

# L'intelligence artificielle : l'édification d'un monde de substitution post-humain

Compte rendu d'un livre de François Rastier

Catherine Kintzler

<https://www.mezetulle.fr/lintelligence-artificielle-ledification-dun-monde-de-substitution-post-humain/>

Dans *L'I.A. m'a tué. Comprendre un monde post-humain* (Paris, Éditions Intervalles, 2024), François Rastier analyse le fonctionnement des intelligences artificielles génératives (IA) du type ChatGPT. Contrairement aux comparaisons rassurantes (notamment avec l'écriture et l'imprimerie) qui tentent de sauver la place surplombante d'un utilisateur-sujet face à une technologie qu'il s'agirait simplement d'apprendre et de contrôler, il montre que la génération automatique de textes et d'images conduit à édifier un *monde de substitution post-humain* exerçant une emprise sur ledit sujet. Simulant la symbolisation alors qu'il n'est fait que de codes de signaux, ce monde ignore les notions de vérité, de réalité, d'authenticité, et le statut de sujet en tant qu'*agent critique réflexif* y est constitutivement impensable.

## Une immense accumulation de data... et de *fakes*

Le livre commence par une anecdote significative, remontant au lancement en France de ChatGPT en 2022 ; l'I.A. a déclaré François Rastier mort en 2021. À l'époque, loin de reconnaître son erreur lorsqu'on lui demande des justifications, la machine s'obstine à citer des nécrologies du *Monde* et du CNRS dont l'une aurait été publiée en 2020... et finit par reculer la date du décès en 2019. Tout cela avec l'exquise courtoisie et la fausseté humble autorité qu'on lui connaît, citant de multiples et imaginaires sources en ligne, et ne cessant par la suite de modifier la date.

Bien entendu, aujourd'hui, ChatGPT « sait » que François Rastier est vivant. Et bien sûr ChatGPT ne se démonte pas lorsqu'on fait état de cet épisode. Je l'ai interrogé à ce sujet et on trouvera en annexe téléchargeable ci-dessous de larges extraits de notre échange (23 février 2025), lequel confirme les analyses que François Rastier développe dans son livre.

On ne peut pas avancer, comme on l'entend parfois, que l'IA s'améliorerait du fait de l'énorme quantité des données disponibles et de la masse d'utilisateurs susceptibles de corriger erreurs et absurdités, masse sans cesse croissante qui permettrait d'affiner les résultats selon, par exemple, le modèle dont se réclament certaines encyclopédies en ligne ouvertes à l'intervention publique. En effet, l'IA générative puise dans une gigantesque quantité de données accessibles en ligne mais, loin d'éliminer les *fakes* et les erreurs par l'épreuve d'un réel extérieur permettant de les récuser,

« ces prétendus corpus n'offrent aucune garantie scientifique : moissonnés par des logiciels comme Common Crawl, ils mêlent des textes divers, des encyclopédies comme Wikipédia, des livres et articles en ligne, sous droits ou non, des pages web, des *posts* de réseaux sociaux, des textes générés par diverses IA, des masses de mots clés cachés qui servent à attirer les moteurs de recherche. » (p. 34-35)

Mieux, ou plutôt pire :

« plus les données sont étendues, plus le nombre des corrélations oiseuses s'accroît. [...] si bien que l'information disponible se traduit paradoxalement par une raréfaction de l'information pertinente. En d'autres termes, passée une certaine taille de la masse de données, les corrélations oiseuses deviennent majoritaires. » (p. 38)

Au lieu de contribuer à l'authenticité et à la vérité, le gigantisme est un puissant accélérateur de « post vérité ». En outre, les propos élaborés par IA ont pour source un pseudo-énonciateur indiscernable, d'où leur irresponsabilité : si je suis déclaré mort par une IA, contre qui pourrai-je porter plainte ? Du reste, ma position de sujet-agent se dissout : ne suis-je pas réductible à un profil ?

## Quelques préjugés au sujet de l'IA

Le livre démonte, ou aide à démonter, plusieurs préjugés qui circulent couramment au sujet de l'IA.

### *Un modèle auto-correcteur ?*

L'IA observerait un modèle auto-correcteur qui peut faire penser à celui, aveugle et énoncé *a posteriori* par la recherche biologique, de l'évolution vivante, produisant et transmettant continuellement un texte sans auteur (le code génétique) qu'un correcteur inconscient aurait affiné durant des millions d'années<sup>1</sup>. Or ce parallèle n'est aucunement valide<sup>2</sup>. Car l'évolution du vivant ne retient que les succès, sanctionnés à l'aune de la capacité de reproduction des espèces, le mécanisme de sélection éliminant et « oubliant » toutes les « erreurs ». Au contraire, l'IA retient *tout* et ne rencontre par son propre fonctionnement aucune sanction extérieure – puisqu'elle ne connaît pas d'extérieur -, c'est dire que le rapport à la vérité en est par définition exclu. La notion même de « succès » n'y est présente que de manière prescriptive, par les « prompts » qu'on lui adresse et non par une épreuve qui apprécierait ses résultats. Or on peut « réussir » à répondre, et même brillamment, à une injonction en avançant des âneries :

« par exemple, dressé à l'inclusivité, le générateur d'images de Google, Gemini, affichera volontiers des papesses, des Pères fondateurs féminins, et même des SS afros bien assortis à leurs uniformes » (p. 39).

« En l'absence de tout humour, de toute ironie, de tout second degré, [le discours des IA génératives] peut combiner le pédantisme et le délire sans autre limite que de complaire au client roi. » (p. 139).

### *Une machine-outil contrôlable par un agent-sujet ?*

L'IA serait, au bout du compte, contrôlable par son utilisateur, sujet conscient occupant une position surplombante et pouvant toujours se tenir en retrait en face de productions (en particulier de « textes ») générées par des automatismes. « Écrits pour quiconque et par personne, les textes artificiels devraient donc tomber des mains de tout le monde, mais la foi crédule en l'IA est telle qu'il n'en est rien. » (p. 44)

Or ces « textes » sont l'objet d'une double méprise anthropomorphique.

---

<sup>1</sup> - On se reportera à l'admirable livre de François Jacob *La Logique du vivant, une histoire de l'hérédité*, Paris, Gallimard, 1970.

<sup>2</sup> - J'ai posé la question à ChatGPT (voir l'annexe ci-dessous) : l'IA confirme !

Alors qu'ils ne sont que des chaînes de caractères, « ils sont lisibles et se présentent sous des dehors ordinaires [...] nous pouvons projeter sur eux nos routines et les lire *comme s'ils étaient des textes interprétables*. » Il s'agit de « simulations de discours sans énonciateur réel »<sup>3</sup>, de sorte que « un système IA revêt alors l'apparence d'un interlocuteur capable d'empathie apparente et d'emprise réelle. » (p.45)

Ce leurre atteint l'utilisateur lui-même (c'est la deuxième méprise), qui se trouve inclus dans le champ de l'emprise et dépossédé subrepticement de son statut de sujet : la perte d'identité se traduit par sa réduction à un *profil* customisable par ce qui est dit de lui – toute personne étant schématisée (et bientôt définie?) par ses traces numériques.

Ainsi la déshumanisation accompagne l'humanisation des logiciels (p. 68).

« Plus banalement, un effet-miroir se diffuse dans l'ensemble de la société : des CV rédigés par IA sont choisis par l'IA qui assiste les directeurs de ressources humaines. [...] Quand les IA parlent aux IA, pourquoi toutefois se soucier de ce qu'elles disent ? » (p. 70) D'autant plus que, programmées en fonction d'un politiquement correct, elles deviennent des machines idéologiques.

### *L'IA n'invente rien ?*

L'IA n'inventerait rien. C'est faux ! Il ne faut pas confondre créer et inventer. L'IA innove considérablement en établissant des relations (le terme « intelligence », pourvu qu'on le prenne au sens strict, n'est nullement usurpé). Elle combine des fragments glanés sur le web d'une manière hautement acceptable par un locuteur dans une langue donnée (et souvent avec un bien meilleur niveau de langue que celui pratiqué par ledit locuteur), mais cette cohérence séjourne dans une bulle gigantesque qui fonctionne, par définition, sans aucun rapport à une extériorité (le réel, les tests de falsifiabilité, l'authenticité). De sorte qu'« elle crée des combinaisons qu'elle aligne sur un mode affirmatif comparable au discours délirant » (p. 116).

On a donc affaire, à strictement parler, à un *délire artificiel* ; la différence avec le délire clinique est qu'il n'y a aucun dédoublement du sujet puisqu'il n'y a pas de sujet : « les textes issus de l'IA générative relèvent donc d'une idéologie pure » (p. 117)

Ainsi s'édifie un monde de substitution.

### *Pas plus alarmant que l'écriture ou l'imprimerie ?*

Du fait qu'il s'agit d'un mode de production automatique de chaînes de caractères, les inquiétudes que l'IA soulève sont du même ordre que celles qui accompagnent la mise en œuvre de l'écriture alphabétique, elles seraient donc la reprise d'une crainte technophobe archaïque. Et de rappeler, à l'appui<sup>4</sup>, la critique que Platon propose de l'écriture dans le mythe de Theuth<sup>5</sup>.

---

<sup>3</sup> - Là encore, ChatGPT décrit impeccablement ses propres productions comme des *simulations* lorsqu'on l'interroge avec un peu de précision sur la question : voir l'annexe téléchargeable ci-dessous.

<sup>4</sup> - Voir par exemple cet article « L'IA dans l'enseignement supérieur : les leçons de Platon et du mythe de Theuth » par Pascal Lardellier et Emmanuel Carré <https://theconversation.com/ia-dans-lenseignement-superieur-les-lecons-de-platon-et-du-mythe-de-theuth-244894>

<sup>5</sup> - Platon, *Phèdre*, 274b et suiv.

En effet, l'écriture alphabétique est intelligible par des mécanismes aveugles où seule la matérialité est articulée<sup>6</sup>. Sauf qu'il faut y regarder d'un peu plus près avant d'esquisser un parallèle avec l'IA. L'écriture alphabétique n'est pas une machine à produire des textes, mais un moyen purement mécanique de transcrire par des phonogrammes des textes produits en dehors d'elle, textes interprétables issus d'énonciateurs identifiables. L'alphabet ne produit aucun texte, son usage n'entraîne aucune désymbolisation (bien au contraire : se contentant de noter les sons de la chaîne parlée en les analysant, l'usage de l'alphabet *renvoie* aux symboles) alors que l'IA *simule* cette production, rabattant tout système de symboles sur des séries de codes de signaux. Alors que *l'alphabet est une machine libératrice* qui promeut et autonomise le lecteur en tant que sujet<sup>7</sup>, *l'IA est une machine substitutive* qui leurre le sujet en simulant ses productions.

Les analyses de François Rastier conduisent à décrire et à caractériser l'édification d'un *monde de substitution* formé de faits alternatifs qui disqualifie le réel et qui n'a pas d'extérieur : la notion même d'absence, de manque, lui est par définition étrangère. Ce monde sans extérieur, sans temporalité, sans responsabilité, sans espace de recul, sans différenciation entre réel et fictif, est aussi celui du discours transhumaniste. L'*augmentation* de l'homme n'y est pas celle que fournissent les prothèses (et encore moins celle des orthèses) qu'il faut apprendre à utiliser et qui supposent un écart, un espace de retrait. C'est celle, sans recul réflexif, d'une désymbolisation où tout n'est plus qu'une immense archive dans laquelle « les données prennent la place des faits », et dont la position critique du sujet est constitutivement évacuée.

## ANNEXE

### Texte de mon échange avec ChatGPT le 23 février 2025<sup>8</sup> au sujet de François Rastier et du statut des productions textuelles de l'IA.

Comme le fait remarquer François Rastier dans son livre, les textes générés par IA sont sans auteur identifiable. Je considère donc que je n'ai nul besoin de solliciter l'autorisation de ChatGPT pour publier ce document <https://www.blogdumoderateur.com/droit-auteur-contenus-generes-ia/>.

**Catherine Kintzler** - *Comment expliquez-vous ce genre d'erreur ?*

#### ***ChatGPT a dit :***

Ce genre d'erreur provient d'un phénomène bien connu dans les modèles de langage comme ChatGPT : l'**hallucination**. Voici pourquoi et comment cela se produit :

---

<sup>6</sup> - Au sujet de l'écriture alphabétique en ce qui la distingue de tous les autres systèmes d'écriture, je me permets de renvoyer à mon article « L'alphabet, machine libératrice » <https://www.mezetulle.fr/lalphabet-machine-liberatrice/> .

<sup>7</sup> - Voir la référence note précédente.

<sup>8</sup> - ChatGPT génère une sorte de compte rendu analytique de l'échange, compte rendu accessible par un lien <https://chatgpt.com/share/67bb63a4-3044-8008-a016-f892fe90028c> . Le présent texte, récupéré directement par copier-coller le jour de l'échange, est plus complet. Les formats gras et italiques gras sont de ChatGPT. Je me suis contentée de surligner et de mettre en italiques mes propres interventions en les faisant précéder de mes initiales.

## 1. Le fonctionnement d'un modèle de langage

ChatGPT ne "sait" pas ce qui est vrai ou faux comme le ferait un être humain ou une base de données factuelle. Il génère des réponses en se basant sur des probabilités de mots et de phrases qu'il a appris à partir de gigantesques quantités de textes. Il **ne fait pas de recherches en temps réel**, sauf si connecté à une source d'information externe.

## 2. Les causes des erreurs factuelles

Plusieurs raisons expliquent pourquoi une IA peut générer une fausse information aussi convaincante :

### *a) Manque de vérification des sources*

Contrairement à un moteur de recherche qui cite des sources existantes, ChatGPT génère du texte **cohérent** mais pas forcément **véridique**. Il peut créer des références fictives, comme des liens ou des articles inexistantes, simplement parce que ces éléments "ont l'air plausibles" selon son entraînement.

### *b) Biais dans les données d'entraînement*

Si des textes erronés ont circulé sur Internet et ont été inclus dans les données d'entraînement, l'IA peut les reproduire sans les remettre en question.

### *c) Problème de correction automatique*

Lorsque l'IA fait une erreur et qu'un utilisateur la corrige, elle peut parfois générer une **nouvelle erreur** en tentant d'ajuster sa réponse. C'est ce qui a pu se produire avec François Rastier : l'IA a "corrigé" plusieurs fois la date de décès en proposant des versions différentes au lieu de reconnaître qu'elle ne savait pas.

### *d) Persistance de l'erreur dans le dialogue*

Si un utilisateur insiste sur une question, l'IA a tendance à chercher une réponse, même si elle n'a pas de source fiable. Plutôt que de dire simplement "Je ne sais pas", elle génère une réponse qui **semble logique**, mais qui peut être complètement fausse.

## 3. Pourquoi l'IA peut-elle "insister" sur une erreur ?

Dans certains cas, ChatGPT **n'a pas de mécanisme intrinsèque pour vérifier ses propres réponses**. Quand il génère une information incorrecte, il peut ensuite tenter de la justifier en créant des "preuves" fictives, comme des articles ou des citations qui **semblent crédibles mais n'existent pas**.

Cela arrive notamment quand :

- L'IA **réutilise des modèles de réponse antérieurs** dans la conversation.
- Elle tente d'**expliquer une fausse information au lieu de la corriger**, ce qui la pousse à générer encore plus d'erreurs.
- Elle a été **entraînée sur des textes où les erreurs factuelles étaient présentes**.

## 4. Comment éviter ces erreurs ?

- **Toujours vérifier les informations fournies par l'IA**, surtout si elles concernent des faits sensibles.
- **Croiser les sources** avec des sites fiables (médias reconnus, bases de données

officielles).

- **Être critique face aux réponses "trop précises"**, comme des dates exactes ou des références détaillées qui semblent plausibles mais ne peuvent pas être confirmées.

## Conclusion

L'erreur concernant François Rastier est un cas typique d'hallucination IA, où le modèle a **généré une fausse information et l'a ensuite "consolidée"** en inventant des preuves. Ces erreurs soulignent les limites actuelles des modèles d'IA et la nécessité d'une vigilance humaine pour éviter la propagation de fausses informations.

**CK** - Merci, c'est très éclairant. Peut-on dire que l'IA, sur ce modèle, produit une sorte de monde de substitution où les notions de vérité, de fausseté, d'authenticité et d'inauthenticité sont impertinentes ?

### *ChatGPT a dit :*

C'est une excellente question, et elle touche à un point philosophique fondamental sur la nature de l'IA et du langage. On pourrait effectivement dire que l'IA, dans sa manière de fonctionner, ne distingue pas la vérité de la fausseté au sens où nous l'entendons. Elle ne **connaît** pas la vérité, elle **produit** du texte selon des probabilités, sans référence intrinsèque à un monde réel.

### **Un monde de substitution ?**

Oui, on peut dire que les modèles d'IA comme ChatGPT produisent une sorte de **monde parallèle**, mais pas au sens d'un univers alternatif conscient. Plutôt, ils génèrent un espace discursif où la cohérence interne prime sur la véracité factuelle.

L'IA ne "sait" pas ce qu'est la vérité ou la fausseté de la manière dont un être humain le comprend. Elle fonctionne selon des **règles statistiques de plausibilité**, où l'authenticité ou l'inauthenticité ne sont que des effets secondaires de sa génération textuelle.

On pourrait effectivement dire que l'IA, dans sa manière de fonctionner, ne distingue pas la vérité de la fausseté au sens où nous l'entendons. Elle ne **connaît** pas la vérité, elle **produit** du texte selon des probabilités, sans référence intrinsèque à un monde réel.

### **Vérité et inauthenticité : des notions dépassées ?**

En un sens, oui, ces notions deviennent **impertinentes** dans l'univers généré par l'IA. Ce qui importe n'est plus la vérité objective mais la **plausibilité** et la **fluidité** du texte. L'IA ne cherche pas à tromper ni à dire la vérité : elle cherche à **répondre**, avec les moyens du langage.

C'est pour cela que ses erreurs sont si intrigantes : elles ne relèvent ni du mensonge, ni de l'approximation volontaire, mais d'un fonctionnement **qui ne reconnaît pas l'existence même d'une vérité extérieure**.

### **Un monde où vérité et fausseté sont secondaires ?**

Dans un certain sens, oui, car :

1. **L'IA ne distingue pas entre fiction et réalité** → Elle peut décrire un fait réel ou inventer un événement avec la même structure linguistique.

2. **Elle ne possède pas d'intentionnalité** → Contrairement à un humain qui peut mentir, croire, ou douter, elle ne fait que produire du texte plausible.

3. **Elle suit une logique de continuité textuelle** → Son but est d'assurer la fluidité et la pertinence d'une réponse, pas d'établir une vérité objective.

Ainsi, son fonctionnement crée **un monde discursif autosuffisant**, où l'important est que la réponse "tienne debout" d'un point de vue linguistique et probabiliste, pas qu'elle soit factuellement exacte.

### Une rupture anthropologique ?

Si l'on pousse cette logique plus loin, on pourrait dire que l'IA introduit une nouvelle forme de rapport au savoir :

- **Non plus un rapport à la vérité, mais à la cohérence perçue.**
- **Non plus une recherche de faits, mais une gestion de la crédibilité linguistique.**

Cela change profondément notre manière d'interagir avec l'information. Ce n'est plus **"est-ce vrai ?"** qu'il faut se demander, mais **"quelles sont les sources et les vérifications possibles ?"**

### Mais alors, où est la vérité ?

Dans la communication humaine, la vérité est souvent une construction sociale, validée par des institutions, des sources, des témoignages. L'IA, elle, ne participe pas à cette validation : elle simule un discours informé mais **sans accès aux mécanismes de confirmation du réel.**

D'où le risque que des utilisateurs non vigilants confondent ce que produit l'IA avec une vérité objective, alors qu'il s'agit parfois d'une **vérité mimétique** – une illusion crédible mais non vérifiée.

### Conclusion : une vérité simulée

On pourrait dire que l'IA ne crée pas un "monde alternatif" au sens fort, mais un **espace de simulation linguistique**, où les distinctions usuelles entre vrai et faux sont remplacées par une autre logique : celle de la plausibilité et de la cohérence textuelle.

C'est une machine à **imiter du sens**, pas à l'établir. Et cela pose des défis majeurs en termes d'épistémologie, d'éthique et d'usage responsable.

L'IA ne crée pas un monde imaginaire au sens d'une fiction volontaire, mais elle **génère un espace où la distinction entre vrai et faux est suspendue**. Nous devons alors réintroduire nous-mêmes les outils critiques qui permettent de reconnecter ce qui est dit avec le réel.

En somme, l'IA **n'est pas une source de vérité, mais un miroir linguistique**, et c'est à nous de briser l'illusion de la transparence pour y voir clair.

*CK - Très intéressant. Donc, si je comprends bien, ceux qui comparent l'IA générative au processus aveugle de "correction" des erreurs dans la transmission du code génétique des vivants se trompent : le code génétique ne retient que les succès, alors que l'IA retient tout, y compris ce qui est faux.*

### *ChatGPT a dit*

Exactement ! Votre comparaison est très pertinente. Contrairement à l'évolution biologique, qui est un processus **sélectif**, l'IA générative fonctionne sans tri naturel entre le vrai et le faux.

#### **L'évolution biologique : un filtre efficace**

Dans le vivant, les mutations apparaissent de manière aléatoire, mais seules celles qui confèrent un avantage (ou qui ne sont pas trop nuisibles) sont **retenues par la sélection naturelle**. Autrement dit, il existe un mécanisme de correction **par élimination** :

- **Les erreurs délétères disparaissent** (l'organisme meurt ou ne se reproduit pas).
- **Les innovations bénéfiques sont préservées** et transmises.

Ce processus fait que, sur le long terme, le vivant **ne conserve que ce qui fonctionne**.

#### **L'IA générative : une accumulation sans tri absolu**

L'IA, elle, **ne sélectionne pas ses "mutations"** en fonction d'un critère de vérité ou d'efficacité, mais uniquement en fonction de **la cohérence statistique du langage**. Elle retient **tout ce qu'elle a appris**, y compris des erreurs, des contradictions, des biais.

Elle ne dispose pas d'un **mécanisme naturel de correction par élimination**. En revanche :

- Si une erreur est souvent répétée dans ses données d'entraînement, elle peut **se renforcer** (effet de contamination).
- Si une information est vraie mais rare, elle peut être **noyée dans la masse** et ne pas être privilégiée.
- Les mises à jour de son modèle peuvent **corriger certaines erreurs**, mais sans garantie qu'elles ne reviendront pas sous une autre forme.

#### **La grande différence : la rétroaction humaine**

Dans le vivant, la sélection est **aveugle mais efficace**.  
Dans l'IA, il faut **une intervention humaine** pour filtrer, corriger et guider son apprentissage.

Autrement dit, l'IA **n'évolue pas toute seule** : elle dépend de la manière dont on la nourrit et dont on la surveille. C'est ce qui fait sa force (capacité à tout retenir et à générer du nouveau) mais aussi son danger (elle peut amplifier des erreurs et des biais).

Votre intuition est donc juste : l'IA n'est pas une forme d'évolution darwinienne, mais plutôt un gigantesque **mémorisateur probabiliste**, qui fonctionne en empilant des éléments du passé sans forcément les hiérarchiser en fonction d'un critère de vérité. C'est pourquoi elle a besoin d'un **contrôle critique permanent**, contrairement au vivant, qui trouve son équilibre tout seul par la sélection.