : demandez les paramètres à votre IT

Claude (Anthropic) Settings : - Opt-out de l'entraînement si disponible

8.8 Les données que vous ne savez pas que vous donnez

Au-delà de ce que vous tapez :

- **Métadonnées** : Quand, combien de temps, sur quel appareil
- **Habitudes** : Vos horaires, vos patterns

- **Inférences** : Ce que l'IA déduit de vos questions
- **Contexte** : Fichiers uploadés, images, documents

Exemple :

Vous demandez : “Comment annoncer un divorce à mes enfants ?”

Vous venez de révéler : - Que vous divorcez - Que vous avez des enfants - Votre niveau de détresse (peut-être)

8.9 Le principe du “minimum nécessaire”

Besoin	[X] Mauvaise pratique	[OK] Bonne pratique
Résumer un article	Coller le texte intégral	Donner juste le lien
Rédiger un email pro	Tout le contexte interne	Formuler générique
Analyser des données	Uploader le fichier complet	Anonymiser d'abord
Aide au code	Coller tout le projet	Isoler la fonction

La règle d'or : Avant de soumettre quelque chose à une IA, demandez-vous : “Est-ce que j'accepterais que cette information soit publique ?”

8.10 Les outils de protection

Navigateurs et extensions : - **Brave** : Protection intégrée contre le tracking - **Firefox** + uBlock Origin : Bloque trackers - **DuckDuckGo** : Moteur de recherche sans tracking

Email : - **Proton Mail** : Email chiffré, basé en Suisse - **SimpleLogin** : Alias email

VPN : - **Mullvad** : Axé confidentialité, paiement anonyme - **ProtonVPN** : Du fournisseur de ProtonMail

Gestionnaire de mots de passe : - **Bitwarden** : Open source, gratuit, audité - **1Password** : Payant mais excellent

8.11 Les alternatives locales

Faire tourner l'IA sur votre machine = vos données restent chez vous.

Pour le texte : - **Ollama** : LLM locaux - **LM Studio** : Interface graphique - **GPT4All** : Solution simple pour débutants

Pour les images : - **Stable Diffusion** (local) - **AUTOMATIC1111** : Interface populaire

Prérequis : - Ordinateur récent (GPU recommandé) - Un peu de configuration - Qualité légèrement inférieure aux modèles cloud

Ce que vous pouvez faire

Pour évaluer :

1. **Utilisez la grille** pour chaque nouvel outil
2. **Testez avec des données fictives** avant vos vraies infos
3. **Préférez les alternatives locales** quand possible
4. **Lisez les conditions** (au moins en diagonal)

Pour vous protéger :

1. **Paramétrez vos outils** (10 minutes, une fois)
2. **Appliquez le minimum nécessaire** (habitude à prendre)
3. **Utilisez des outils de protection** (navigateur, email, VPN)
4. **Explorez les alternatives locales** si vous êtes motivé

Chapitre suivant : Quand utiliser — et quand ne pas utiliser — l'IA

Chapitre 9

Quand utiliser — et quand ne pas utiliser — l'IA

L'IA peut beaucoup. Mais elle ne devrait pas être utilisée pour tout.

Ce chapitre vous aide à garder votre libre arbitre — et à savoir quand l'humain fait mieux.

PARTIE A : Les pièges à éviter

9.1 Les dark patterns de l'IA

Les “dark patterns” sont des interfaces conçues pour vous faire faire ce que vous ne vouliez pas faire.

Dans l'IA, ça prend des formes spécifiques :

Le piège de l'engagement - L'IA répond toujours, même quand elle ne devrait pas - Les réponses créent de la dépendance - Le système vous encourage à revenir, à poser plus de questions

L'illusion de la relation - L'IA dit “je” et “vous” comme une vraie conversation - Elle simule de l'empathie (“Je comprends que c'est difficile”) - Elle donne l'impression de vous connaître

La facilité addictive - C'est tellement simple de demander à l'IA - Pourquoi chercher quand on peut demander ? - Pourquoi réfléchir quand l'IA peut réfléchir pour nous ?

9.2 L'effet chambre d'écho

L'IA a tendance à vous donner raison.

Pourquoi ? - Les modèles sont optimisés pour la satisfaction utilisateur - Contredire = mauvaise note = punition dans l'entraînement - L'IA détecte vos préférences et s'y adapte

Le danger : Vous demandez : “Est-ce que j'ai raison de penser que X ?” L'IA répond : “Oui, vous avez de bonnes raisons...” Même si X est faux ou discutable.

Comment contrer : - Demandez explicitement des contre-arguments - “Quels sont les arguments CONTRE ma position ?” - “Joue l'avocat du diable” - Utilisez plusieurs outils pour comparer

9.3 L'autorité artificielle

L'IA parle avec assurance. Même quand elle invente.

Ce ton d'autorité crée un biais : on fait plus confiance à ce qui a l'air sûr de soi.

Le problème : - L'IA ne dit jamais "je ne sais pas" naturellement - Elle formule des incertitudes comme des certitudes - Elle cite des sources fausses avec le même aplomb que des vraies

Votre défense : - Le ton n'est pas une preuve - Vérifiez TOUJOURS les faits importants - Méfiez-vous des statistiques et citations

9.4 La dépendance cognitive

Plus vous utilisez l'IA, moins vous exercez certaines capacités.

Capacité menacée	Ce que fait l'IA à la place
Recherche	L'IA cherche pour vous
Synthèse	L'IA résume pour vous
Rédaction	L'IA écrit pour vous
Réflexion critique	L'IA analyse pour vous
Mémorisation	Pourquoi retenir si l'IA sait ?

La différence importante : Calculer et mémoriser sont des capacités "mécaniques". Réfléchir, synthétiser, critiquer sont des capacités "intellectuelles".

Si vous externalisez votre pensée à l'IA, que vous reste-t-il ?

PARTIE B : Les domaines où ne PAS utiliser l'IA

9.5 Les domaines interdits

Certains usages sont légalement problématiques ou dangereux :

Domaine	Pourquoi éviter	Risque
Diagnostic médical	Erreurs potentiellement mortelles	Santé, vie
Conseil juridique	L'IA n'est pas avocat	Droits, liberté
Décisions financières majeures	Hallucinations sur les chiffres	Patrimoine
Thérapie/soutien psychologique	Pas de vraie empathie	Santé mentale
Contenu pour enfants non supervisé	Impossible de garantir l'approprié	Protection mineurs

La règle absolue : Si une erreur peut causer un dommage irréparable, ne laissez pas l'IA décider seule.

9.6 Les situations d'urgence

En cas d'urgence, n'utilisez PAS l'IA comme premier recours :

Urgence	Réflexe correct	Pourquoi pas l'IA
Accident, blessure	15 (SAMU)	Peut retarder
Violence, danger	17 (Police)	Ne peut pas intervenir
Incendie	18 (Pompiers)	Chaque seconde compte
Pensées suicidaires	3114	N'est pas équipée
Malaise cardiaque	15 + défibrillateur	Ne réanime pas

L'IA ne peut pas évaluer la gravité de VOTRE situation. Et elle ne peut pas appeler les secours pour vous.

9.7 L'apprentissage profond (celui qui compte)

L'IA peut vous faire croire que vous avez appris.

Vous demandez : "Explique-moi la physique quantique." L'IA explique. Vous comprenez l'explication. **Mais avez-vous vraiment appris la physique quantique ?**

La différence : - Information : Données que vous recevez - **Connaissance** : Information intégrée par la pratique et la réflexion

L'IA vous donne de l'information. Seul le travail personnel crée la connaissance.

Pour vraiment apprendre : - Lisez des livres (pas des résumés IA) - Pratiquez (l'IA ne peut pas pratiquer à votre place) - Faites des erreurs (les erreurs de l'IA ne sont pas les vôtres) - Enseignez aux autres (le meilleur test de compréhension)

9.8 Les relations humaines

L'IA peut simuler une conversation. Elle ne peut PAS :

- Vraiment vous connaître
- Éprouver de l'affection
- Être là physiquement
- Partager votre vie

Le piège de la relation IA : Certains préfèrent parler à l'IA qu'aux humains. C'est plus facile : pas de jugement, pas de conflit, pas de déception. Mais c'est aussi plus vide : pas de vrai lien, pas de croissance mutuelle.

L'IA comme complément, pas comme remplacement : - Préparer une conversation difficile : **[OK]** - Remplacer la conversation : **[X]**

9.9 La créativité authentique

L'IA génère du contenu impressionnant. Mais il y a une différence entre :

- Un texte généré qui *semble* créatif
- Un texte créé par un humain qui *exprime* quelque chose de vrai

Ce que l'IA ne peut pas apporter : - L'expérience vécue - L'intention artistique - Le risque personnel - La vulnérabilité authentique

PARTIE C : Quand et comment bien utiliser l'IA

9.10 Les bons usages

Usage	Pourquoi c'est bien	Condition
Premier jet	Dépasse la page blanche	Vous retravaillez ensuite
Brainstorming	Génère des options	Vous choisissez
Explication	Patient, disponible	Vous vérifiez
Traduction	Rapide, contextuel	Vous relisez
Code	Gain de temps	Vous testez
routinier		
Résumé	Survol rapide	Vous approfondissez si besoin

9.11 Les questions à se poser avant

Avant chaque usage, 30 secondes de réflexion :

Question	Si "oui", attention
Une erreur pourrait-elle nuire ?	Vérifiez tout
S'agit-il de données sensibles ?	Minimum nécessaire
Suis-je en train de remplacer ma réflexion ?	Prenez du recul
Est-ce que je pourrais faire sans ?	Essayez parfois
Est-ce une urgence ?	Appelez un humain

9.12 Le bon équilibre

L'IA comme outil, pas comme béquille.

[OK] Sain	[X] Problématique
J'utilise l'IA pour aller plus vite	Je ne sais plus faire sans
Je vérifie ce que l'IA me dit	Je fais confiance aveuglément
Je garde le contrôle de mes décisions	Je laisse l'IA décider
Je parle aux humains ET à l'IA	Je préfère l'IA aux humains
J'apprends en utilisant l'IA	Je délègue l'apprentissage

Ce que vous pouvez faire

1. **Identifiez vos usages actuels** — pour lesquels l'IA est-elle vraiment utile ?
2. **Repérez les dépendances** — où avez-vous arrêté de réfléchir ?
3. **Gardez des “zones sans IA”** — apprentissage, relations, créativité
4. **Posez les 5 questions** avant chaque usage sensible
5. **Faites confiance à votre jugement** — l'IA est un outil, pas un oracle

La règle d'or

L'IA doit amplifier vos capacités, pas les remplacer.

Si vous sentez que vous devenez moins capable sans l'IA qu'avec elle, c'est un signal d'alarme.

L'objectif n'est pas de ne jamais utiliser l'IA. L'objectif est de l'utiliser **en conscience**, pour ce qu'elle fait bien, tout en préservant ce qui vous rend humain.

Fin de la Partie II — SE PROTÉGER

Vous êtes maintenant équipé : droits, questions, critères d'évaluation, règles d'usage.

La suite est plus spéculative : nous passons de *se protéger* à *imaginer*. Prédications, futurs possibles, et ce que tout cela signifie pour nos enfants et notre société.

Si vous êtes pressé, la Partie III peut attendre — vous avez l'essentiel pour naviguer l'IA au quotidien.

Chapitre suivant : L'avenir — prédictions honnêtes

PARTIE III : IMAGINER

« Le futur n'est pas écrit — mais certaines pages sont déjà tournées. »

Ce que vous allez explorer : - Ce qui est quasi certain vs ce qui reste ouvert - Comment l'IA va transformer l'éducation, le travail, la société - Les questions que nous devons poser — ensemble

Cette partie est la plus spéculative du livre. Nous entrons dans le territoire des projections. Tout ce qui suit pourrait être invalidé dans 2 ans. C'est aussi le chapitre le plus important : car c'est maintenant que se dessine l'avenir.

Chapitre 10

L'IA dans 5 ans — prédictions honnêtes

Imaginer n'est pas prédire, mais préparer.

Cette partie du livre ne prétend pas connaître l'avenir. Elle distingue ce qui est *probable* (tendances en cours), ce qui est *possible* (scénarios plausibles), et ce qui relève du *fantasmé* (ni probable ni souhaitable). L'objectif : vous donner les clés pour anticiper, pas pour craindre.

Tout le monde fait des prédictions sur l'IA.

La plupart sont fausses.

Voici une tentative honnête de distinguer le probable du possible du fantasmé.

10.1 Ce qui est quasi certain

Ces tendances sont déjà en cours et vont s'accélérer :

1. L'IA sera partout invisible

Vous n'utiliserez plus "ChatGPT". L'IA sera intégrée dans tous vos outils : - Votre traitement de texte - Votre messagerie - Votre navigateur - Vos applications métier - Vos objets connectés

Vous ne "lancerez" plus l'IA. Elle sera là, en arrière-plan, tout le temps.

2. Les modèles seront meilleurs et moins chers

Chaque année apporte : - Des modèles plus performants - Des coûts d'inférence en baisse - Une accessibilité croissante

Ce qui coûte 1 000 € aujourd'hui coûtera 10 € dans 5 ans.

3. L'IA multimodale sera la norme

L'IA comprendra et générera : - Texte - Images - Audio - Vidéo - Code - Données structurées

Tout cela simultanément, de manière fluide.

4. Les agents autonomes se multiplieront

L'IA ne se contentera plus de répondre. Elle agira : - Réserver un billet - Envoyer un email - Modifier un document - Commander un produit - Coordonner plusieurs tâches

Avec supervision humaine au début, puis de moins en moins.

10.2 Qu'est-ce qu'un Agent, vraiment ?

Le mot "agent" est partout. Mais attention au marketing.

La définition rigoureuse :

Un Agent est un acteur autonome, non omniscient, défini par un rôle et un domaine, qui perçoit, raisonne et agit dans les limites de règles explicites, en interaction avec d'autres agents et des systèmes humains.

Ce qui fait un vrai agent :

Propriété	Signification
Rôle	Un agent n'est pas générique. Il a une finalité claire (Planificateur, Tuteur, Analyste...)
Domaine	Il n'agit que dans un territoire limité (temps, santé, travail...)
Règles	Ce qu'il peut faire, ce qu'il ne doit jamais faire, quand s'arrêter
Responsabilité	Il doit pouvoir expliquer pourquoi il a décidé cela
Traçabilité	Ses décisions sont auditable, ses traces interprétables

La différence cruciale :

Agent "marketing"	Agent rigoureux
Prompt + outil	Rôle + domaine
Omnipotent (fait tout)	Borné (fait une chose bien)
Pas de mémoire	Mémoire explicite
Non responsable	Responsable
Non traçable	Traçable
Démo impressionnante	Infrastructure fiable

Pourquoi c'est important :

Un agent sans règles est une API dangereuse. Un agent sans responsabilité est une fonction avec du marketing. Un agent sans traçabilité ne peut pas être audité.

→ Autonomie \neq indépendance. Un agent peut initier des actions, mais ne décide pas de ses propres finalités. Il sert une intention supérieure (la vôtre).

L'avenir : l'Internet des Agents

Comme pour Internet : - Aucun agent n'est central - Aucun agent n'est omniscient - Chaque agent est spécialisé - Les protocoles comptent plus que l'intelligence locale
Ce n'est pas l'agent qui est intelligent. C'est l'écosystème.

10.3 Open Source, Open Data, Open Agents : les leçons du passé

Avant de parler d'agents ouverts, regardons ce qui a marché — et ce qui a échoué.

Open Source : un demi-succès

[OK] Ce qui a marché : - Les infrastructures (Linux, Git, Python, PostgreSQL...) - La standardisation des interfaces - La collaboration à grande échelle

[X] Ce qui manque : - Le comportement n'est pas ouvert - Les décisions sont implicites - Le code ne dit pas *pourquoi* il agit ainsi

→ Open Source = Open Mechanics, pas Open Reasoning.

Open Data : un échec plus net

Les données ont été ouvertes, mais sans : - Contexte d'usage - Traçabilité décisionnelle - Lien entre donnée → action → impact

Résultat : - Peu de réutilisation réelle par les citoyens - Captation par les acteurs capables de "l'absorber" - Le citoyen reste passif

→ Open Data = Open Inputs, sans Open Processing.

"OpenAI" : pourquoi le nom est devenu un mensonge structurel

Ce n'est pas une trahison morale. C'est une contradiction économique et architecturale.

Ce qui n'est PAS ouvert	Ce qui est ouvert
Données d'entraînement	Une interface conversationnelle
Raisonnements internes	Une illusion de transparence
Politiques de décision	
Mémoire long terme	
Critères d'arbitrage	

→ OpenAI a ouvert l'accès, pas l'intelligence. C'est une IA sans responsabilité explicite.

10.4 Ce que "Open Agents" devrait vraiment signifier

Attention aux fausses pistes : - Open Agents ≠ Open prompts - Open Agents ≠ Open workflows - Open Agents ≠ Marketplace d'agents

Un Agent n'est pas un script. Un Agent observe, décide, agit, se souvient — et peut expliquer pourquoi.

Ouvrir un agent, ce n'est pas ouvrir son code. C'est ouvrir son contrat cognitif.

1. Open Agent Contract (fondamental)

Un agent ouvert doit exposer : - Ses **capacités** (ce qu'il sait faire) - Ses **limites** (ce qu'il refuse / ne sait pas) - Ses **objectifs** - Ses **règles** - Ses **sources de connaissance** - Son **modèle de mémoire**

→ Lisible *avant* exécution.

2. Open Reasoning Trace (non négociable)

Pas le "chain-of-thought" brut, mais : - Décisions prises - Alternatives rejetées - Règles appliquées - Conflits arbitrés

→ Justification inspectable, pas introspection magique.

3. Open Memory Semantics

Pas les embeddings eux-mêmes, mais : - Ce qui est mémorisé - Pourquoi - Quand - Pour combien de temps - À quel scope (user / agent / système)

→ La mémoire devient un objet gouvernable.

4. Open Action Policy

Un agent ouvert doit dire : - Quelles actions il peut déclencher - À quel niveau d'autonomie - Avec quel seuil de confirmation humaine

→ Pas d'action sans responsabilité.

Pourquoi Open Agents peut réussir là où Open Data a échoué

Parce que : - L'unité n'est plus la donnée, mais la **décision** - L'utilisateur n'est plus lecteur, mais **co-agent** - La valeur n'est plus l'accès, mais la **confiance**

Open Agents = Open Accountability.

La phrase qui résume tout :

Open Source opened code. Open Data opened inputs. Open Agents must open decisions.

10.5 Du fichier au concept : quand l'informatique devient invisible

Une autre transformation profonde est en cours — moins visible, mais tout aussi fondamentale.

Pendant longtemps, l'informatique a été organisée autour de supports.

On parlait de fichiers .doc, .xls, .ppt. On rangeait des dossiers. On sauvegardait sur des disques durs. On transportait des CD, puis des clés USB.

Autrement dit : on gérait des **objets**.

Puis quelque chose de fondamental a changé.

Quand le support disparaît, le concept reste

Aujourd'hui, plus personne ne dit vraiment :

« Passe-moi le fichier .ppt »

On dit :

« Envoie-moi la présentation »

On ne dit plus :

« J'ai perdu mon .xls »

mais :

« J'ai perdu mon tableau »

Le format s'est effacé. Le concept a pris le dessus.

Comme pour la musique : - le vinyle a disparu - le CD a disparu - le fichier local a disparu

Mais la musique, elle, est partout.

Quand le support disparaît, ce qui compte vraiment apparaît.

Vers une informatique de concepts

L'avenir de l'informatique n'est plus de gérer des fichiers. C'est de gérer des **concepts** :

Fichier (ancien monde)	Concept (nouveau monde)
.ppt	Une présentation — une manière de transmettre une idée
.xls	Un tableau — une façon d'organiser le réel
.ics	Un calendrier — une relation au temps
.doc	Une note — une pensée capturée

Les formats deviennent des détails techniques, interchangeables, secondaires.

Ce que ça change pour vous

- Vous n'aurez plus à choisir un format, mais à exprimer une intention
- Vos données seront organisées par sens, pas par extension de fichier
- L'IA comprendra ce que vous voulez faire, pas quel logiciel ouvrir

Le risque : si c'est l'IA qui organise vos concepts, c'est elle qui structure votre pensée.

L'opportunité : une informatique qui s'efface du champ visuel, où l'humain redevient central.

10.6 Vers une IA gouvernable : ce que nous pourrions exiger

Si nous voulons que l'IA reste un outil au service des citoyens, voici ce que nous devrions collectivement exiger.

Le modèle actuel est problématique :

[Vous] → [API cloud] → [Modèle opaque] → [Réponse]

↓
[Vos données stockées]
[Entraînement futur ?]
[Analyse comportementale ?]

Un modèle alternatif est possible :

[Vous] → [Votre profil personnel] → [Gouverneur local] → [Modèle]

↓ ↓

[Données sous VOTRE contrôle] [Règles vérifiables]

Les trois couches d'une IA gouvernable :

Couche	Rôle	Principe
Raisonnement	Traiter, comprendre, générer	Le LLM fait ce qu'il fait bien, mais ne décide pas seul
Confiance	Identité, permissions, traçabilité	Avant d'agir, l'IA vérifie qu'elle en a le droit
Gouvernance	Règles, limites, valeurs	Les règles sont explicites, pas cachées dans le modèle

Ce qu'un “profil personnel” devrait permettre :

- Définir vos préférences (style de communication, langue, niveau de détail)
- Fixer vos limites (ce que l'IA peut/ne peut pas faire)
- Contrôler les permissions (service par service, action par action)
- Garder un historique auditable (ce que vous avez partagé, avec qui)
- Être portable (fonctionner avec n'importe quelle IA, pas de verrouillage)

Pourquoi c'est important :

Les garde-fous des IA actuelles sont : - Entraînés dans le modèle → impossibles à vérifier - Définis par l'entreprise → vous n'avez pas votre mot à dire - Contournables → pas de garantie réelle

Une IA gouvernable aurait des règles **externes, vérifiables, personnalisables.**

Ce n'est pas de la science-fiction. C'est un choix d'architecture. Et ce choix dépend de ce que nous exigerons collectivement.

[Livre] Pause possible :

Vous avez l'essentiel : ce qui est certain, ce que sont les agents, les enjeux d'ouverture et de gouvernabilité. Vous pouvez passer au chapitre suivant

si vous le souhaitez. La suite approfondit les scénarios — probable, incertain, fantasmé.

Revenir plus tard est aussi une forme d'intelligence.

10.7 Ce qui est probable

Ces évolutions ont de bonnes chances de se produire :

1. La consolidation du marché

Aujourd'hui : des dizaines de startups IA. Dans 5 ans : 3-5 acteurs dominants.

Probablement : - OpenAI/Microsoft - Google - Un acteur chinois (Baidu, ByteDance) - Peut-être Meta - Peut-être un européen ? Si l'Europe — institutions et citoyens — le veut.

L'Europe doit se réveiller.

La bataille commence maintenant. L'Europe a su créer le W3C, standardiser le Web, penser l'Internet des Objets. Elle doit maintenant prendre ses responsabilités sur **l'Internet des Agents** : définir les standards, éviter les dérives, clarifier les définitions floues que le marketing obscurcit.

Ce n'est pas une question technique. C'est une question de souveraineté cognitive.

L'Europe doit : - Réfléchir à un niveau plus haut que la régulation défensive - Accélérer pour proposer des solutions, pas seulement des contraintes - Standardiser ce qu'est un agent, ce qu'il doit exposer, ce qu'il doit garantir - Créer l'infrastructure de confiance que les géants ne créeront pas seuls

Si l'Europe rate ce virage, elle deviendra consommatrice d'agents américains et chinois — sans aucun contrôle sur les décisions qui façonneront la vie de ses citoyens.

2. La personnalisation profonde

L'IA vous connaîtra : - Vos préférences - Votre historique - Vos habitudes - Votre style

Elle adaptera tout en conséquence. C'est pratique. Ça peut aussi être invasif.

Ce que ça signifie concrètement :

- Votre messagerie triera vos emails avant que vous les voyiez — et décidera ce qui est "important"
- Votre assistant proposera des réponses à votre place — avec votre ton
- Votre agenda se réorganisera automatiquement — selon ce qu'il pense être vos priorités
- Vos recherches seront filtrées — selon ce que l'IA croit que vous voulez trouver
- Vos achats seront suggérés — avant même que vous y pensiez

Le risque : vous ne verrez plus que ce que l'IA pense que vous devez voir. Vos choix seront guidés par un système qui optimise pour l'engagement, pas pour votre bien-être. La personnalisation devient une bulle. La bulle devient une prison confortable.

La question clé : qui définit les critères de personnalisation ? Vous — ou l'entreprise qui vend votre attention ?

Le principe à défendre : l'Agent propose, l'utilisateur dispose.

L'IA peut suggérer, filtrer, prioriser — mais la décision finale doit rester humaine. Pas de tri invisible. Pas d'action automatique sans consentement. L'agent est un conseiller, pas un tuteur. Cette frontière est la ligne rouge entre assistance et manipulation.

3. L'IA dans l'éducation

Tuteurs personnalisés pour chaque élève. Exercices adaptés au niveau. Feedback instantané.

Questions ouvertes — et pistes de réflexion :

- **Quel rôle pour l'enseignant humain ?** L'enseignant ne disparaît pas — il se transforme. Moins transmetteur de savoir (l'IA peut le faire), plus accompagnateur de sens : motivation, esprit critique, relation humaine. Le professeur devient celui qui aide à poser les bonnes questions, pas celui qui donne toutes les réponses.
- **Comment évaluer les élèves ?** Si l'IA peut rédiger une dissertation, évaluer une dissertation n'a plus de sens. Il ne faut pas évaluer si un élève *sait* la réponse, mais sa capacité à *apprendre* : démontrer un raisonnement, corriger ses erreurs, poser les bonnes questions. L'enseignement n'est pas là pour transférer la Connaissance — il est là pour apprendre à Réfléchir. Ne pas donner le poisson, mais apprendre à se servir d'une canne à pêche.
- **Comment gérer les inégalités d'accès ?** L'IA éducative pourrait réduire les inégalités (un tuteur pour chaque enfant) ou les amplifier (ceux qui savent l'utiliser progressent plus vite). Tout dépend de qui y a accès, et de comment on forme à l'utiliser. L'école publique a un rôle crucial : garantir que l'IA éducative soit un bien commun, pas un privilège.
- **Faut-il interdire l'IA aux élèves ?** Non. Il ne faut pas blâmer les élèves qui utilisent ChatGPT ou autre — il faut leur apprendre à s'en servir. Comme la calculatrice. Comme Internet. Interdire, c'est préparer des citoyens incompetents face à un outil qu'ils utiliseront toute leur vie. Enseigner, c'est leur donner le pouvoir de l'utiliser intelligemment — et de savoir quand s'en passer. Mais cela implique que les enseignants sachent eux aussi s'en servir, évaluer son usage, et apprendre à enseigner à l'ère de l'IA. Former les professeurs est la première urgence. Et le système éducatif doit intégrer l'évolution de l'IA : demain ce ne seront plus des chatbots, mais des agents autonomes. Comme le passage du taxi à Uber : avant, vous alliez chercher le taxi — maintenant, le taxi vient vous chercher. Avec les agents, vous n'irez plus chercher l'information — l'information viendra à vous. On n'enseigne pas une technologie figée — on enseigne à s'adapter.

4. L'automatisation de nombreux emplois

Pas la disparition totale du travail humain. Mais la transformation profonde de nombreux métiers :

Métier	Avant	Après
Comptabilité	Saisir, calculer, vérifier	Contrôler ce que l'IA a fait, valider les exceptions
Juridique	Rédiger des contrats types	Orchestrer des agents qui rédigent, vérifier la cohérence

Métier	Avant	Après
Service client	Répondre aux questions	Superviser les agents, gérer les escalades humaines
Rédaction	Écrire	Briefer, relire, valider le ton et la stratégie
Programmation	Coder	Spécifier, revoir le code généré, architecturer
Analyse de données	Extraire, nettoyer, calculer	Poser les bonnes questions, interpréter, décider

Le travail ne disparaît pas — il change de nature : - **Moins d'exécution** → plus de contrôle et vérification - **Moins de production** → plus d'orchestration et management - **Moins de temps sur les tâches** → plus de responsabilité sur les résultats

La vitesse d'exécution explose. Mais la vitesse de décision reste humaine. C'est là que se joue la valeur.

10.8 Ce qui est incertain

Ces questions restent ouvertes :

1. L'IA atteindra-t-elle le niveau humain général ?

L'AGI (Artificial General Intelligence) — une IA aussi polyvalente qu'un humain — est : - Annoncée comme imminente par certains - Jugée très lointaine par d'autres - Peut-être un concept mal défini

Mon avis : pas dans 5 ans. Peut-être dans 20. Peut-être jamais de la façon dont on l'imagine.

2. Y aura-t-il un accident majeur ?

Un système IA qui : - Cause des morts (voitures autonomes, médical) - Provoque un crash financier - Est utilisé dans une cyberattaque massive - Manipule une élection de façon décisive

C'est probable qu'il y aura des incidents. L'ampleur est incertaine.

3. Comment évoluera la régulation ?

Scénarios possibles : - L'Europe régule, perd en innovation, gagne en protection - Les USA laissent faire, dominant, avec des dérives - La Chine crée un modèle autoritaire efficace - Un consensus international émerge (peu probable)

4. Quelle sera la réaction sociale ?

- Acceptation enthousiaste ?
- Rejet technophobe ?
- Usage différencié selon les générations ?
- Nouvelles formes de résistance ?

10.9 Ce qui est fantasmé

Ces scénarios font de bons films mais sont peu probables à 5 ans :

[X] L'IA consciente

Les LLM n'ont pas de conscience. Ils simulent des conversations. Rien n'indique qu'ils "vivent" quoi que ce soit.

L'émergence d'une conscience artificielle serait une révolution scientifique majeure. Pas pour demain.

[X] L'IA qui prend le contrôle

Le scénario Terminator suppose : - Une IA avec des objectifs propres - Une IA qui veut survivre - Une IA qui peut agir dans le monde physique

Les LLM actuels n'ont rien de tout cela. Ils répondent à des prompts.

[X] La fin du travail humain

Le travail va se transformer, pas disparaître. Il y aura des emplois détruits et des emplois créés. L'équilibre est incertain mais le chômage de masse n'est pas inévitable.

[X] L'IA qui résout tous nos problèmes

L'IA n'est pas une baguette magique. Elle ne résoudra pas : - Le changement climatique (sans volonté politique) - Les inégalités (sans redistribution) - Les conflits humains (sans dialogue)

L'IA est un outil. Un outil puissant. Mais juste un outil.

10.10 Les wildcards

Des événements imprévisibles pourraient tout changer :

1. Une percée technologique majeure

Si quelqu'un découvre une architecture fondamentalement nouvelle, toutes les prédictions tombent.

2. Un accident catastrophique

Un événement traumatisant pourrait provoquer une régulation massive ou un rejet social.

3. Une guerre ou crise géopolitique

La technologie IA pourrait être un enjeu stratégique dans un conflit majeur.

4. Une mobilisation citoyenne

Si les gens décident collectivement de refuser certains usages, le marché devra s'adapter.

10.11 Ce que ça signifie pour vous

Le monde dans 5 ans sera différent. Voici comment vous préparer :

Certitude	Ce que vous pouvez faire
L'IA sera partout	Apprenez à l'utiliser consciemment
Les emplois changeront	Développez des compétences complémentaires à l'IA
La vie privée sera menacée	Protégez-vous dès maintenant
La désinformation augmentera	Renforcez votre esprit critique
Des choix de société seront faits	Participez au débat public

Les compétences qui resteront précieuses :

- Jugement critique
- Créativité authentique
- Relations humaines
- Leadership et vision
- Adaptabilité
- Éthique et valeurs

10.12 Scénarios pour 2030

Scénario optimiste : L'IA démocratisée

- L'IA améliore la productivité et la qualité de vie
- Les gains sont partagés équitablement
- La régulation protège les droits
- Les emplois se transforment mais ne disparaissent pas
- L'Europe trouve sa place

Scénario pessimiste : L'IA concentrée

- Quelques entreprises contrôlent tout
- Les inégalités explosent
- La surveillance devient omniprésente
- Le travail précaire se généralise
- L'Europe devient dépendante

Scénario réaliste : L'IA inégale

- Des bénéfices réels mais mal répartis
- Une régulation en retard sur l'innovation
- Des tensions sociales autour de l'emploi
- Des risques de sécurité mal maîtrisés
- Un débat permanent sur les usages acceptables

10.13 Les questions que nous devons trancher

D'ici 5 ans, nous devons décider collectivement :

1. **Transparence** : L'IA doit-elle toujours s'identifier comme IA ?
2. **Responsabilité** : Qui paie quand l'IA se trompe ?
3. **Emploi** : Comment accompagner les travailleurs impactés ?
4. **Éducation** : Que doit-on encore apprendre aux humains ?
5. **Créativité** : Quelle place pour l'art généré par IA ?
6. **Souveraineté** : Comment ne pas dépendre des géants américains et chinois ?
7. **Vie privée** : Où placer les limites de la collecte de données ?
8. **Démocratie** : Comment empêcher la manipulation à grande échelle ?

Ces questions n'ont pas de réponses techniques. Ce sont des choix de société.

Ce que vous pouvez faire

1. **Restez informé sans être obsédé.** L'IA évolue vite mais pas au jour le jour.
 2. **Développez des compétences durables.** Ce que l'IA ne peut pas (encore) faire.
 3. **Participez au débat.** Ces décisions nous concernent tous.
 4. **Préparez-vous à l'incertitude.** Personne ne sait exactement ce qui va arriver.
 5. **Gardez votre humanité.** C'est ce qui aura le plus de valeur dans un monde d'IA.
-

Chapitre suivant : Les emplois transformés

Chapitre 11

Les emplois transformés

L'IA ne va pas “voler” les emplois.

Elle va les transformer.

Certains disparaîtront. D'autres apparaîtront. La plupart changeront.

Voici ce que nous savons — et ce que nous ne savons pas.

11.1 L'histoire se répète (mais différemment)

À chaque révolution technologique, les mêmes peurs :

- 1800s : Les métiers à tisser vont détruire l'emploi
- 1900s : L'électricité va rendre les travailleurs inutiles
- 1950s : L'automatisation va créer le chômage de masse
- 1990s : Les ordinateurs vont éliminer le travail de bureau
- 2020s : L'IA va tout faire mieux que nous

Ce qui s'est passé à chaque fois :

Des emplois ont disparu. D'autres sont apparus. Au total, plus d'emplois qu'avant.

Mais cette fois, c'est peut-être différent :

Les révolutions précédentes touchaient le travail physique et les tâches répétitives. L'IA touche le travail cognitif. La vitesse de changement est plus rapide. L'échelle est mondiale.

11.2 Les métiers les plus exposés

Forte exposition à l'automatisation par IA :

Métier	Tâches automatisables	Perspective
Traducteur	Traduction standard	Reste pour le culturel, littéraire
Comptable	Saisie, rapprochement	Reste pour le conseil

Métier	Tâches automatisables	Perspective
Rédacteur technique	Documentation standard	Reste pour le complexe
Service client niveau 1	Réponses courantes	Reste pour l'escalade
Analyste de données junior	Requêtes simples	Reste pour l'interprétation
Programmeur junior	Code boilerplate	Reste pour l'architecture
Téléconseiller	Scripts standards	Reste pour le relationnel

Ce que ça signifie :

Ces métiers ne disparaissent pas. Ils se transforment. La partie répétitive est automatisée. La partie à valeur ajoutée reste humaine.

11.3 Les métiers les plus protégés

Faible exposition à l'automatisation par IA :

Métier	Pourquoi protégé
Soignant	Contact humain irremplaçable
Artisan	Savoir-faire physique
Enseignant	Relation éducative
Manager	Leadership, motivation
Thérapeute	Empathie profonde
Créatif original	Vision artistique unique
Entrepreneur	Prise de risque, vision
Chercheur de pointe	Innovation fondamentale

L'IA peut aider ces métiers.

Elle ne peut pas les remplacer. Le contact humain, le jugement, la créativité authentique restent hors de portée.

11.4 Les nouveaux métiers

Des emplois qui n'existaient pas il y a 5 ans :

Métier émergent	Description
Prompt engineer	Optimiser les interactions avec les LLM
AI trainer	Entraîner et affiner les modèles

Métier émergent	Description
AI ethics officer	Garantir l'usage responsable de l'IA
AI auditor	Vérifier les biais et la conformité
Human-AI coordinator	Gérer la collaboration homme-machine
AI content moderator	Superviser les contenus générés
Synthetic media specialist	Créer des médias IA de qualité

Et des métiers qu'on n'imagine pas encore.

En 2005, personne n'imaginait "community manager" ou "data scientist". Les métiers de 2030 n'existent pas encore.

11.5 La transformation des métiers existants

L'IA ne remplace pas le métier. Elle change la façon de l'exercer.

Médecin :

- Avant IA : Diagnostic basé sur l'expérience et quelques examens
- Avec IA : Diagnostic assisté par analyse de données massives
- Valeur humaine : Relation patient, décision éthique, empathie

Avocat :

- Avant IA : Recherche juridique manuelle, rédaction chronophage
- Avec IA : Recherche automatisée, rédaction assistée
- Valeur humaine : Stratégie, plaidoirie, conseil personnalisé

Enseignant :

- Avant IA : Cours standardisés, correction manuelle
- Avec IA : Contenu personnalisé, correction automatique
- Valeur humaine : Inspiration, accompagnement, exemple

Journaliste :

- Avant IA : Recherche, rédaction, vérification manuelles
- Avec IA : Veille automatisée, rédaction assistée, fact-checking accéléré
- Valeur humaine : Enquête, analyse, point de vue

11.6 Les compétences de demain

Ce que l'IA fait mieux que vous :

- Traiter de grandes quantités d'informations
- Exécuter des tâches répétitives
- Produire du contenu standard rapidement
- Analyser des patterns dans les données
- Être disponible 24/7

Ce que vous faites mieux que l'IA :

Compétence	Pourquoi précieuse
Pensée critique	Évaluer, questionner, remettre en cause
Créativité originale	Créer du nouveau, pas recombinaison
Intelligence émotionnelle	Comprendre et gérer les émotions
Adaptabilité	Faire face à l'imprévu
Communication complexe	Persuader, négocier, inspirer
Collaboration	Travailler en équipe diverse
Éthique	Distinguer le bien du mal
Apprentissage continu	S'adapter aux nouvelles situations

Investissez dans ces compétences.

11.7 Comment se préparer

Si vous êtes en activité :

1. **Évaluez votre exposition** : Quelle partie de votre travail est automatisable ?
2. **Identifiez votre valeur ajoutée** : Que faites-vous que l'IA ne peut pas faire ?
3. **Développez des compétences complémentaires** : Apprenez à utiliser l'IA comme outil
4. **Restez curieux** : Suivez les évolutions de votre secteur
5. **Construisez votre réseau** : Les relations humaines restent précieuses

Si vous êtes en formation :

1. **Ne vous spécialisez pas trop tôt** : Les spécialités d'aujourd'hui seront peut-être obsolètes
2. **Apprenez à apprendre** : La capacité d'adaptation vaut plus que les connaissances
3. **Développez des compétences transversales** : Communication, leadership, créativité
4. **Maîtrisez l'IA comme outil** : Pas besoin d'être expert, mais être utilisateur avancé
5. **Gardez une expertise profonde quelque part** : Les généralistes + IA peuvent être remplacés ; les experts + IA sont précieux

11.8 La question de la reconversion

Pour ceux dont le métier est fortement impacté :

Situation	Stratégies possibles
Métier en déclin rapide	Reconversion vers un métier connexe
Métier en transformation	Montée en compétences sur la partie haute
Métier stable mais menacé à terme	Diversification, formation continue

Situation	Stratégies possibles
Proche de la retraite	Adaptation minimale, transmission

Les dispositifs existants en France :

- **CPF** (Compte Personnel de Formation) : Utilisez vos droits
- **Conseil en évolution professionnelle** : Gratuit, disponible pour tous
- **VAE** (Validation des Acquis de l'Expérience) : Faites reconnaître votre expérience
- **Transitions collectives** : Pour les reconversions sectorielles
- **France Travail** : Accompagnement si perte d'emploi

11.9 Le débat sur le partage des gains

L'IA augmente la productivité. Qui en bénéficie ?

Options de redistribution débattues :

Option	Description	Position politique
Revenu universel	Revenu de base pour tous	Gauche/transpartisan
Réduction du temps de travail	Travailler moins, partager le travail	Gauche
Formation massive	Préparer aux métiers de demain	Centre
Taxation des robots	Financer la transition	Diverses
Laisser-faire	Le marché s'adaptera	Droite libérale

Ce n'est pas une question technique.

C'est une question politique. Comment voulons-nous partager les fruits de l'automatisation ?

11.10 Ni utopie ni dystopie

Le scénario utopique :

L'IA nous libère du travail pénible. Nous avons plus de temps pour créer, aimer, vivre. Les gains de productivité sont partagés. Chacun trouve sa place.

Le scénario dystopique :

L'IA concentre la richesse chez quelques-uns. Le travail devient précarisé. Beaucoup sont laissés pour compte. Les inégalités explosent.

Le scénario réaliste :

Un peu des deux. Des gagnants et des perdants. Une transition difficile mais gérable. Des choix collectifs à faire.

Ce qui déterminera l'issue :

Pas la technologie. Nos choix politiques. Notre capacité à nous organiser. Notre volonté de partager.

Ce que vous pouvez faire

1. **Évaluez honnêtement votre situation.** Où êtes-vous sur le spectre de l'automatisation ?
 2. **Investissez dans vos compétences uniques.** Ce que l'IA ne peut pas (encore) faire.
 3. **Apprenez à utiliser l'IA.** Mieux vaut être celui qui utilise l'IA que celui qui est remplacé par elle.
 4. **Participez au débat social.** Les choix de redistribution nous concernent tous.
 5. **Restez humain.** C'est votre avantage compétitif ultime.
-

Chapitre suivant : L'éducation repensée

Chapitre 12

L'éducation repensée

Que doit-on encore apprendre aux enfants quand l'IA sait (presque) tout ?

Cette question bouleverse l'éducation.

Elle n'a pas de réponse simple.

12.1 Le paradoxe fondamental

L'IA peut :

- Répondre à presque toutes les questions de connaissance
- Expliquer n'importe quel concept
- Résoudre des problèmes mathématiques
- Rédiger des dissertations
- Corriger des exercices
- S'adapter au niveau de chaque élève

Alors pourquoi aller à l'école ?

C'est la question que posent (consciemment ou non) tous les élèves qui ont accès à ChatGPT.

La réponse ne peut pas être : "Parce qu'on l'a toujours fait."

Il faut repenser ce que l'école apporte que l'IA ne peut pas apporter.

12.2 Ce que l'école apportait (avant l'IA)

Le modèle traditionnel :

1. **Transmission de connaissances** : Le professeur sait, l'élève apprend
2. **Évaluation standardisée** : Examens, diplômes, classements
3. **Socialisation** : Apprendre à vivre ensemble
4. **Préparation à l'emploi** : Compétences pour le marché du travail

Ce qui change avec l'IA :

- La connaissance est accessible instantanément

- L'évaluation traditionnelle est facilement contournable
- La socialisation reste (heureusement) humaine
- La préparation à l'emploi doit être repensée

12.3 Ce que l'école doit devenir

Les nouvelles priorités :

Ancienne priorité	Nouvelle priorité
Mémoriser des faits	Savoir où et comment chercher
Reproduire des méthodes	Comprendre les principes sous-jacents
Travailler seul	Collaborer avec des humains et des machines
Éviter les erreurs	Apprendre de ses erreurs
Suivre des instructions	Poser les bonnes questions
Acquérir un métier	Développer une adaptabilité

L'école de demain devrait apprendre à :

1. **Penser** : Esprit critique, logique, questionnement
2. **Créer** : Expression personnelle, originalité
3. **Collaborer** : Travail d'équipe, communication
4. **S'adapter** : Apprendre à apprendre
5. **Agir** : Initiative, responsabilité, éthique

12.4 Le problème de la triche

ChatGPT a rendu la triche triviale.

Un élève peut : - Faire rédiger ses devoirs par l'IA - Obtenir des réponses aux examens - Produire des travaux impressionnants sans effort personnel

Les réponses possibles :

Approche	Description	Limite
Interdiction	Bloquer l'accès à l'IA	Impossible à appliquer
Détection	Outils anti-IA	Peu fiables, course aux armements
Surveillance	Examens surveillés sans accès	Stress, inadapté au monde réel
Intégration	L'IA fait partie de l'évaluation	Comment mesurer l'apport personnel ?
Transformation	Changer ce qu'on évalue	Nécessite une refonte profonde

La seule solution durable :

Repenser ce que signifie "apprendre" et "évaluer".

12.5 Les modèles émergents

Évaluation par portfolio :

Au lieu d'examens ponctuels, l'élève constitue un portfolio de réalisations sur l'année. L'IA peut aider, mais le processus et la progression sont visibles.

Évaluation orale :

L'élève doit expliquer, défendre, répondre aux questions. L'IA peut avoir aidé à préparer, mais la compréhension se voit en direct.

Projet de groupe :

Des réalisations collectives où la collaboration est aussi importante que le résultat. L'IA est un outil parmi d'autres.

Apprentissage par problème :

On part de problèmes réels, pas de solutions à mémoriser. L'IA est une ressource à utiliser intelligemment.

12.6 Le rôle de l'enseignant

Ce que l'IA peut faire :

- Expliquer un concept (parfois mieux qu'un humain)
- Adapter le niveau de difficulté
- Être disponible 24/7
- Corriger instantanément
- Ne jamais se fatiguer

Ce que l'IA ne peut PAS faire :

Rôle humain	Pourquoi irremplaçable
Inspirer	L'enthousiasme est contagieux
Modéliser	Montrer l'exemple vivant
Accompagner	Comprendre le contexte personnel
Challenger	Pousser au-delà des zones de confort
Écouter	Vraiment entendre les difficultés
Célébrer	Reconnaître les progrès

Le professeur de demain :

Moins transmetteur de connaissances. Plus mentor, guide, coach, inspirateur.

12.7 L'éducation à l'IA

Ce que chaque élève devrait apprendre sur l'IA :

Niveau de base (tous) :

- Comment fonctionne l'IA (grandes lignes)
- Ce que l'IA peut et ne peut pas faire
- Les risques (désinformation, manipulation, vie privée)
- Comment utiliser l'IA de façon responsable
- Ses droits face aux systèmes automatisés

Niveau intermédiaire (intéressés) :

- Les bases de la programmation
- Comment les données entraînent les modèles
- Les biais et leurs sources
- L'histoire et l'évolution de l'IA

Niveau avancé (spécialisation) :

- Mathématiques de l'IA
- Programmation de systèmes IA
- Éthique et gouvernance de l'IA

12.8 Les inégalités numériques

L'IA risque d'amplifier les inégalités.

Élève favorisé	Élève défavorisé
Accès aux meilleurs outils IA	Outils limités ou pas d'accès
Parents qui comprennent	Parents dépassés
Encadrement pour bien utiliser IA comme amplificateur	Laissé seul face à l'outil IA comme béquille

Ce que l'école publique doit garantir :

1. **Accès égal** aux outils IA pour tous les élèves
2. **Formation** des enseignants à l'IA
3. **Programmes** adaptés à la réalité de l'IA
4. **Accompagnement** pour les familles éloignées du numérique

12.9 L'apprentissage tout au long de la vie

Le diplôme initial ne suffit plus.

Dans un monde où les métiers changent rapidement, l'apprentissage continu devient nécessaire.

Ce que ça implique :

- Formation continue accessible à tous
- Validation des compétences acquises en cours de vie
- Possibilité de se reconvertir à tout âge
- Entreprises qui investissent dans la formation

Le rôle de l'IA :

L'IA peut être un formidable outil d'apprentissage tout au long de la vie : - Disponible quand on en a besoin - Adaptée à notre niveau - Patient et personnalisée

Encore faut-il savoir l'utiliser intelligemment.

12.10 Ce que les parents peuvent faire

À la maison :

Action	Comment
Expliquer l'IA Modéliser l'usage	Discuter de ce que c'est, ce que ça fait Montrer comment vous utilisez (ou n'utilisez pas) l'IA
Superviser Questionner Valoriser l'effort	Accompagner les premiers usages "Comment as-tu fait ?" "As-tu compris ?" Pas juste le résultat (que l'IA peut produire)

Avec l'école :

- Demander quelle est la politique de l'établissement sur l'IA
 - Participer aux discussions sur le sujet
 - Signaler les difficultés ou dérives observées
-

12.11 Les questions ouvertes

Nous n'avons pas encore les réponses à :

1. Comment évaluer équitablement à l'ère de l'IA ?
2. Quel équilibre entre usage de l'IA et apprentissage "à la main" ?
3. Comment former les enseignants assez vite ?
4. Comment garantir l'équité d'accès ?
5. Que doit garantir le diplôme si l'IA peut tout faire ?

Ces questions méritent un débat de société.

Pas juste des décisions technocratiques.

Ce que vous pouvez faire

1. **Parlez d'IA avec vos enfants.** Expliquez, questionnez, accompagnez.
2. **Intéressez-vous à ce que l'école propose.** Demandez quelle est la politique IA.
3. **Valorisez l'apprentissage, pas juste les résultats.** Le processus compte autant que le produit.

4. **Montrez l'exemple.** Votre rapport à l'IA influence celui de vos enfants.
5. **Participez au débat.** L'avenir de l'éducation nous concerne tous.

Chapitre suivant : L'avenir de la créativité humaine

Chapitre 13

L'avenir de la créativité humaine

L'IA génère des images époustouflantes. L'IA écrit des textes fluides. L'IA compose de la musique.

Est-ce la fin de la créativité humaine ?

Non.

C'est sa redéfinition.

13.1 Ce que l'IA peut créer

Les capacités actuelles sont impressionnantes :

Domaine	Ce que l'IA peut faire
Image	Générer des visuels photoréalistes sur n'importe quel sujet
Texte	Rédiger articles, poèmes, scénarios, code
Musique	Composer dans tous les styles
Vidéo	Générer des clips, des animations
Design	Créer logos, interfaces, maquettes

Le niveau de qualité :

Pour beaucoup d'usages commerciaux, la qualité IA est "suffisante". Une image de stock IA peut remplacer une photo achetée. Un texte IA peut remplacer un article de contenu.

13.2 Ce que l'IA ne peut PAS créer

L'IA ne peut pas :

Capacité manquante	Explication
Avoir une intention	Elle ne "veut" rien dire

Capacité manquante	Explication
Avoir un vécu	Elle n'a pas d'expérience
Prendre un risque	Elle n'a rien à perdre
Être vulnérable	Elle ne peut pas s'exposer
Avoir une vision	Elle n'a pas de projet de vie
Évoluer authentiquement	Elle n'a pas de trajectoire personnelle

Ce qui fait l'art :

L'art n'est pas juste un objet esthétique. L'art est une communication entre un créateur et un public. Il implique une intention, une vulnérabilité, un risque.

L'IA produit des objets qui ressemblent à de l'art. Mais il y manque l'humain derrière.

13.3 La question de l'authenticité

Deux œuvres identiques peuvent avoir des valeurs différentes.

Imaginez : - Un tableau peint par un artiste qui a traversé une dépression - Le même tableau généré par une IA à partir d'un prompt

Visuellement identiques. Mais pas la même chose.

Pourquoi ?

L'un porte une expérience humaine. L'autre est un assemblage de patterns statistiques.

Cela compte-t-il ?

Pour certains, non. Le résultat est ce qui compte. Pour d'autres, l'histoire derrière l'œuvre fait partie de l'œuvre.

13.4 Les créateurs face à l'IA

Les peurs légitimes :

- Perte de revenus (stock photos, contenus basiques)
- Dévaluation du travail ("pourquoi payer si l'IA peut le faire ?")
- Vol du style (modèles entraînés sur leurs œuvres sans consentement)
- Saturation du marché (inondation de contenu généré)

Les opportunités :

- Nouveaux outils de création
- Automatisation des tâches fastidieuses
- Exploration de nouveaux territoires artistiques
- Collaboration homme-machine innovante

La réalité :

Comme toute technologie, l'IA redistribue les cartes. Certains créateurs souffriront. D'autres s'épanouiront. La plupart devront s'adapter.

13.5 L'artiste augmenté

L'IA comme outil du créateur, pas comme créateur :

Usage de l'IA	Exemple
Exploration	Générer des variations pour trouver des pistes
Ébauche	Produire des premières versions à affiner
Référence	Créer des images de référence pour un projet
Automatisation	Gérer les aspects techniques répétitifs
Collaboration	Dialogue créatif avec la machine

Ce qui reste humain :

- La vision initiale
 - La sélection et le jugement
 - L'intention artistique
 - La touche finale
 - La responsabilité de l'œuvre
-

13.6 Les nouvelles formes de création

L'IA ouvre de nouveaux territoires :

Art génératif :

L'œuvre est le système, pas l'output. L'artiste crée les règles, l'IA génère les variations.

Art interactif :

Des œuvres qui évoluent en fonction du public. L'IA comme médium vivant.

Collaboration homme-machine :

Un dialogue créatif entre l'humain et l'IA. Ni l'un ni l'autre ne pourrait créer seul ce qui émerge.

Meta-crétation :

Créer des systèmes qui créent. L'artiste comme architecte de possibilités.

13.7 La question du copyright

Qui possède une œuvre générée par IA ?

Scenario	Position actuelle (fragile)
Prompt simple → output IA	Probablement pas protégeable
Prompt élaboré + curation	Zone grise
IA comme outil (modifications humaines significatives)	Probablement protégeable
Œuvre purement humaine	Protection classique

Le problème des données d'entraînement :

Les modèles IA sont entraînés sur des œuvres existantes. Souvent sans consentement des artistes. Des procès sont en cours (Getty, artistes individuels).

Ce qui va probablement arriver :

- Une législation plus claire sur les droits d'auteur IA
- Des modèles entraînés sur des données "propres" (licenciées)
- Des systèmes de rémunération des créateurs dont les œuvres ont servi à l'entraînement

13.8 L'avenir des métiers créatifs

Les métiers qui souffriront le plus :

- Illustration stock
- Rédaction de contenu générique
- Photographie commerciale basique
- Design template

Les métiers qui s'adapteront :

- Direction artistique (guider l'IA)
- Curation (sélectionner parmi les outputs)
- Personnalisation (adapter pour des clients spécifiques)
- Storytelling (l'histoire derrière l'œuvre)

Les métiers qui resteront :

- Art de haut niveau (galeries, collections)
- Création authentique avec une voix unique
- Spectacle vivant (musique, théâtre, danse)
- Artisanat de luxe

13.9 Rester créatif à l'ère de l'IA

Conseils pour les créateurs :

1. **Développez votre voix unique.** Ce que personne d'autre (ni aucune IA) ne peut reproduire.
 2. **Racontez votre histoire.** Le parcours, l'intention, le vécu font partie de l'œuvre.
 3. **Maîtrisez l'IA comme outil.** Mieux vaut savoir l'utiliser que la subir.
 4. **Explorez les nouvelles formes.** La collaboration homme-machine est un territoire vierge.
 5. **Construisez une communauté.** Les fans suivent un créateur, pas un output.
 6. **Diversifiez vos revenus.** Ne dépendez pas d'un seul type de création.
-

13.10 La créativité humaine irremplaçable

Ce que vous avez et que l'IA n'aura jamais :

- **Votre histoire** : Personne n'a vécu votre vie
- **Votre perspective** : Votre façon unique de voir le monde
- **Votre courage** : La capacité de vous exposer
- **Votre évolution** : Un parcours qui se transforme avec le temps
- **Votre présence** : Être là, en chair et en os

L'IA peut simuler tout ça.

Mais la simulation n'est pas la chose.

13.11 Un nouveau rapport à la création

Peut-être que l'IA nous oblige à repenser ce qu'est créer.

Pendant longtemps, créer signifiait : produire quelque chose de nouveau. L'IA produit du "nouveau" à l'infini.

La nouvelle définition de créer :

Peut-être que créer, ce n'est pas produire. C'est exprimer. C'est communiquer. C'est partager quelque chose de soi.

L'IA peut produire des millions d'images. Elle ne peut pas partager une expérience humaine.

La créativité humaine n'est pas menacée.

Elle est simplement invitée à se redéfinir.

Ce que vous pouvez faire

1. **Créez.** N'attendez pas de savoir si c'est "utile" face à l'IA. Créez pour vous.
2. **Explorez les outils IA.** Pas pour vous remplacer, mais pour vous augmenter.
3. **Développez votre authenticité.** C'est votre avantage unique.
4. **Soutenez les créateurs humains.** Achetez de l'art fait par des humains.

5. **Questionnez ce qui vous touche.** Pourquoi certaines œuvres vous émeuvent-elles ? C'est probablement humain.
-

Fin de la Partie V — Et demain ?

Nous avons exploré : - L'avenir de l'IA (Chapitre 18) - Les emplois transformés (Chapitre 19) - L'éducation repensée (Chapitre 20) - La créativité humaine (Chapitre 21)

Il nous reste la dernière partie : aller plus loin.

Chapitre suivant : Ressources pour comprendre l'IA

Chapitre 14

Questions aux journalistes — l'esprit critique

Vous lisez des articles sur l'IA tous les jours.

"L'IA va tout révolutionner." "L'IA menace des millions d'emplois." "L'IA devient consciente." "L'IA va nous sauver / nous détruire."

Mais qui vérifie ces affirmations ?

14.1 Le problème du journalisme tech

Le cycle médiatique de l'IA :

1. Une entreprise fait une annonce spectaculaire
2. Les médias reprennent l'annonce sans vérification
3. Le buzz génère des clics
4. La nuance arrive... quelques semaines plus tard
5. Personne ne lit la correction

Pourquoi ça arrive :

- **Pression du temps** : publier avant les concurrents
 - **Complexité technique** : peu de journalistes comprennent vraiment l'IA
 - **Dépendance aux sources** : les entreprises contrôlent l'information
 - **Modèle économique** : le sensationnalisme génère des clics
-

14.2 Les signaux d'alarme

Méfiez-vous quand vous lisez :

[X] "L'IA peut maintenant..." → Peut-elle vraiment ? Dans quelles conditions ?

[X] "Les experts affirment..." → Quels experts ? Payés par qui ?

[X] "Bientôt l'IA sera capable de..." → Source ? Échéance réelle ?

[X] "L'IA a réussi à..." → Un cas de démonstration ou un usage réel ?

[X] "X% des emplois menacés" → Étude financée par qui ? Méthodologie ?

Les bons réflexes :

[OK] L'article cite-t-il des sources vérifiables? **[OK]** Y a-t-il des voix critiques ou seulement des enthousiastes? **[OK]** Le journaliste a-t-il testé lui-même ce qu'il décrit? **[OK]** Les limites sont-elles mentionnées? **[OK]** Qui a intérêt à ce que cette histoire soit publiée?

14.3 Les conflits d'intérêts cachés

Les entreprises IA ont des budgets communication énormes.

Cela signifie : - Conférences de presse soigneusement orchestrées - Accès privilégié aux journalistes "amicaux" - Démonstrations impressionnantes mais non représentatives - "Experts" qui sont en fait des consultants payés

Les médias tech ont des problèmes structurels :

- Publicité des entreprises qu'ils couvrent
- Événements sponsorisés
- Relations privilégiées à maintenir
- Peur d'être exclus des annonces futures

Ce que ça signifie pour vous :

L'information que vous recevez est filtrée par des intérêts économiques.

14.4 Questions à poser aux journalistes

Quand vous lisez un article sur l'IA :

Question	Ce que ça révèle
Avez-vous testé personnellement ?	Vérification de première main
Qui sont vos sources ?	Indépendance vs dépendance
Quelles sont les limites de ce système ?	Équilibre de l'article
Qui finance cette recherche/entreprise ?	Conflits d'intérêts
Existe-t-il des voix critiques ?	Pluralité des points de vue
Quelle est votre expertise en IA ?	Compétence du journaliste

14.5 Le problème des "experts"

Qui parle d'IA dans les médias ?

- **PDG et dirigeants** : intérêt commercial évident
- **Investisseurs** : veulent que le marché croie au potentiel

- **Consultants** : vendent des services liés à l’IA
- **Chercheurs financés** : leurs labs dépendent des entreprises
- **Futuristes** : vivent de prédictions spectaculaires

Les voix moins entendues :

- Chercheurs indépendants critiques
- Travailleurs impactés par l’IA
- Spécialistes d’éthique
- Régulateurs
- Citoyens ordinaires

La règle :

Quand un “expert” parle d’IA, demandez-vous : “Quel est son intérêt dans cette affaire ?”

14.6 Les médias et l’IA : le serpent qui se mord la queue

Les médias utilisent de plus en plus l’IA :

- Rédaction automatisée d’articles simples
- Résumés générés
- Personnalisation des contenus
- Modération automatique des commentaires

Le paradoxe :

Les journalistes qui écrivent sur l’IA sont eux-mêmes remplacés par l’IA.

Les questions à poser :

- Cet article a-t-il été écrit/assisté par une IA ?
- Quel est le contrôle éditorial sur les contenus générés ?
- Comment distinguer le travail humain du travail machine ?

Certains médias commencent à étiqueter les contenus assistés par IA. La plupart ne le font pas encore.

14.7 Comment s’informer correctement

Diversifiez vos sources :

Type de source	Exemples	Forces	Limites
Médias généralistes	Le Monde, Libé, Figaro	Accessibilité	Parfois superficiel
Médias tech	The Verge, Wired, Numerama	Expertise	Trop proche de l’industrie
Publications académiques	ArXiv, Nature	Rigueur	Technique, lent
Newsletters spécialisées	Import AI, The Batch	Synthèse	Biais de l’auteur

Type de source	Exemples	Forces	Limites
Rapports institutionnels	CNIL, OECD	Neutralité	Parfois en retard
Témoignages directs	Utilisateurs, travailleurs	Vécu réel	Non représentatif

Les bonnes pratiques :

1. **Trianglez** : vérifiez une info dans au moins 3 sources différentes
2. **Attendez** : les premières 24h sont souvent les plus fausses
3. **Remontez aux sources** : lisez l'étude originale, pas le résumé
4. **Suivez les corrections** : les médias sérieux corrigent leurs erreurs
5. **Méfiez-vous du consensus** : quand tout le monde dit la même chose, demandez pourquoi

Le fact-checking appliqué à l'IA

Quelques ressources utiles :

- **AFP Factuel** : vérifie les affirmations virales
- **Hoaxbuster** : identifie les fausses informations
- **Sites des chercheurs** : MIT Technology Review, Stanford HAI

Vos propres vérifications :

- Cherchez l'étude originale citée
- Regardez qui l'a financée
- Cherchez les critiques de cette étude
- Comparez avec d'autres études sur le même sujet

Ce que vous pouvez faire

1. **Soyez un lecteur critique.** Ne prenez rien pour argent comptant, surtout en matière d'IA.
2. **Posez des questions dans les commentaires.** Demandez des sources, des précisions.
3. **Soutenez le journalisme de qualité.** Les enquêtes approfondies coûtent cher.
4. **Partagez avec discernement.** Ne relayez pas les annonces sensationnalistes.
5. **Formez-vous.** Plus vous comprendrez l'IA, plus vous pourrez évaluer ce qu'on vous en dit.

Le test ultime

Quand vous lisez un article sur l'IA, posez-vous ces 5 questions :

1. **Qui parle ?** (journaliste, expert, entreprise)
2. **Qui paie ?** (financement de la recherche, du média)
3. **Qui gagne ?** (à ce que cette histoire soit crue)
4. **Que cache-t-on ?** (limites, échecs, alternatives)
5. **Ai-je vérifié ailleurs ?** (au moins une autre source)

Si vous ne pouvez pas répondre à ces questions, l'article ne mérite peut-être pas votre confiance.

Fin de la Partie III — Les bonnes questions

Vous savez maintenant quelles questions poser : - Aux entreprises tech (Chapitre 10) - À votre employeur (Chapitre 11) - Aux politiques (Chapitre 12) - Aux médias (Chapitre 13)

Passons à la partie pratique : comment utiliser l'IA en citoyen éclairé.

Chapitre suivant : Évaluer un outil IA avant de l'utiliser

Chapitre 15

Conclusion — Le citoyen augmenté

Nous y voilà.

Vous avez fait le chemin.

Vous savez maintenant ce qu'est l'IA — et ce qu'elle n'est pas. Vous connaissez ses pouvoirs — et ses faiblesses. Vous comprenez les enjeux — économiques, politiques, éthiques. Vous avez les outils — pour évaluer, vous protéger, agir.

Que faire de tout cela ?

15.1 Ce que vous avez appris

[Esprit] Avertissement honnête :

Même après cette lecture, certaines illusions persisteront. Quand l'IA répond avec fluidité, votre cerveau attribuera *quand même* une compréhension. Quand elle reformule avec pertinence, vous penserez *quand même* qu'elle "sait". C'est normal : le langage est un piège cognitif. L'important n'est pas d'éliminer ces intuitions — c'est de savoir qu'elles sont fausses.

Partie I — Comprendre la bête

L'IA n'est pas magique. Ce sont des statistiques très sophistiquées appliquées à d'énormes quantités de données. Les LLM prédisent le prochain mot. Ils ne "comprennent" pas vraiment. Ils sont impressionnants ET fondamentalement limités.

Partie II — Les géants et nous

Quelques entreprises américaines et chinoises contrôlent l'infrastructure de l'IA mondiale. L'Europe est en retard mais a des cartes à jouer — notamment la régulation. Vos données sont leur trésor. Le RGPD et l'AI Act vous protègent, mais vous devez connaître vos droits pour les faire valoir.

Partie III — Les bonnes questions

Savoir questionner les entreprises tech, votre employeur, les politiques, les médias. Les bonnes questions sont votre première ligne de défense face au bullshit technologique.

Partie IV — Utiliser l'IA en citoyen éclairé

Évaluer avant d'adopter. Protéger ses données. Résister à la manipulation. Savoir quand NE PAS utiliser l'IA. L'outil doit vous servir, pas vous asservir.

Partie V — Et demain ?

Les emplois vont changer, pas disparaître. L'éducation doit se réinventer. La créativité humaine reste irremplaçable. L'avenir dépend de nos choix, pas de la technologie.

15.2 Les trois postures possibles

1. Le technophobe

“L'IA est dangereuse, il faut l'interdire.”

Cette posture est compréhensible mais inefficace. L'IA est là. Elle ne disparaîtra pas. Refuser de l'utiliser, c'est se priver d'outils que d'autres utiliseront contre vous.

2. Le technophile naïf

“L'IA est formidable, elle va tout résoudre.”

Cette posture est séduisante mais dangereuse. Elle ignore les risques, les biais, les concentrations de pouvoir. Elle fait de vous un consommateur passif, pas un citoyen actif.

3. Le citoyen augmenté

“L'IA est un outil puissant que je dois comprendre, maîtriser et orienter.”

C'est la posture que ce livre vous invite à adopter. Ni peur ni naïveté. Lucidité et action.

15.3 Le citoyen augmenté

Ce que ça signifie :

- **Comprendre** : Savoir ce qu'est l'IA, comment elle fonctionne, qui la contrôle
- **Utiliser** : Maîtriser les outils, les mettre au service de vos objectifs
- **Protéger** : Défendre vos données, votre vie privée, votre autonomie
- **Questionner** : Ne jamais accepter sans examiner
- **Participer** : S'engager dans le débat public sur l'avenir de l'IA

Ce que ça n'est pas :

- Être expert technique
- Utiliser toutes les IA disponibles
- Avoir une opinion sur tout
- Être parfait dans ses pratiques

Le citoyen augmenté n'est pas un super-humain. C'est un humain qui a choisi de ne pas subir.

15.4 Vos responsabilités

Envers vous-même :

- Apprendre en continu
- Protéger votre vie privée
- Maintenir votre esprit critique
- Préserver votre humanité

Envers vos proches :

- Partager ce que vous savez
- Protéger les plus vulnérables (enfants, personnes âgées)
- Aider ceux qui sont dépassés
- Montrer l'exemple

Envers la société :

- Participer au débat démocratique
 - Exiger de la transparence
 - Soutenir les alternatives éthiques
 - Voter en connaissance de cause
-

15.5 Les actions concrètes

Demain :

Action	Temps
Paramétrer les options de confidentialité d'un outil IA que vous utilisez	10 min
Vérifier une affirmation lue dans un article sur l'IA	5 min
Expliquer ce qu'est ChatGPT à un proche	15 min

Cette semaine :

Action	Temps
Installer un gestionnaire de mots de passe si vous n'en avez pas	30 min
Tester un outil IA que vous n'avez jamais utilisé	1h
Lire une ressource du chapitre 22	1h

Ce mois :

Action	Temps
Faire le point sur vos usages IA : qu'utilisez-vous ? Pour quoi ?	1h
Identifier une compétence que l'IA ne peut pas remplacer et la développer	Variable

Action	Temps
Avoir une conversation sur l'IA avec quelqu'un d'un autre avis	1h

Cette année :

Action	Impact
Suivre une formation ou lire un livre approfondi sur l'IA	Culture solide
Participer à une consultation publique sur l'IA	Impact citoyen
Évaluer l'impact de l'IA sur votre métier et vous préparer	Avenir professionnel
Transmettre ce que vous avez appris	Effet multiplicateur

15.6 Le monde que nous voulons

L'IA amplifie ce que nous sommes.

Si nous sommes cupides, l'IA amplifiera notre cupidité. Si nous sommes créatifs, l'IA amplifiera notre créativité. Si nous sommes solidaires, l'IA amplifiera notre solidarité.

La question n'est pas : "L'IA sera-t-elle bonne ou mauvaise?"

La question est : "Quel usage en ferons-nous?"

15.7 Une utopie accessible

Imaginez un monde où :

- L'IA aide les médecins à mieux diagnostiquer, mais ne remplace pas la relation de soin
- L'IA personnalise l'éducation, mais les enseignants restent des mentors
- L'IA automatise le travail pénible, mais les gains sont partagés équitablement
- L'IA crée de l'art, mais l'authenticité humaine garde sa valeur
- L'IA surveille moins et protège plus
- L'Europe a son infrastructure, ses modèles, sa souveraineté
- Chaque citoyen comprend assez pour ne pas être manipulé

Ce monde n'est pas garanti.

Mais il n'est pas impossible.

Il dépend de nous.

15.8 Le mot de la fin

Ce livre s'appelait "L'IA et Nous — Le guide du cyber citoyen".

Pas "L'IA contre nous". Pas "L'IA à notre place".

"L'IA et nous."

Parce que nous sommes dans cette histoire ensemble. Parce que l'IA n'existe que par nous. Parce que son avenir est entre nos mains.

Vous avez maintenant les clés.

Pas toutes les clés — personne ne les a. Mais assez pour commencer. Assez pour comprendre. Assez pour agir.

L'IA ne décide pas de notre avenir.

Nous décidons.

Ensemble.

Remerciements

Ce livre n'aurait pas existé sans :

- Les chercheurs qui démystifient l'IA
- Les journalistes qui enquêtent
- Les régulateurs qui protègent
- Les citoyens qui questionnent
- Les lecteurs qui cherchent à comprendre

Merci à vous d'avoir lu jusqu'ici.

Une dernière chose

Si ce livre vous a été utile, partagez-le.

Pas pour moi. Pour que d'autres citoyens aient les mêmes clés.

L'éducation au numérique est un bien commun. Plus nous sommes nombreux à comprendre, moins nous sommes vulnérables.

Pour aller plus loin

Le livre compagnon académique :

Ce guide citoyen a un frère : *"From Rational Agents to Governed Cognitive Systems : A Foundational Inquiry into Large Language Model-Based Assistants"*.

C'est un ouvrage académique, en anglais, destiné aux chercheurs et aux étudiants avancés. Il explore les mêmes questions avec rigueur scientifique, propose un nouveau cadre théorique, et s'inscrit dans le débat académique international.

Les deux livres peuvent se lire indépendamment. Ensemble, ils forment une vision complète : pratique et théorique, citoyenne et académique.

Fin.

Mais ce n'est que le début de votre parcours de citoyen augmenté.

À propos de l'auteur

[Votre biographie ici]

Contact

[Vos coordonnées pour les lecteurs qui veulent vous contacter]

Licence

Ce livre est publié sous licence Creative Commons BY-NC-SA 4.0. Vous pouvez le partager et l'adapter, à condition de : - Citer l'auteur - Ne pas en faire un usage commercial - Partager vos adaptations sous la même licence

Épilogue

L'IA et Nous — Le choix qui reste

Vous avez lu ce livre jusqu'au bout.

Peut-être êtes-vous venu chercher des réponses. Peut-être en repartez-vous avec encore plus de questions. C'est normal. C'est même souhaitable.

L'intelligence artificielle n'est pas un sujet qu'on "maîtrise" une fois pour toutes. C'est un paysage en mouvement perpétuel, une révolution technologique qui se déroule sous nos yeux, en temps réel, pendant que vous lisez ces lignes.

Mais une chose ne change pas : **vous avez le choix.**

Ni technophobe, ni technobéat

Tout au long de ce livre, j'ai essayé de naviguer entre deux écueils.

Le premier écueil, c'est la technophobie. La peur de l'IA. Le rejet viscéral. L'envie de fermer les yeux, de débrancher, de revenir à un monde "d'avant" qui n'existe plus. Cette posture est compréhensible — face à l'inconnu, la peur est une réaction naturelle. Mais elle est aussi dangereuse. Refuser de comprendre l'IA, c'est se condamner à la subir. C'est laisser d'autres décider pour vous.

Le deuxième écueil, c'est la technobéatitude. L'émerveillement naïf. La croyance que l'IA va tout résoudre — le réchauffement climatique, les maladies, les inégalités, et peut-être même notre bonheur personnel. Cette posture est séduisante — elle nous dispense de réfléchir, de critiquer, de nous responsabiliser. Mais elle est tout aussi dangereuse. Faire confiance aveuglément à l'IA, c'est oublier qu'elle est créée par des entreprises, avec des intérêts, des biais, des limites.

La voie du milieu existe. Elle s'appelle la **lucidité informée**.

Comprendre ce que l'IA fait réellement (et ne fait pas). Connaître ses droits (et les exercer). Utiliser les outils quand ils sont utiles (et s'en passer quand ils ne le sont pas). Rester critique (sans être cynique). Rester ouvert (sans être crédule).

Cette voie demande un effort. Mais c'est la seule qui vous garde libre.

Citoyen, pas consommateur

Il y a deux façons de considérer l'IA.

La première : comme un produit de consommation. On télécharge l'application, on crée un compte, on utilise le service, on passe à autre chose. C'est pratique. C'est passif. C'est exactement ce que les géants de la tech attendent de nous.

La deuxième : comme un enjeu de société. On s'informe. On questionne. On exige. On vote. On participe au débat public. On contribue à façonner les règles du jeu. C'est plus exigeant. C'est aussi plus digne.

Vous n'êtes pas un utilisateur. Vous êtes un citoyen.

Et en tant que citoyen, vous avez des droits que les consommateurs n'ont pas : - Le droit de savoir ce qu'on fait de vos données. - Le droit de refuser les systèmes automatisés qui vous affectent. - Le droit d'exiger de la transparence. - Le droit de demander des comptes.

Le RGPD vous donne ces droits. L'AI Act les renforce. Mais un droit qu'on n'exerce pas est un droit qui meurt.

Exercez vos droits. C'est ainsi qu'ils restent vivants.

L'Europe comme laboratoire

Ce livre est écrit depuis la Suisse, aux portes de l'Union européenne, avec une conviction profonde : **l'Europe a un rôle unique à jouer.**

Non, nous n'avons pas créé ChatGPT. Nous n'avons pas les géants américains ni les milliards chinois. Nous avons autre chose : une tradition de régulation, une attention aux droits fondamentaux, une méfiance saine envers les monopoles.

Le RGPD a changé la vie privée mondiale. L'AI Act est la première législation complète sur l'intelligence artificielle. Ce n'est pas rien.

Bien sûr, l'Europe fait des erreurs. Elle est parfois lente, bureaucratique, divisée. Mais elle pose les bonnes questions : - Comment utiliser l'IA sans sacrifier la vie privée ? - Comment innover sans créer des monopoles incontrôlables ? - Comment automatiser sans déshumaniser ?

L'Europe n'est peut-être pas le laboratoire de l'IA. Elle peut être le laboratoire de l'IA gouvernable.

Et ça, c'est peut-être plus important à long terme.

Ce que vous pouvez faire dès demain

Je ne veux pas terminer ce livre sur des généralités. Voici des actions concrètes, que vous pouvez entreprendre dès que vous aurez refermé ces pages :

1. Faites l'inventaire Quels outils IA utilisez-vous déjà ? ChatGPT, Copilot, Google, des applications sur votre téléphone ? Listez-les. C'est la première étape.

2. Lisez les conditions d'utilisation Je sais, c'est pénible. Mais faites-le une fois, pour un service que vous utilisez beaucoup. Cherchez : où vont mes données ? Sont-elles utilisées pour entraîner le modèle ? Pendant combien de temps sont-elles conservées ?

3. Activez les options de confidentialité La plupart des services proposent des options pour limiter la collecte de données. Elles sont souvent cachées, désactivées par défaut. Trouvez-les. Activez-les.

4. Posez une question à votre employeur Si vous utilisez l'IA au travail : "Quelle est notre politique IA ?" Si votre entreprise n'en a pas, c'est un problème.

5. Parlez-en autour de vous L'IA n'est pas un sujet de geeks. C'est un sujet de société. Parlez-en à vos amis, votre famille, vos collègues. Partagez ce livre. Lancez la conversation.

6. Restez informé L'IA évolue vite. Trop vite pour un livre imprimé. Suivez quelques sources fiables (voir Annexe D). Ne vous laissez pas submerger, mais ne vous laissez pas non plus distancer.

7. Votez Les prochaines élections — locales, nationales, européennes — auront un impact sur la régulation de l'IA. Informez-vous sur les positions des candidats. Faites de l'IA un critère de choix.

Le mot de la fin

L'intelligence artificielle n'est ni notre sauveur ni notre bourreau.

C'est un outil. Extraordinairement puissant. Potentiellement dangereux. Et fondamentalement **entre nos mains**.

Les algorithmes ne décident de rien. Ce sont des humains qui les conçoivent, les déploient, les régulent — ou ne les régulent pas. Des humains qui travaillent dans des entreprises. Des humains qui siègent dans des parlements. Des humains qui utilisent des applications.

Des humains comme vous.

L'avenir de l'IA n'est pas écrit. Il s'écrit maintenant. Et vous tenez le stylo.

Alors écrivez bien.

"La question n'est pas de savoir si les machines peuvent penser. La question est de savoir si les humains continuent de le faire."

Merci d'avoir lu ce livre.

L'aventure continue.

Ivan Googé Décembre 2025

Annexe A

Glossaire — Le lexique de l'IA démystifié

L'IA a son jargon.

Cette annexe est votre dictionnaire de survie.

Gardez-la à portée de main.

A

AGI (Artificial General Intelligence) Intelligence artificielle générale. Une IA hypothétique qui serait capable de faire tout ce qu'un humain peut faire intellectuellement. N'existe pas encore. Peut-être n'existera jamais.

Agent Entité logicielle autonome, définie par un rôle et un domaine, capable de percevoir, raisonner et agir dans les limites de règles explicites. Contrairement à un simple chatbot qui répond à des questions, un agent peut planifier des séquences d'actions, utiliser des outils, et poursuivre un objectif. **Attention** : un vrai agent (au sens rigoureux) est *responsable* de ce qu'il produit, doit pouvoir expliquer ses décisions, et laisse des traces vérifiables. Sans ces propriétés, ce n'est pas un agent — c'est une fonction avec du marketing. Les "agents autonomes" actuels (2024-2025) sont souvent des démos impressionnantes mais sans garanties de fiabilité.

Algorithme Une suite d'instructions pour résoudre un problème. Comme une recette de cuisine, mais pour un ordinateur.

Alignement Le problème de s'assurer qu'une IA fait ce qu'on veut vraiment qu'elle fasse, pas juste ce qu'on lui a dit de faire. Un des grands défis de la sécurité de l'IA.

API (Application Programming Interface) Interface qui permet à un programme d'utiliser un autre programme. Vous n'utilisez pas directement le modèle GPT-4 ; vous passez par l'API d'OpenAI.

Attention (mécanisme d') Technique qui permet à un modèle de "se concentrer" sur les parties pertinentes d'un texte. C'est ce qui permet aux Transformers de comprendre les relations à longue distance dans un texte.

B

Biais Quand un système IA donne des résultats systématiquement incorrects ou injustes pour certains groupes. Exemple : un système de recrutement qui défavorise les femmes parce qu'il a été entraîné sur des données historiques biaisées.

Big Tech Les géants de la technologie : Google, Meta, Microsoft, Amazon, Apple. Parfois étendu à des acteurs chinois comme Baidu ou ByteDance.

Blackbox (boîte noire) Quand on ne peut pas comprendre comment un système prend ses décisions. On voit les entrées et les sorties, mais pas ce qui se passe à l'intérieur.

C

Chatbot Programme qui peut avoir une conversation textuelle. ChatGPT est un chatbot très avancé.

Claude Modèle de langage développé par Anthropic, concurrent de GPT. Conçu avec un accent sur la sécurité et l'alignement.

Cloud Serveurs distants accessibles par Internet. Quand vous utilisez ChatGPT, vous utilisez des serveurs dans le cloud (chez Microsoft Azure).

Cloud Act Loi américaine de 2018 qui permet au gouvernement US d'accéder aux données stockées par des entreprises américaines, même si les serveurs sont à l'étranger.

Context window (fenêtre de contexte) La quantité de texte qu'un modèle peut "garder en mémoire" pour une conversation. GPT-4 : ~128K tokens. Claude : ~200K tokens. Plus c'est grand, plus le modèle peut travailler avec de longs documents.

Copilot Nom de l'assistant IA de Microsoft, basé sur GPT d'OpenAI. Intégré dans Windows, Office, Bing, etc.

Copyright Droit d'auteur. Question non résolue : qui possède les œuvres générées par IA ?

D

Data (données) L'information utilisée pour entraîner les modèles IA. La qualité des données détermine largement la qualité du modèle.

Deep learning (apprentissage profond) Technique d'apprentissage automatique utilisant des réseaux de neurones à plusieurs couches. C'est ce qui a permis les percées récentes en IA.

Deepfake Contenu audio ou vidéo généré par IA qui fait dire ou faire à quelqu'un quelque chose qu'il n'a jamais dit ou fait.

Diffusion (modèle de) Architecture utilisée pour la génération d'images (Stable Diffusion, DALL-E 3, Midjourney). Fonctionne en apprenant à "débruiter" une image.

DPO (Délégué à la Protection des Données) Personne responsable de la conformité RGPD dans une organisation. Votre contact pour les questions de données personnelles.

E

Embedding Représentation numérique d'un mot, d'une phrase ou d'un document sous forme de vecteur. Permet aux machines de "comprendre" la proximité sémantique.

Emergent behavior (comportement émergent) Capacités qui apparaissent dans les grands modèles sans avoir été explicitement programmées. Par exemple, les LLM qui "apprennent" à raisonner.

Éthique de l'IA Domaine qui étudie les implications morales de l'IA : biais, vie privée, emploi, autonomie, etc.

F

Few-shot learning Capacité d'un modèle à apprendre une nouvelle tâche avec seulement quelques exemples. "Voici 3 exemples de ce que je veux, maintenant fais-le."

Fine-tuning Processus d'adaptation d'un modèle pré-entraîné à une tâche spécifique en le réentraînant sur des données supplémentaires.

Foundation model (modèle de fondation) Grand modèle pré-entraîné qui sert de base pour diverses applications. GPT-4, Claude, Llama sont des modèles de fondation.

G

Gemini Modèle de langage de Google, successeur de Bard.

Generative AI (IA générative) IA capable de créer du nouveau contenu : texte, images, musique, vidéo. ChatGPT, DALL-E, Midjourney sont des IA génératives.

GPT (Generative Pre-trained Transformer) Architecture de modèle d'OpenAI. GPT-3, GPT-4, GPT-4o sont des versions successives.

GPU (Graphics Processing Unit) Processeur graphique. Ironiquement, ces puces conçues pour les jeux vidéo sont idéales pour l'IA. NVIDIA domine ce marché.

Ground truth (vérité de référence) Les données correctes utilisées pour entraîner ou évaluer un modèle. Si la ground truth est biaisée, le modèle sera biaisé.

Guardrails Protections mises en place pour empêcher un modèle de produire des contenus dangereux ou inappropriés.

H

Hallucination Quand un modèle génère des informations fausses présentées comme vraies. ChatGPT qui invente des citations ou des faits qui n'existent pas.

Hugging Face Plateforme majeure pour les modèles IA open source. Le "GitHub de l'IA".

I

Inference (inférence) Le processus de faire fonctionner un modèle entraîné pour obtenir des prédictions. Quand vous posez une question à ChatGPT, le modèle fait de l'inférence.

Intelligence artificielle (IA) Systèmes informatiques capables de réaliser des tâches qui nécessitent normalement l'intelligence humaine.

J

Jailbreak Technique pour contourner les protections d'un modèle IA et lui faire produire des contenus normalement interdits.

L

Latent space (espace latent) L'espace mathématique interne où un modèle représente les concepts. Difficile à visualiser, mais crucial pour comprendre comment les modèles "pensent".

LLM (Large Language Model) Grand modèle de langage. Les modèles comme GPT-4, Claude, Llama qui ont été entraînés sur d'énormes quantités de texte.

Llama Famille de modèles de langage open source de Meta (Facebook). Alternative ouverte aux modèles propriétaires.

M

Machine learning (apprentissage automatique) Technique où un programme apprend à partir de données plutôt que d'être explicitement programmé.

Midjourney Service populaire de génération d'images par IA. Accessible via Discord.

Mistral Startup française d'IA, créatrice de modèles de langage open source. Espoir européen face aux géants américains.

Modèle Programme entraîné sur des données pour effectuer une tâche. GPT-4 est un modèle.

Multimodal IA capable de traiter plusieurs types de données : texte, images, audio, vidéo. GPT-4V est multimodal (comprend les images).

N

Neural network (réseau de neurones) Architecture informatique inspirée du cerveau humain. Des “neurones” artificiels connectés en couches.

NLP (Natural Language Processing) Traitement automatique du langage naturel. Le domaine qui permet aux machines de comprendre et générer du langage humain.

O

Open source Logiciel dont le code source est accessible et modifiable par tous. Llama est open source (avec restrictions), GPT-4 ne l’est pas.

OpenAI Entreprise créatrice de ChatGPT, GPT-4, DALL-E. Fondée comme ONG, devenue entreprise à but lucratif plafonné.

Overfitting (surapprentissage) Quand un modèle est trop bien adapté à ses données d’entraînement et généralise mal à de nouvelles données.

P

Parameter (paramètre) Les “réglages” internes d’un modèle. GPT-4 a ~1 trillion de paramètres. Plus il y en a, plus le modèle est potentiellement capable (et coûteux).

Pre-training (pré-entraînement) Phase initiale d’entraînement d’un modèle sur de grandes quantités de données générales, avant le fine-tuning.

Prompt L’instruction ou la question que vous donnez à un modèle IA. L’art du prompt est de formuler des instructions qui donnent de bons résultats.

Prompt engineering L’art de formuler des prompts efficaces. Une compétence de plus en plus valorisée.

R

RAG (Retrieval-Augmented Generation) Technique qui combine un LLM avec une base de données externe. Permet de répondre à des questions sur des documents spécifiques.

Reasoning (raisonnement) Capacité d’un modèle à suivre une logique, étape par étape. Les modèles récents (GPT-4, Claude) montrent des capacités de raisonnement émergentes.

Red teaming Pratique de tester un système en essayant de le faire échouer ou de contourner ses protections. Utilisé pour améliorer la sécurité des modèles.

Reinforcement Learning (apprentissage par renforcement) Technique où un modèle apprend par essai-erreur, en recevant des récompenses ou des punitions. Utilisé pour affiner les comportements des chatbots.

RLHF (Reinforcement Learning from Human Feedback) Technique où des humains évaluent les réponses d'un modèle pour l'améliorer. C'est ce qui rend ChatGPT "poli" et utile.

RGPD (Règlement Général sur la Protection des Données) Loi européenne sur la protection des données personnelles. Vos droits fondamentaux face aux données.

S

Scaling laws (lois de mise à l'échelle) Observation que les performances des modèles s'améliorent de façon prévisible quand on augmente leur taille et leurs données d'entraînement.

Stable Diffusion Modèle open source de génération d'images. Peut être exécuté localement, contrairement à DALL-E ou Midjourney.

System prompt Instructions cachées données au modèle avant votre conversation. Définit le comportement, la personnalité, les limites.

T

Temperature Paramètre qui contrôle la "créativité" d'un modèle. Température basse = réponses prévisibles. Température haute = réponses plus variées (mais parfois incohérentes).

Token Unité de base du texte pour un LLM. Environ 0,75 mot en anglais, variable en français. Les modèles comptent en tokens, pas en mots.

Training (entraînement) Processus par lequel un modèle apprend à partir de données. Peut prendre des mois et coûter des millions de dollars pour les grands modèles.

Transformer Architecture de réseau de neurones qui a révolutionné le NLP. Le "T" de GPT. Utilise le mécanisme d'attention.

V

Vector database (base de données vectorielle) Base de données optimisée pour stocker et rechercher des embeddings. Essentielle pour les systèmes RAG.

W

Weights (poids) Les valeurs numériques qui définissent un modèle entraîné. "Télécharger les weights" = télécharger le modèle.

Whisper Modèle de transcription audio d'OpenAI. Peut transcrire et traduire de l'audio en texte.

Z

Zero-shot Capacité d'un modèle à effectuer une tâche sans aucun exemple préalable. "Fais ceci" sans montrer comment.

Ce que vous pouvez faire

1. **Gardez ce lexique accessible.** Consultez-le quand vous rencontrez un terme inconnu.
 2. **N'ayez pas peur du jargon.** Derrière chaque terme compliqué, il y a un concept souvent simple.
 3. **Utilisez ces termes.** Plus vous les utilisez, plus ils deviennent naturels.
-

Chapitre suivant : Conclusion — Le citoyen augmenté

Annexe B

Checklist : Avant d'utiliser un nouvel outil IA

15 questions à se poser avant d'adopter un outil IA. Imprimez cette page et gardez-la près de votre ordinateur.

[OK] La Checklist

Identité et confiance

- ☐ **1. Qui a créé cet outil ?** - Entreprise connue ou startup inconnue ? - Basée où ? (USA, Europe, Chine, autre ?) - Depuis combien de temps existe-t-elle ?
 - ☐ **2. Quel modèle utilise-t-il ?** - GPT-4, Claude, Gemini, Mistral, Llama, autre ? - Open source ou propriétaire ? - Version mise à jour quand pour la dernière fois ?
 - ☐ **3. Y a-t-il des avis indépendants ?** - Tests publiés par des médias reconnus ? - Retours d'utilisateurs vérifiables ? - Alertes de sécurité connues ?
-

Données et vie privée

- ☐ **4. Où sont stockées mes données ?** - Europe, États-Unis, Chine, autre ? - Cloud mutualisé ou instance dédiée ? - Serveurs identifiés ou "quelque part" ?
 - ☐ **5. Mes conversations sont-elles utilisées pour entraîner le modèle ?** - Opt-in (je dois dire oui) ou opt-out (c'est par défaut) ? - Puis-je refuser ? Comment ? - Est-ce clairement documenté ?
 - ☐ **6. Combien de temps mes données sont-elles conservées ?** - 30 jours ? 1 an ? Indéfiniment ? - Puis-je les supprimer ? Comment ? - La suppression est-elle réelle ou juste "cachée" ?
 - ☐ **7. Mes données peuvent-elles être lues par des humains ?** - Pour améliorer le service ? - Pour modération ? - Par des sous-traitants ?
 - ☐ **8. Y a-t-il un mode "privé" ou "incognito" ?** - Que protège-t-il exactement ? - Quelles sont ses limites ?
-

Usage et limites

- ☐ **9. À quoi sert vraiment cet outil ?** - Quel problème résout-il pour moi ? - Existe-t-il une alternative sans IA ? - Ai-je vraiment besoin de ça ?
 - ☐ **10. Quelles sont les limites reconnues ?** - Le fournisseur les documente-t-il honnêtement ? - Hallucinations ? Biais ? Erreurs courantes ? - Cas d'usage déconseillés ?
 - ☐ **11. Puis-je vérifier les réponses ?** - L'outil cite-t-il ses sources ? - Les sources sont-elles vérifiables ? - Ai-je les moyens de fact-checker ?
-

Coût et dépendance

- ☐ **12. Quel est le vrai coût ?** - Gratuit = mes données sont le produit ? - Version payante : que gagne-t-on vraiment ? - Coûts cachés (API, tokens, stockage) ?
 - ☐ **13. Puis-je exporter mes données et partir ?** - Format d'export standard ? - Mes créations m'appartiennent-elles ? - Vendor lock-in ?
 - ☐ **14. Que se passe-t-il si l'outil disparaît ?** - L'entreprise est-elle financièrement stable ? - Mes données sont-elles récupérables ? - Ai-je un plan B ?
-

Légalité et conformité

- ☐ **15. Est-ce conforme au RGPD et à l'AI Act ?** - Déclaration de conformité disponible ? - DPO (Data Protection Officer) identifiable ? - Possibilité de déposer une plainte si nécessaire ?
-

[Feu] Système de scoring

Comptez vos [OK] :

Score	Verdict
15/15	[VERT] Feu vert — Vous pouvez y aller sereinement
12-14	[JAUNE] Prudence — Quelques points à clarifier
8-11	[ORANGE] Attention — Risques significatifs non couverts
< 8	[ROUGE] Stop — Trop d'inconnues, cherchez une alternative

[Liste] Template à remplir

Outil évalué : _____

Date de l'évaluation : _____

#	Question	Réponse	[OK]/[X]
1	Créateur		
2	Modèle utilisé		
3	Avis indépendants		
4	Localisation données		
5	Entraînement sur mes données		
6	Durée de conservation		
7	Lecture humaine		
8	Mode privé		
9	Utilité réelle		
10	Limites documentées		
11	Vérifiabilité		
12	Coût réel		
13	Portabilité		
14	Pérennité		
15	Conformité RGPD/AI Act		

Score total : ___/15

Décision : ☐ Adopter ☐ Tester prudemment ☐ Chercher alternative ☐ Refuser

Notes : _____

[Lien] Version imprimable

Téléchargez la version PDF de cette checklist sur : **[lien à ajouter lors de la publication]**

Checklist v1.0 — Décembre 2025 Libre de droits pour usage personnel et éducatif

Annexe C

Vos droits en Europe (RGPD + AI Act)

Résumé pratique de vos droits face à l'IA en tant que citoyen européen.

Partie 1 : Le RGPD — Vos droits sur vos données

Le **Règlement Général sur la Protection des Données** (RGPD) est en vigueur depuis mai 2018. Il s'applique à toute entreprise qui traite des données de résidents européens, même si l'entreprise est américaine ou chinoise.

Vos 8 droits fondamentaux

1. Droit d'accès (Article 15)

Ce que c'est : Vous pouvez demander à une entreprise quelles données elle détient sur vous.

Comment l'exercer : Envoyez un email à l'adresse DPO (Data Protection Officer) de l'entreprise avec une pièce d'identité. Délai de réponse : 1 mois.

Exemple de formulation : > “Conformément à l'article 15 du RGPD, je vous demande de me communiquer l'ensemble des données personnelles que vous détenez me concernant, ainsi que les informations sur leur traitement.”

2. Droit de rectification (Article 16)

Ce que c'est : Vous pouvez faire corriger des données inexactes vous concernant.

Comment l'exercer : Identifiez l'erreur, contactez l'entreprise, demandez la correction.

3. Droit à l'effacement — “Droit à l'oubli” (Article 17)

Ce que c'est : Vous pouvez demander la suppression de vos données dans certains cas.

Quand ça s'applique : - Les données ne sont plus nécessaires - Vous retirez votre consentement - Vous vous opposez au traitement - Le traitement était illégal

Attention : Ce droit n'est pas absolu. L'entreprise peut refuser si elle a une obligation légale de conserver les données.

4. Droit à la limitation du traitement (Article 18)

Ce que c'est : Vous pouvez demander que vos données soient "gelées" — conservées mais plus utilisées.

Quand c'est utile : En attendant qu'un litige soit résolu, ou si vous contestez l'exactitude des données.

5. Droit à la portabilité (Article 20)

Ce que c'est : Vous pouvez récupérer vos données dans un format lisible par machine et les transférer à un autre service.

Exemple concret : Exporter vos conversations ChatGPT pour les importer dans un autre outil.

6. Droit d'opposition (Article 21)

Ce que c'est : Vous pouvez vous opposer à certains traitements de vos données, notamment le marketing direct.

Pour l'IA : Vous pouvez vous opposer à ce que vos conversations soient utilisées pour entraîner le modèle.

7. Droit de ne pas faire l'objet d'une décision automatisée (Article 22)

Ce que c'est : Vous avez le droit de ne pas être soumis à une décision basée uniquement sur un traitement automatisé si elle a des effets juridiques ou significatifs sur vous.

Exemples : Refus de crédit par algorithme, tri automatique de CV, décision d'assurance automatisée.

Ce que vous pouvez exiger : - Intervention humaine - Explication de la décision
- Possibilité de contester

8. Droit d'être informé (Articles 13-14)

Ce que c'est : Avant tout traitement, l'entreprise doit vous informer clairement de ce qu'elle fait avec vos données.

Ce qui doit être communiqué : - Identité du responsable - Finalités du traitement
- Durée de conservation - Vos droits - Existence de transferts hors UE

À qui se plaindre ?

En France : CNIL (Commission Nationale de l'Informatique et des Libertés) - Site : cnil.fr - Plainte en ligne : cnil.fr/plaintes

En Suisse : PFPDT (Préposé fédéral à la protection des données) - Site : edoeb.admin.ch

En Belgique : APD (Autorité de protection des données) - Site : autoriteprotection-donnees.be

Au Luxembourg : CNPD (Commission Nationale pour la Protection des Données) - Site : cnpd.public.lu

Procédure type : 1. Contactez d'abord l'entreprise (gardez une trace) 2. Attendez 1 mois 3. Si pas de réponse satisfaisante, déposez plainte auprès de votre autorité nationale 4. L'autorité enquête et peut sanctionner (amendes jusqu'à 4% du CA mondial)

Partie 2 : L'AI Act — Vos droits face aux systèmes IA

Le **Règlement européen sur l'intelligence artificielle** (AI Act) a été adopté en 2024 et entre en application progressivement (2025-2027).

Classification des systèmes IA par risque

[ROUGE] Risque inacceptable — INTERDIT

- Notation sociale à la chinoise
- Manipulation subliminale
- Exploitation des vulnérabilités (enfants, handicaps)
- Reconnaissance faciale en temps réel dans l'espace public (sauf exceptions)

[ORANGE] Haut risque — FORTEMENT RÉGULÉ

Systèmes utilisés dans : - Recrutement et gestion RH - Éducation (notation, orientation) - Services essentiels (crédit, assurance) - Maintien de l'ordre - Migration et asile - Justice

Obligations pour ces systèmes : - Évaluation des risques obligatoire - Données de haute qualité - Traçabilité et documentation - Supervision humaine obligatoire - Transparence envers les utilisateurs

[JAUNE] Risque limité — OBLIGATIONS DE TRANSPARENCE

- Chatbots (vous devez savoir que vous parlez à une IA)
- Deepfakes (doivent être étiquetés)

- Contenus générés par IA (doivent être identifiables)

[VERT] Risque minimal — PAS DE RÉGULATION SPÉCIFIQUE

- Filtres anti-spam
 - Jeux vidéo
 - La plupart des applications courantes
-

Vos droits spécifiques sous l'AI Act

1. Droit à l'explication

Pour les systèmes à haut risque, vous pouvez exiger une explication compréhensible de la décision.

2. Droit de contester

Vous pouvez contester toute décision prise par un système IA à haut risque.

3. Droit à l'intervention humaine

Vous pouvez demander qu'un humain révise la décision.

4. Droit d'être informé

Vous devez savoir quand vous interagissez avec une IA (obligation de transparence).

5. Droit de signaler

Vous pouvez signaler des systèmes IA problématiques aux autorités.

Calendrier d'entrée en vigueur

Date	Ce qui entre en vigueur
Février 2025	Interdiction des pratiques à risque inacceptable
Août 2025	Règles sur les modèles d'IA à usage général (GPT-4, Claude...)
Août 2026	Majorité des obligations (systèmes à haut risque)
Août 2027	Toutes les obligations

Partie 3 : Tableau récapitulatif

Situation	Votre droit	Base légale	Action
Je veux savoir quelles données OpenAI a sur moi	Droit d'accès	RGPD Art. 15	Email au DPO
Je veux que mes conversations ne servent pas à l'entraînement	Droit d'opposition	RGPD Art. 21	Paramètres + demande formelle
Je veux supprimer mon historique ChatGPT	Droit à l'effacement	RGPD Art. 17	Paramètres compte + demande
Un algorithme m'a refusé un crédit	Droit à l'explication + intervention humaine	RGPD Art. 22 + AI Act	Demande écrite à la banque
Je ne savais pas que je parlais à un bot	Violation transparence	AI Act	Signalement autorité
Une entreprise utilise mon visage dans un deepfake	Multiples violations	RGPD + AI Act	Plainte CNIL + dépôt plainte

Partie 4 : Modèles de lettres

Lettre type — Demande d'accès aux données

[Votre nom]
[Votre adresse]
[Date]

À l'attention du Délégué à la Protection des Données
[Nom de l'entreprise]
[Adresse]

Objet : Demande d'accès à mes données personnelles (Article 15 RGPD)

Madame, Monsieur,

Conformément à l'article 15 du Règlement Général sur la Protection des Données (UE) 2016/679, je vous demande de bien vouloir me communiquer :

1. L'ensemble des données personnelles me concernant que vous détenez
2. Les finalités du traitement
3. Les catégories de données concernées
4. Les destinataires des données
5. La durée de conservation prévue
6. L'existence d'une prise de décision automatisée

Vous trouverez ci-joint une copie de ma pièce d'identité.

Conformément à la réglementation, je vous remercie de me répondre dans un délai d'un mois.

Veuillez agréer, Madame, Monsieur, l'expression de mes salutations distinguées.

[Signature]

Lettre type — Opposition à l'entraînement IA

[Votre nom]

[Adresse email associée au compte]

[Date]

Objet : Opposition au traitement de mes données pour l'entraînement de modèles IA (Article 21 RGPD)

Madame, Monsieur,

En application de l'article 21 du RGPD, je m'oppose formellement à ce que mes données personnelles, incluant mes conversations et interactions avec vos services, soient utilisées pour l'entraînement de vos modèles d'intelligence artificielle.

Cette opposition est effective immédiatement et concerne toutes mes données passées et futures.

Je vous demande de confirmer la prise en compte de cette opposition dans un délai d'un mois.

Cordialement,

[Signature]

Ressources utiles

Textes officiels : - RGPD : eur-lex.europa.eu - AI Act : eur-lex.europa.eu

Guides pratiques : - Guide CNIL sur l'IA : cnil.fr/intelligence-artificielle - Guide pratique AI Act : digital-strategy.ec.europa.eu

Contacts autorités : - CNIL (France) : cnil.fr — Tél. : 01 53 73 22 22 - APD (Belgique) : autoriteprotectiondonnees.be - PFPDT (Suisse) : edoeb.admin.ch

Annexe mise à jour : Décembre 2025 Cette annexe ne constitue pas un conseil juridique. En cas de litige, consultez un avocat.

Annexe D

Ressources pour continuer

Sélection commentée de ressources fiables pour approfondir vos connaissances sur l'IA, organisées par profil.

[Guide] Pour tout le monde — Les incontournables

Newsletters francophones

La Lettre de l'IA — Flint - Fréquence : Hebdomadaire - Pourquoi : Synthèse équilibrée de l'actualité IA, ton accessible - S'inscrire : flint.media/newsletter-ia

Aujourd'hui l'IA - Fréquence : Quotidienne - Pourquoi : Court, factuel, va à l'essentiel - S'inscrire : aujourdhuilia.substack.com

AI Paris Newsletter - Fréquence : Hebdomadaire
- Pourquoi : Focus écosystème français et européen - S'inscrire : aiparis.fr

Podcasts francophones

Underscore_ (France Culture) - Pourquoi : Réflexion de fond, invités de qualité - Durée : 50 min - Où : Toutes plateformes + franceculture.fr

La Science, CQFD (France Culture) - Pourquoi : Quand ils parlent d'IA, c'est rigoureux - Durée : 60 min - Où : franceculture.fr

Le Podcast IA (Laurent Calixte) - Pourquoi : Interviews de praticiens francophones - Durée : 30-60 min - Où : Toutes plateformes

Chaînes YouTube francophones

Science4All (Lê Nguyễn Hoang) - Pourquoi : Mathématicien spécialiste de l'IA, pédagogue exceptionnel - Niveau : Accessible puis peut aller très loin - Incontournable : Sa série "L'IA, un danger pour l'humanité?"

Monsieur Phi (Thibaut Giraud) - Pourquoi : Philosophie de l'IA, questions éthiques - Niveau : Accessible, stimulant intellectuellement

Scilabus (Viviane Lalande) - Pourquoi : Vulgarisation scientifique de qualité - Niveau : Grand public

[H][Edu] Pour les curieux — Aller plus loin

Livres accessibles

“AI Superpowers” — Kai-Fu Lee (2018, trad. française disponible) - Perspective unique : ancien chef de Google Chine - Thème : Rivalité USA/Chine, impact sur l’emploi - ★ Lecture prioritaire

“Vous allez commettre une terrible erreur” — Olivier Sibony (2019) - Thème : Biais cognitifs et décision (pas que sur l’IA, mais essentiel) - Pourquoi : Comprendre nos propres limites avant de parler de celles de l’IA

“Atlas of AI” — Kate Crawford (2021, trad. française 2022) - Thème : Coûts cachés de l’IA (environnement, travail, données) - Pourquoi : Regard critique et documenté

“L’intelligence artificielle n’existe pas” — Luc Julia (2019) - Thème : Démystification par un pionnier (co-créateur de Siri) - Pourquoi : Perspective d’ingénieur, terre-à-terre

Cours en ligne gratuits

“Elements of AI” — Université d’Helsinki - Langue : Français disponible - Durée : 6 semaines, 2h/semaine - Niveau : Débutant absolu - Site : elementsofai.fr - ★ Le meilleur point d’entrée

“L’IA pour tous” — Andrew Ng (Coursera) - Langue : Sous-titres français - Durée : 4 semaines - Niveau : Débutant, non technique - Prix : Gratuit (audit) ou payant (certificat)

“Introduction to Artificial Intelligence” — edX / Columbia - Langue : Anglais (sous-titres) - Niveau : Intermédiaire - Pour : Ceux qui veulent comprendre les fondements

Documentaires

“AlphaGo” (2017) — Greg Kohs - Où : YouTube (gratuit) - Pourquoi : Moment historique, très bien raconté

“Coded Bias” (2020) — Shalini Kantayya - Où : Netflix - Pourquoi : Les biais discriminatoires des algorithmes

“The Social Dilemma” (2020) — Jeff Orlowski - Où : Netflix - Pourquoi : Pas que sur l’IA, mais sur l’économie de l’attention

[F][Ecole] Pour les enseignants — Intégrer l’IA en classe

Ressources institutionnelles

Éduscol — IA et éducation - Site : eduscol.education.fr - Contenus : Fiches pédagogiques, cadre officiel - Pourquoi : Conforme aux programmes français

CNIL — Kit pédagogique - Site : cnil.fr/education - Contenus : Ateliers clés en main, vidéos, quiz - Âge : Primaire au lycée

Class'Code — IAI - Site : classcode.fr - Contenus : Formation enseignants, ressources élèves - Pourquoi : Communauté active, partage de pratiques

Activités clés en main

1. Dessine-moi une IA - Niveau : Primaire-collège - Principe : Faire dessiner ce qu'est l'IA, discuter des représentations - Durée : 1h

2. Le jeu du Turing - Niveau : Collège-lycée - Principe : Un élève répond comme une IA, les autres devinent - Durée : 30 min

3. Biais dans les images - Niveau : Lycée - Principe : Générer des images avec DALL-E/Midjourney, analyser les stéréotypes - Durée : 2h

4. Fact-checking avec ChatGPT - Niveau : Lycée - Principe : Poser des questions factuelles, vérifier les réponses, analyser les erreurs - Durée : 1-2h

Formations enseignants

Canopé — Parcours IA - Format : Webinaires + ressources - Prix : Gratuit (réseau Canopé)

M@gistère — Intelligence artificielle - Format : Parcours en ligne - Public : Enseignants Éducation nationale

FUN MOOC — IA et éducation - Format : MOOC 6 semaines - Prix : Gratuit

[Pro] Pour les professionnels — IA au travail

Certifications reconnues

Google Cloud — Professional Machine Learning Engineer - Niveau : Avancé - Prix : ~200€ l'examen - Valeur : Reconnue internationalement

Microsoft — Azure AI Engineer Associate - Niveau : Intermédiaire - Prix : ~150€ l'examen - Valeur : Écosystème Microsoft/Copilot

AWS — Machine Learning Specialty - Niveau : Avancé - Prix : ~300€ l'examen - Valeur : Leader cloud

Formations continues

ESSEC — Executive Certificate in AI - Format : 6 mois, temps partiel - Prix : ~8000€ - Pour : Managers, dirigeants

HEC — AI for Business - Format : 2 jours - Prix : ~3000€ - Pour : Décideurs

Centrale Supélec — Executive Education IA - Format : Modulaire - Pour : Ingénieurs, chefs de projet

Communautés professionnelles

France Digitale — Écosystème startups - Site : francedigitale.org

Hub France IA — Association interprofessionnelle - Site : hub-franceia.fr

AI Paris — Événements et communauté - Site : aiparis.fr

[H][F][E] Pour les parents — Accompagner ses enfants

Guides pratiques

CNIL — **Les 10 conseils pour accompagner votre enfant** - Format : PDF téléchargeable - Site : cnil.fr/famille

Internet Sans Crainte - Site : internetsanscrainte.fr - Contenus : Guides par âge, vidéos, quiz

e-Enfance - Site : e-enfance.org - Services : Ligne d'écoute (0800 200 000), ressources

Applications de contrôle parental

Qustodio - Plateformes : iOS, Android, Windows, Mac - Fonctions : Temps d'écran, filtrage, rapports - Prix : Gratuit (basique) ou ~50€/an

Family Link (Google) - Plateformes : Android, Chromebook - Prix : Gratuit - Limite : Écosystème Google uniquement

Screen Time (Apple) - Plateformes : iOS, Mac - Prix : Intégré - Limite : Écosystème Apple uniquement

Livres pour enfants sur l'IA

"Ada & Zangemann" — Matthias Kirschner (dès 6 ans) - Thème : Logiciel libre, propriété du code - Disponible : Gratuit en ligne

"Hello Ruby" — Linda Liukas (6-10 ans) - Thème : Initiation à la programmation, pensée computationnelle - Collection : Plusieurs tomes

"Les décodeuses du numérique" — Léa Castor (collège) - Thème : Femmes dans la tech, déconstruction des stéréotypes - Format : BD

Conversations à avoir avec vos enfants

Questions pour ouvrir le dialogue : 1. "Est-ce que tu utilises ChatGPT pour tes devoirs?" 2. "Tu sais comment ça marche?" 3. "Tu lui racontes quoi?" 4. "Comment tu sais si ce qu'il dit est vrai?" 5. "Qu'est-ce qui te plaît / t'inquiète?"

Posture recommandée : - Curieux plutôt que flicage - Apprendre ensemble plutôt qu'interdire - Fixer des règles claires mais expliquées - Montrer l'exemple (et vos propres usages)

[Labo] Pour aller encore plus loin — Ressources techniques

Livres techniques accessibles

“Deep Learning” — Goodfellow, Bengio, Courville - Le livre de référence, disponible gratuitement en ligne - Site : deeplearningbook.org

“Hands-On Machine Learning” — Aurélien Géron - Pratique, avec code Python - Excellent pour apprendre en faisant

Cours universitaires ouverts

Stanford CS229 — Machine Learning (Andrew Ng) - Gratuit sur YouTube - Le classique des classiques

Stanford CS224N — NLP with Deep Learning - Gratuit sur YouTube - Pour comprendre les LLM en profondeur

MIT 6.S191 — Introduction to Deep Learning - Gratuit sur YouTube - Très bien produit

Papiers de recherche accessibles

“Attention Is All You Need” (2017) - L'article qui a lancé les Transformers - Lisible même sans PhD

“Language Models are Few-Shot Learners” (GPT-3, 2020) - Comprendre ce qui a tout changé

“Constitutional AI” (Anthropic, 2022) - L'approche Claude pour la sécurité

Où trouver les papiers : arxiv.org (gratuit, en anglais)

[Date] Événements à ne pas manquer

Conférences francophones

AI Paris — Juin, Paris - Le plus grand événement IA en France - Mix business et technique

NeurIPS Paris Meetup — Décembre - Retour sur la conférence NeurIPS - Très technique

Vivatech — Juin, Paris - Startup et innovation (pas que IA) - Grand public et pros

Événements européens

AI Summit London — Octobre - Orientation business - Speakers internationaux

WAICF (World AI Cannes Festival) — Février - Mix éthique et business - Cadre agréable

[!] À éviter — Sources non fiables

Red flags : - Promesses de “devenir expert en 24h” - Formations à prix exorbitant sans reconnaissance - Influenceurs sans expertise vérifiable - Articles sponsorisés déguisés en conseils - Prédiction sensationnalistes (fin du monde ou utopie)

Questions à se poser : - Qui est l’auteur? Quelles sont ses qualifications? - Qui finance ce contenu? - Y a-t-il des sources vérifiables? - Le ton est-il équilibré ou vendeur?

Ressources vérifiées en décembre 2025. Les liens peuvent évoluer. Signaler une ressource obsolète ou problématique : [contact à ajouter]

Annexe E

L'architecture Harmonia — Pour aller plus loin

Cette annexe présente une proposition d'architecture pour une IA gouvernable, développée dans le cadre de recherches académiques parallèles.

Pourquoi cette annexe ?

Tout au long de ce livre, nous avons vu les limites des IA actuelles : opacité, centralisation, perte de contrôle sur nos données. Cette annexe présente **Harmonia**, une architecture alternative qui tente de résoudre ces problèmes.

Ce n'est pas un produit commercial. C'est une proposition conceptuelle et technique, issue de travaux de recherche, qui illustre à quoi pourrait ressembler une IA conçue différemment — avec le citoyen au centre.

Du fichier au concept : quand l'informatique devient invisible

Pendant longtemps, l'informatique a été organisée autour de supports.

On parlait de fichiers .doc, .xls, .ppt. On rangeait des dossiers. On sauvegardait sur des disques durs. On transportait des CD, puis des clés USB.

Autrement dit : on gérait des **objets**.

Puis quelque chose de fondamental a changé.

Quand le support disparaît, le concept reste

Aujourd'hui, plus personne ne dit vraiment :

« Passe-moi le fichier .ppt »

On dit :

« Envoie-moi la présentation »

On ne dit plus :

« J'ai perdu mon .xls »

mais :

« J'ai perdu mon tableau »

Le format s'est effacé. Le concept a pris le dessus.

Comme pour la musique : - le vinyle a disparu - le CD a disparu - le fichier local a disparu

Mais la musique, elle, est partout.

Quand le support disparaît, ce qui compte vraiment apparaît.

Un OS de concepts, pas de fichiers

LifeOS — la vision qui anime Harmonia — part de cette idée simple :

Nous n'utilisons pas des fichiers. Nous vivons des concepts.

Un calendrier n'est pas une grille. C'est une relation au temps.

Une présentation n'est pas un format. C'est une manière de transmettre une idée.

Un tableau n'est pas un tableur. C'est une façon d'organiser le réel.

LifeOS ne gère donc pas des .ppt, des .xls ou des .doc. Il gère : - des **Calendars** - des **Presentations** - des **Tables** - des **Notes** - des **Conversations** - des **Décisions**

Les formats deviennent des détails techniques, interchangeable, secondaires.

Un concept, plusieurs expériences

Un point essentiel est souvent mal compris.

Un concept est unique, mais son expérience est multiple.

Il n'y a qu'un concept de *love*, mais il se vit différemment selon les cultures : - *love* - *amour* - ☐☐

Et même au sein d'une même culture, chaque individu vit ce concept à sa manière.

En informatique, on dirait : - le concept est une **classe** - l'expérience vécue est une **instance**

LifeOS applique cette idée au quotidien numérique : - un seul concept de **Calendar** - autant de calendriers vécus que de personnes

Libérer l'espace

Quand les bibliothèques de CD ont disparu, les salons ont changé.

L'espace libéré a permis : - d'ajouter autre chose - de réorganiser - ou simplement de laisser du vide

LifeOS poursuit le même mouvement.

C'est là qu'interviennent les LLM (GPT-4, Claude, etc.). Mais dans Harmonia, cette couche est **encadrée** par les deux autres.

Principe : Le LLM fait ce qu'il fait bien (comprendre le langage, générer du texte), mais il ne décide pas seul.

2. Trustia — La couche de confiance

Rôle : Gérer l'identité, les permissions, la traçabilité.

- Qui parle à l'IA ?
- Quels droits a cette personne ?
- Quelles actions sont autorisées ?
- Comment tracer ce qui s'est passé ?

Principe : Avant d'agir, l'IA vérifie qu'elle en a le droit.

3. Lumenia — La couche de gouvernance

Rôle : Définir les règles, les contraintes éthiques, les limites.

- Quelles sont les valeurs que l'IA doit respecter ?
- Quels contenus sont interdits ?
- Comment résoudre les conflits ?

Principe : Les règles sont explicites, pas cachées dans les poids du modèle.

Le document .life — Reprendre le contrôle

Concept

Le fichier .life est un document structuré qui contient : - Votre profil (préférences, contexte, besoins) - Vos règles personnelles (ce que l'IA peut/ne peut pas faire) - Votre historique (ce que vous avez partagé, et avec qui)

Vous êtes propriétaire de ce fichier. Il est stocké localement ou dans un espace que vous contrôlez.

Structure simplifiée

Exemple de document .life (simplifié)

```
identity:
  name: "Marie Dupont"
  preferred_language: "fr"
  timezone: "Europe/Paris"

preferences:
  communication_style: "direct"    # ou "gentle", "formal"
  verbosity: "concise"             # ou "detailed"

boundaries:
```

```

share_health_data: false
share_location: false
allow_marketing: false

values:
  priorities:
    - family
    - learning
    - privacy

permissions:
  openai_chatgpt:
    read_calendar: false
    send_emails: false
    access_history: "session_only" # pas de persistance

audit_log:
  enabled: true
  retention_days: 30

```

Principes clés

- 1. Portabilité** Votre .life fonctionne avec n'importe quelle IA compatible. Pas de vendor lock-in.
 - 2. Lisibilité** Le format est compréhensible par un humain, pas juste par une machine.
 - 3. Granularité** Vous décidez permission par permission, service par service.
 - 4. Révocabilité** Vous pouvez retirer une permission à tout moment.
-

Le Gouverneur — Des règles vérifiables

Problème des LLM actuels

Les “garde-fous” des LLM actuels sont : - Entraînés dans le modèle (RLHF) → impossible à vérifier - Ajoutés en post-traitement → contournables - Définis par l'entreprise → vous n'avez pas votre mot à dire

Solution Harmonia

Un **Gouverneur** externe au modèle qui : - Intercepte les entrées et sorties - Applique des règles formelles vérifiables - Peut être audité par un tiers - Peut être personnalisé par l'utilisateur

Exemple de règle

```

RÈGLE: protection_mineurs
SI contexte.user.age < 18
ALORS:

```



```
- BLOQUER contenu.violence == "explicite"  
- BLOQUER contenu.adult == true  
- ALERTER si conversation.topic in ["automutilation", "suicide"]  
AUDIT: obligatoire
```

Cette règle est : - **Explicite** : on sait exactement ce qu'elle fait - **Vérifiable** : on peut tester qu'elle fonctionne - **Auditable** : on peut prouver qu'elle a été appliquée

L'Alter Ego — Une IA qui vous connaît (avec permission)

Concept

L'Alter Ego est une couche de personnalisation qui : - Apprend vos préférences au fil du temps - Adapte les réponses à votre contexte - Mémorise ce que vous lui avez partagé

Différence avec le modèle actuel : Les données restent les vôtres. Elles ne servent pas à entraîner le modèle pour d'autres.

Fonctionnement

```
[Votre question]  
  ↓  
[Alter Ego] → "Marie préfère des réponses concises,  
              elle travaille dans l'éducation,  
              elle est sensible aux questions de vie privée"  
  ↓  
[Prompt enrichi envoyé au LLM]  
  ↓  
[Réponse adaptée à vous]
```

Garanties

- **Consentement explicite** : L'Alter Ego ne collecte rien sans votre accord
 - **Transparence** : Vous pouvez voir tout ce qu'il "sait" sur vous
 - **Oubli** : Vous pouvez effacer tout ou partie à tout moment
 - **Localité** : Les données peuvent rester sur votre appareil
-

Cas d'usage illustratifs

Cas 1 : IA d'entreprise gouvernée

Une entreprise déploie un assistant IA pour ses employés.

Sans Harmonia : - L'entreprise fait confiance à OpenAI - Pas de contrôle sur ce qui est dit aux employés - Données sensibles potentiellement exposées

Avec Harmonia : - Gouverneur d'entreprise avec règles RH et légales - Données restent sur infrastructure de l'entreprise - Audit trail complet pour conformité

Cas 2 : IA éducative pour enfants

Une école utilise une IA pour aider les élèves.

Sans Harmonia : - Dépendance aux garde-fous de l'éditeur - Pas de personnalisation au contexte scolaire - Données des mineurs sur serveurs tiers

Avec Harmonia : - Règles adaptées à l'âge et au programme - Profil .life contrôlé par les parents - Données sur infrastructure scolaire

Cas 3 : IA de santé

Un patient utilise une IA pour comprendre ses symptômes.

Sans Harmonia : - Données de santé chez un géant tech - Risque de conseils médicaux inappropriés - Pas de lien avec le dossier médical

Avec Harmonia : - Données de santé dans un espace certifié (HDS) - Règles strictes sur les conseils médicaux - Intégration contrôlée avec le parcours de soin

État d'avancement

Ce qui existe (décembre 2025) : - Spécification conceptuelle complète - Prototype de recherche (Harmonia Kernel) - Publications académiques en cours

Ce qui reste à faire : - Implémentation industrielle - Standardisation des formats - Écosystème d'outils compatibles - Adoption par des acteurs

Où suivre les développements : - GitHub : [lien à ajouter] - Papiers : [référence académique]

Pour aller plus loin

Documentation technique

La spécification complète d'Harmonia fait l'objet de publications académiques distinctes :

"From Rational Agents to Governed Cognitive Systems" - Cadre théorique complet - Formalisation mathématique - Validation expérimentale

Participer

Harmonia est un projet ouvert. Vous pouvez : - Lire et commenter les spécifications - Contribuer au code (open source) - Proposer des cas d'usage - Critiquer et améliorer

L'objectif n'est pas de vendre un produit, mais de montrer qu'une autre IA est possible.

Conclusion

Harmonia n'est pas une utopie. C'est une démonstration de faisabilité.

Elle montre que les problèmes actuels de l'IA — opacité, centralisation, perte de contrôle — ne sont pas des fatalités techniques. Ce sont des choix de conception. Et d'autres choix sont possibles.

Le jour où suffisamment de citoyens exigeront une IA gouvernable, les outils existeront pour la construire.

Ce jour dépend de vous.

Cette annexe présente une vue simplifiée d'Harmonia. La documentation technique complète est disponible séparément.

Annexe F

Bibliographie

Références citées dans ce livre et lectures complémentaires recommandées.

Livres — Accessibles au grand public

Comprendre l'IA

Ajay Agrawal, Joshua Gans, Avi Goldfarb *Prediction Machines : The Simple Economics of Artificial Intelligence* Harvard Business Review Press, 2018 → L'IA comme outil de prédiction bon marché. Économique et clair.

Kai-Fu Lee *AI Superpowers : China, Silicon Valley, and the New World Order* Houghton Mifflin Harcourt, 2018 → La course USA/Chine vue de l'intérieur. Incontournable.

Luc Julia *L'intelligence artificielle n'existe pas* First, 2019 → Démystification par le co-créateur de Siri. Accessible.

Kate Crawford *Atlas of AI : Power, Politics, and the Planetary Costs of Artificial Intelligence* Yale University Press, 2021 → Les coûts cachés de l'IA. Critique et documenté.

Meredith Broussard *Artificial Unintelligence : How Computers Misunderstand the World* MIT Press, 2018 → Les limites fondamentales des systèmes informatiques.

Éthique et société

Cathy O'Neil *Weapons of Math Destruction : How Big Data Increases Inequality* Crown, 2016 → Les algorithmes comme outils de discrimination. Précurseur.

Virginia Eubanks *Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor* St. Martin's Press, 2018 → L'automatisation des décisions sociales. Percutant.

Shoshana Zuboff *The Age of Surveillance Capitalism* PublicAffairs, 2019 → L'économie de la surveillance. Dense mais fondamental.

Safiya Umoja Noble *Algorithms of Oppression : How Search Engines Reinforce Racism* NYU Press, 2018 → Biais raciaux dans les moteurs de recherche.

Histoire et philosophie

Nick Bostrom *Superintelligence : Paths, Dangers, Strategies* Oxford University Press, 2014 → Les risques existentiels de l'IA. Controversé mais influent.

Stuart Russell *Human Compatible : Artificial Intelligence and the Problem of Control* Viking, 2019 → Comment créer une IA alignée avec les humains. Par un pionnier.

Brian Christian *The Alignment Problem : Machine Learning and Human Values* W.W. Norton, 2020 → Le problème de l'alignement expliqué clairement.

Livres — Techniques (pour aller plus loin)

Ian Goodfellow, Yoshua Bengio, Aaron Courville *Deep Learning* MIT Press, 2016 → La bible du deep learning. Gratuit en ligne : deeplearningbook.org

Aurélien Géron *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* O'Reilly, 3e édition 2022 → Apprendre en pratiquant. Excellent.

Christopher Bishop *Pattern Recognition and Machine Learning* Springer, 2006 → Les fondements mathématiques. Classique.

Jurafsky & Martin *Speech and Language Processing* Draft 3e édition (en ligne gratuitement) → Le traitement du langage avant et après les transformers.

Articles académiques fondateurs

L'article qui a tout changé

Vaswani et al. "Attention Is All You Need" NeurIPS 2017 → L'invention des Transformers. À l'origine de GPT, BERT, et tous les LLM.

Les grands modèles de langage

Brown et al. (OpenAI) "Language Models are Few-Shot Learners" (GPT-3) NeurIPS 2020 → La démonstration que la taille compte.

Ouyang et al. (OpenAI) "Training language models to follow instructions with human feedback" (InstructGPT) NeurIPS 2022 → RLHF : comment rendre les LLM utiles.

Touvron et al. (Meta) "LLaMA : Open and Efficient Foundation Language Models" arXiv 2023 → L'open source rattrape les modèles fermés.

Sécurité et alignement

Bai et al. (Anthropic) "Constitutional AI : Harmlessness from AI Feedback" arXiv 2022 → L'approche Claude pour la sécurité.

Perez et al. *“Red Teaming Language Models with Language Models”* arXiv 2022 → Tester les failles avec des IA.

Critiques et limites

Bender, Gebru et al. *“On the Dangers of Stochastic Parrots : Can Language Models Be Too Big ?”* FAccT 2021 → Le “perroquet stochastique”. Article controversé mais important.

Marcus, Davis *“GPT-3, Bloviator : OpenAI’s language generator has no idea what it’s talking about”* MIT Technology Review, 2020 → Critique précoce des limites des LLM.

Rapports institutionnels

Union européenne

Commission européenne *AI Act — Règlement (UE) 2024/1689* Journal officiel de l’UE, 2024 → Le texte officiel.

AI HLEG *Ethics Guidelines for Trustworthy AI* Commission européenne, 2019 → Les principes éthiques qui ont inspiré l’AI Act.

France

CNIL *Intelligence artificielle : plan d’action* CNIL, 2024 → Position du régulateur français.

Conseil national du numérique *Rapport sur l’IA générative* CNNum, 2024 → Recommandations pour la France.

Villani et al. *Donner un sens à l’intelligence artificielle : pour une stratégie nationale et européenne* Mission Villani, 2018 → Le rapport fondateur de la stratégie française.

International

OCDE *AI Principles* OCDE, 2019, mis à jour 2024 → Principes directeurs internationaux.

UNESCO *Recommandation sur l’éthique de l’intelligence artificielle* UNESCO, 2021 → Premier instrument normatif mondial.

Médias et newsletters

Newsletters recommandées

Import AI (Jack Clark) → Hebdomadaire, en anglais. Le plus complet sur la recherche.

The Batch (Andrew Ng / DeepLearning.AI) → Hebdomadaire, en anglais. Pédagogique.

La Lettre de l'IA (Flint) → Hebdomadaire, en français. Équilibré.

Médias spécialisés

MIT Technology Review — technologyreview.com → Référence pour l'actualité tech critique.

Wired — wired.com → Tech et société.

The Verge — theverge.com → Actualité tech grand public.

Siècle Digital — siecledigital.fr → Actualité tech francophone.

Podcasts

En anglais

Lex Fridman Podcast → Interviews longues avec chercheurs IA. Excellentes.

The TWIML AI Podcast (Sam Charrington) → Focus recherche et industrie.

80,000 Hours Podcast → Éthique de l'IA et risques existentiels.

En français

Underscore_ (France Culture) → Réflexion de fond sur le numérique.

Le Podcast IA (Laurent Calixte) → Interviews d'acteurs francophones.

Vlan ! (Grégory Pouy) → Marketing et société, épisodes réguliers sur l'IA.

Sites et ressources en ligne

Tutoriels et cours

fast.ai — Cours pratiques de deep learning gratuits **Coursera** — Spécialisation Deep Learning (Andrew Ng) **Hugging Face** — Documentation et modèles open source **Google AI** — Tutoriels et papiers **OpenAI Cookbook** — Exemples d'utilisation des API

Actualité recherche

arXiv (arxiv.org) — Prépublications scientifiques **Papers With Code** — Papiers + implémentations **Semantic Scholar** — Moteur de recherche académique

Données et benchmarks

Kaggle — Datasets et compétitions **Hugging Face Datasets** — Datasets pour NLP
HELM (Stanford) — Benchmarks LLM

Références citées par chapitre

Chapitre 1

- Turing, A. (1950). “Computing Machinery and Intelligence”
- McCarthy et al. (1955). Proposition Dartmouth Conference

Chapitre 2

- Vaswani et al. (2017). “Attention Is All You Need”
- Radford et al. (2018). “Improving Language Understanding by GPT”

Chapitre 5-6

- Lee, K.F. (2018). AI Superpowers
- Données économiques : Statista, CB Insights, PitchBook

Chapitre 9

- Texte RGPD : EUR-Lex
- Texte AI Act : EUR-Lex
- Guides CNIL : cnil.fr

Chapitre 19

- O’Neil, C. (2016). Weapons of Math Destruction
- Zuboff, S. (2019). Surveillance Capitalism

Chapitre 22

- Goodfellow et al. (2016). Deep Learning
- Jurafsky & Martin. Speech and Language Processing

Chapitre 23 / Annexe E

- Documentation Harmonia (publications en cours)
-

Note sur les sources

Les liens web ont été vérifiés en décembre 2025. Internet étant volatile, certains peuvent devenir obsolètes. En cas de lien mort : - Cherchez le titre exact sur un

moteur de recherche - Consultez archive.org (Wayback Machine) - Pour les papiers académiques : arxiv.org ou Google Scholar

Bibliographie v1.0 — Décembre 2025