

IA générative et compétences informationnelles

Spokus.eu

0.

Introduction

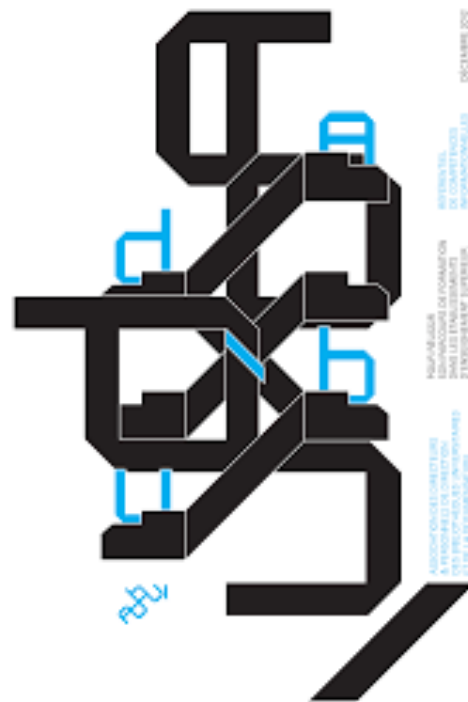
Compétence informationnelle

« Fait de disposer d'un ensemble de connaissances et de compétences permettant de reconnaître l'existence d'un **besoin d'information** et d'avoir la capacité et la discipline de **trouver l'information** adéquate, de la **comprendre**, de **l'exploiter** et de la **communiquer**, dans une perspective de résolution de problème ou de prise de décision. »

Grand dictionnaire terminologique

5 compétences (Référentiel de l'ADBU)

- [1] Identifier un **besoin d'information** et les **sources** disponibles
- [2] **Accéder aux informations**
- [3] **Évaluer** de façon critique l'information
- [4] **Produire et communiquer** le résultat de ses recherches d'information
- [5] **Valoriser** sa recherche et tendre vers la **science ouverte**



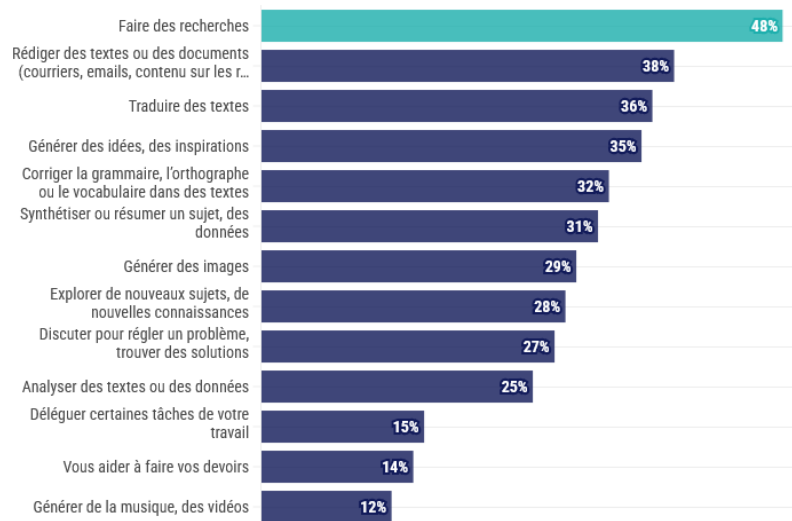
L'IA-G dans les pratiques informationnelles

39 %

des Français utilisent
activement l'IA-G

Recherche d'information [2]
et rédaction [4] en tête des
pratiques

Cas d'utilisation des IA génératives



Source : L'usage de l'intelligence artificielle par les Français | Ipsos-CESI École d'ingénieurs *

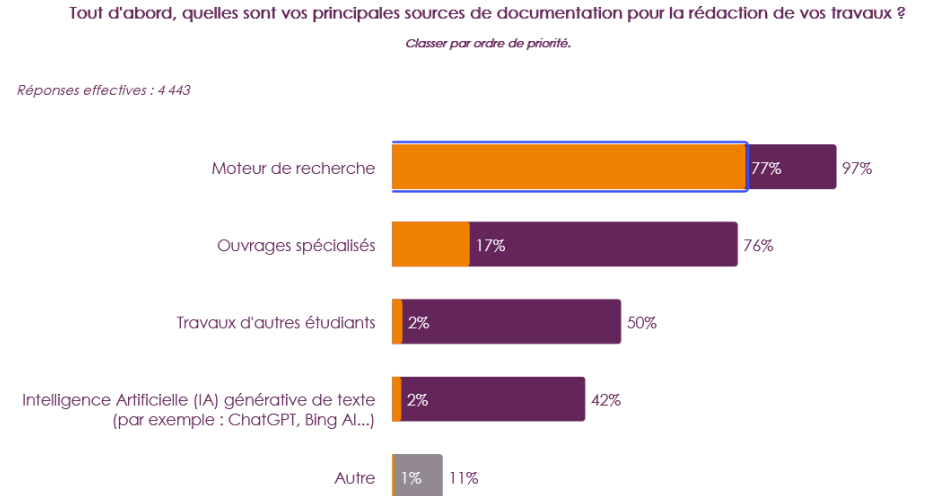
Base : 483 utilisateurs
Échantillon : 1000 personnes, représentatives de la population française âgée de 18 à 75 ans
Méthodologie complète disponible dans le [rapport d'étude](#)

L'IA-G dans les pratiques informationnelles

42 %

des étudiants utilisent l'IA-G
comme une **source de
documentation**

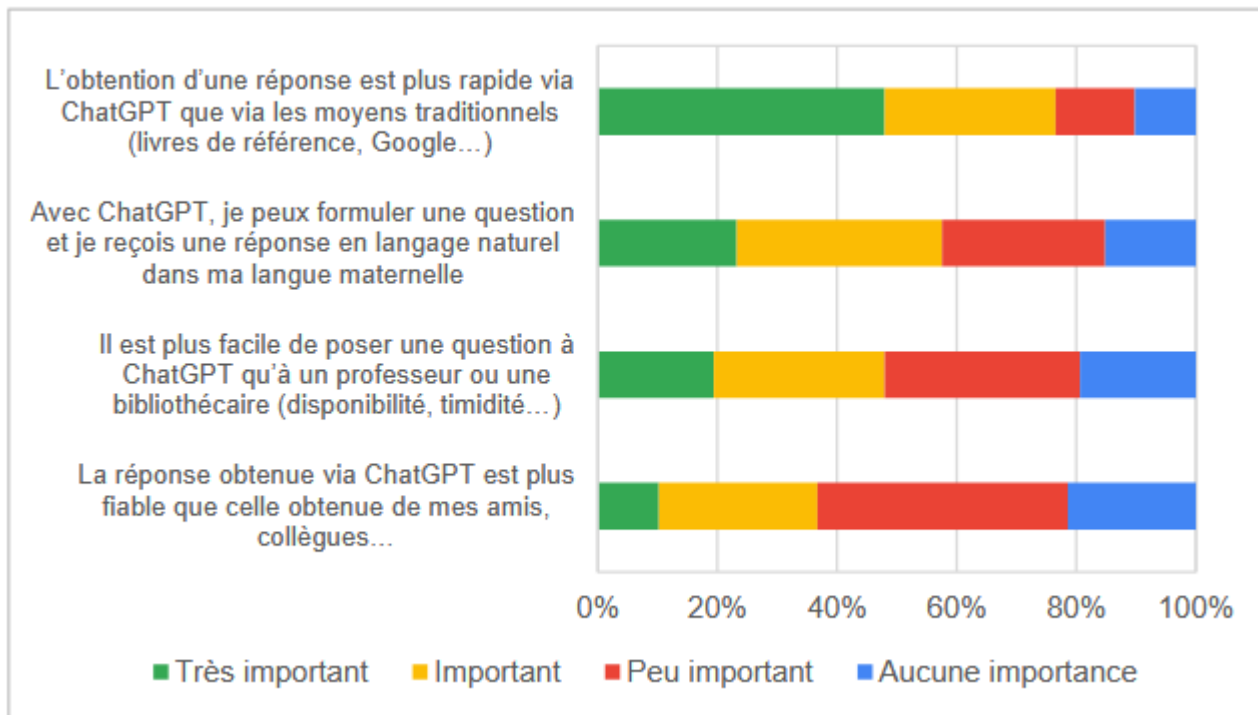
Recherche d'idées ^[1] et
d'informations ^[2] en tête des
pratiques.



(Compilatio, 2023)

L'IA-G dans les pratiques informationnelles

Pour quelle raison utiliser ChatGPT ?



(Lerinckx, 2024)

« **L'intelligence artificielle**
est la science de faire faire à
des machines des choses
qui demanderaient de
l'intelligence si elles étaient
faites par des humains »

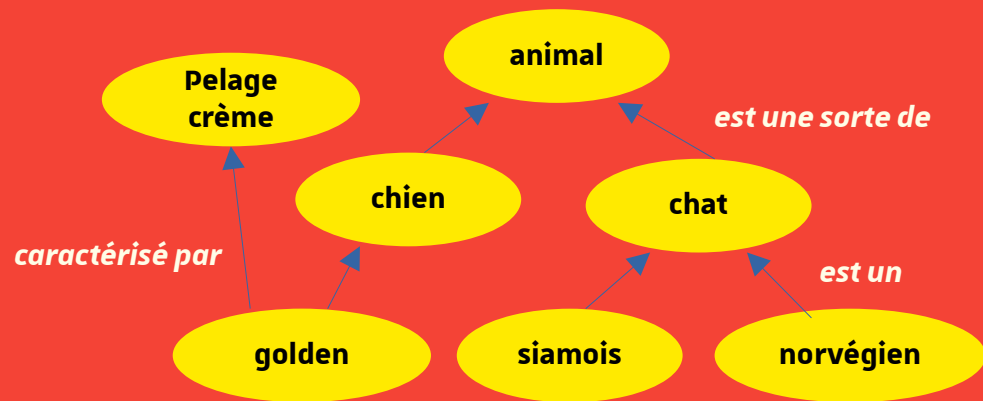
Marvin Minsky, 1968

Marvin Minsky à l'OPLC,
2008 (source : Wikimedia)

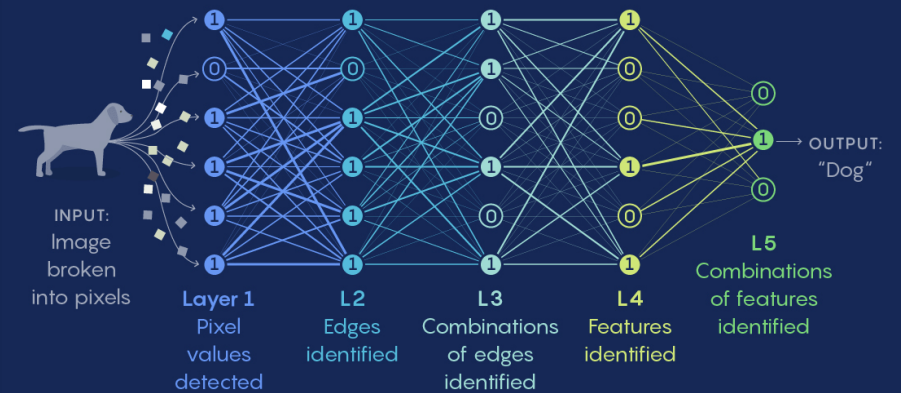




(Mattioli, 2020)



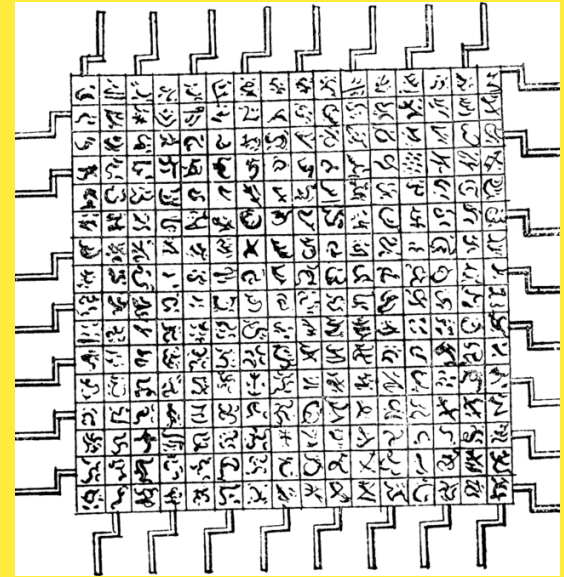
IA symbolique



IA connexionniste

Une **IA générative de texte** est un type d'**IA connexionniste** entraînée à partir d'un **grand nombre de textes**.

Grâce à son **modèle de langage**, elle peut produire tout type de contenus (articles, définitions, poèmes, etc.) en réponse à une demande précise (ou **prompt**).



Dans Les Voyages de Gulliver (1726), Jonathan Swift imagine une machine reposant sur la combinaison des mots, et sensée générer n'importe quel discours.

Pourquoi avoir attendu les années 2020 pour
disposer d'IA génératives de texte performantes ?

Objectifs de la formation

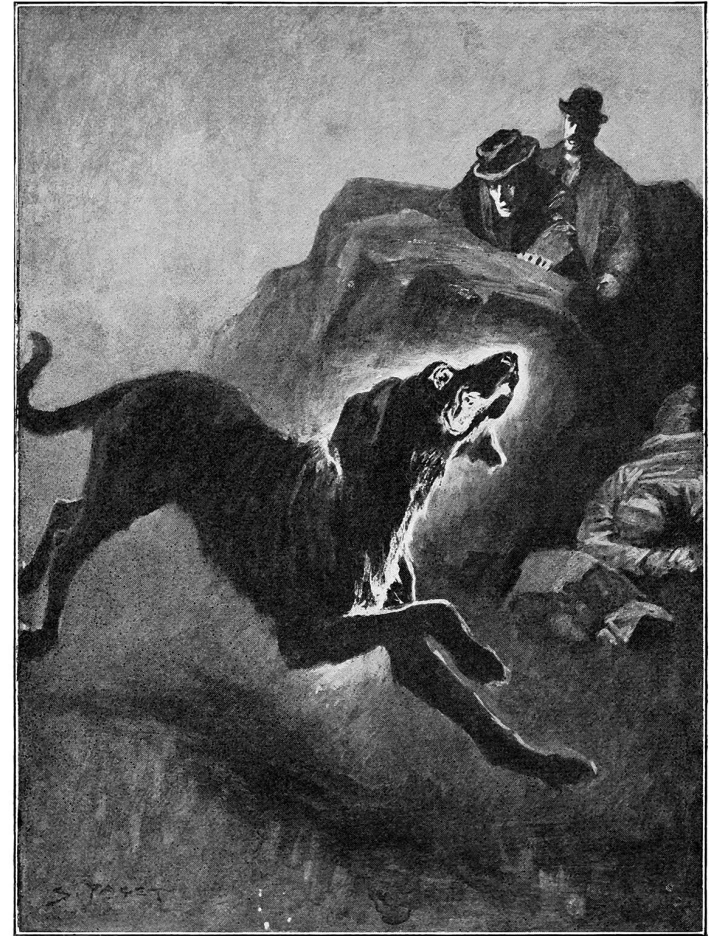
- ✓ Comprendre comment les **IA génératives restituent et produisent de l'information à partir de modèles de langage**, en comparaison avec le modèle des catalogues et bases de données bibliographiques, mais aussi les **interactions possibles entre IA génératives et bases de données**
- ✓ Explorer les **usages des IA génératives dans la recherche, la compréhension, l'analyse, la synthétisation et la production d'information**, et leurs **enjeux éthiques et écologiques** : biais cognitifs, désinformation, citabilité, respect du droit d'auteur, coût écologique, etc.
- ✓ Réfléchir à une **éthique de l'usage des IA génératives**, et comment sensibiliser les usagers à cette éthique dans le cadre de la transmission des compétences informationnelles

1.

L'IA pour
**définir son besoin en
information** ^[11]

Victimes d'une mauvaise
réputation, les chiens noirs
ont-ils moins de chances d'être
adoptés ?

Menons l'enquête !



Sidney Paget's illustration of The Hound of the Baskervilles.

A.

L'IA pour
questionner son sujet ^[111]
construire sa problématique ^[113]

Ex. 1. « *Victimes d'une mauvaise réputation, les chiens noirs ont-ils moins de chances d'être adoptés ?* »

Une question scientifique est une question qui peut être répondue de manière objective et systématique en utilisant des méthodes scientifiques

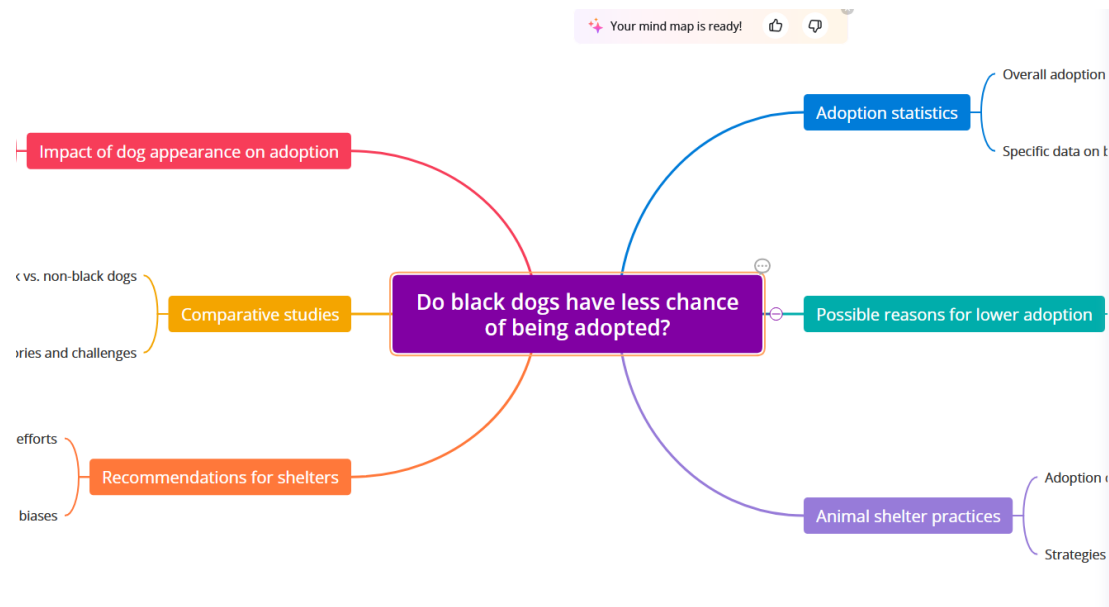
Est-ce une bonne **question scientifique** ?

Soumettez la à *Mistral* (équipe 1), *ChatGPT* (équipe 2) et *Perplexity* (équipe 3) et trouvez des pistes d'amélioration ...

Après cet exercice, quels sont, d'après-vous, les critères d'une bonne **question scientifique** ?
Les éléments à éviter ?



Générer une carte
mentale à partir d'un
sujet, d'une question
scientifique ou d'une
problématique



B.

L'IA pour
identifier des mots-clés [111]

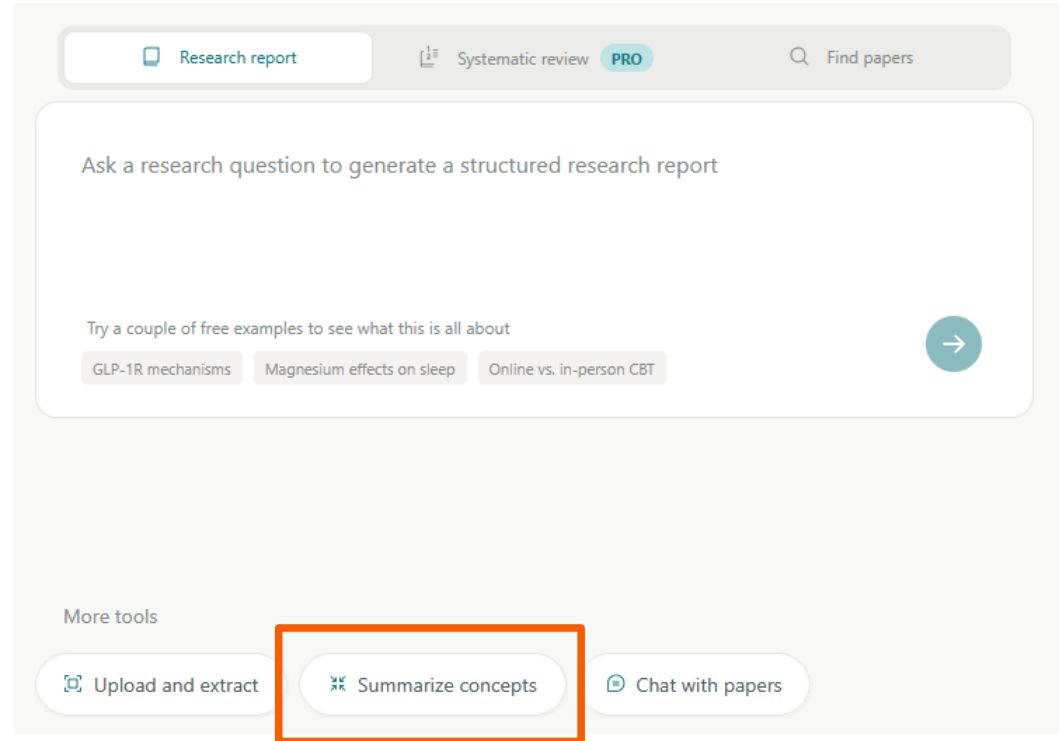
Ex. 2. A partir de votre question scientifique, identifiez des mots-clés, synonymes et termes associés ...

... sans utiliser l'IA ni Internet (équipe 1), sans utiliser l'IA (équipe 2), en utilisant l'IA de votre choix (équipe 3).

Représentez le résultat sous forme de carte mentale, de listes ou de tableau, etc. suivant votre choix.



Identifier des concepts
à partir d'articles
scientifiques via Elicit

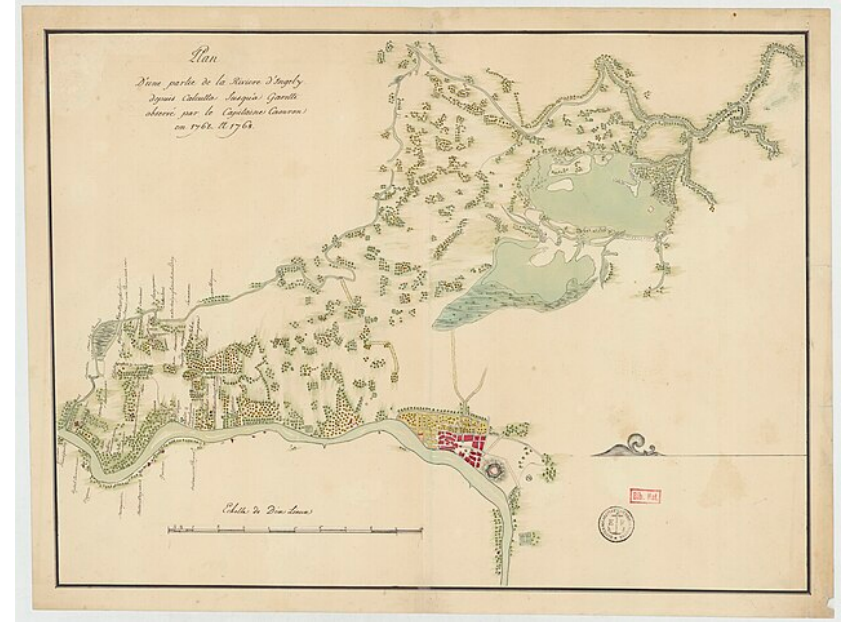


C.

L'IA pour
planifier sa recherche ^[114]

Ex. 3. Maintenant, planifiez
votre recherche ...

... En utilisant l'IA de votre
choix (équipe 1 & 2), sans
utiliser l'IA (équipe 3).



Plan d'une partie de la rivière d'Ingely depuis Calcutta jusqu'à Garetti - observé par le capitaine Camron en 1762 et 1768 (source : Wikimedia)

2.

L'IA pour
accéder aux informations ^[2]

A.

Catalogues,
moteurs, bases de données ...
Et IA générative ?

Ex. 4. Recherchez des informations sur « *Comment survivre en forêt* » en utilisant ...

- équipe 1 : ChatGPT & catalogue
- équipe 2 : Perplexity & catalogue
- équipe 3 : ChatGPT & Google

Puis complétez le tableau suivant ...



Planche hors texte publiée dans *Atala* par le Viconte de Chateaubriand ; Avec les dessins de Gustave Doré.
Louis Hachette (Paris), 1863, p. VI.

	<i>Perplexity</i>	<i>ChatGPT</i>	<i>Google</i>	<i>Sudoc</i>
Interface utilisateur				
Type d'interaction avec l'outil				
Langage utilisé par l'utilisateur (<i>normalisé vs. naturel</i>)				
Format de réponse				
Possibilités d'affiner la réponse / le résultat				
Technologie sous-jacente				
Reproductibilité des résultats				

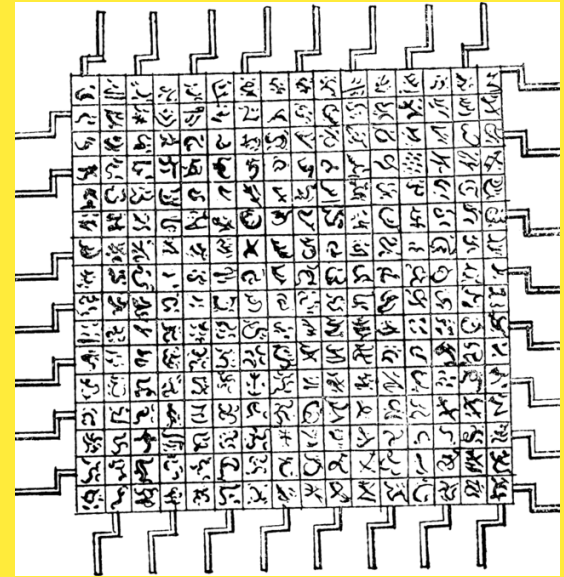
	<i>Perplexity</i>	<i>ChatGPT</i>	<i>Google</i>	<i>Sudoc</i>
Interface utilisateur	Chat	Chat	Moteur de recherche	Moteur de recherche
Type d'interaction avec l'outil	Prompts. Interaction conversationnelle continue.	Prompts. Interaction conversationnelle continue.	Requête	Requête
Langage utilisé par l'utilisateur (<i>normalisé vs. naturel</i>)	Naturel	Naturel	Naturel + normalisé : mots-clés, opérateurs de recherche, etc.	Normalisé : mots-clés, opérateurs de recherche
Format de réponse	Réponse en langage naturel appuyée sur des références.	Réponse en langage naturel.	Liste de résultats, recherches associées, bloc de questions / réponses.	Liste de résultats.
Possibilités d'affiner la réponse / le résultat	Relance ou réponse de l'utilisateur.	Relance ou réponse de l'utilisateur.	Filtres de recherche	Filtres de recherche
Technologie sous-jacente	Modèle de langage, Retrieval-augmented generation (RAG)	Modèle de langage, Retrieval-augmented generation (RAG)	Indexation du web et classement automatiques + moteur de recherche	Base de données bibliographique enrichie manuellement + interface de recherche
Reproductibilité des résultats	Très partielle : chaque prompt donne un résultat ≠ même si des éléments peuvent se recouper	Très partielle : chaque prompt donne un résultat ≠ même si des éléments peuvent se recouper	Partielle car résultats personnalisés	Complète

B.

Comment l'IA-G produit-elle de l'information ?

Une **IA générative de texte** est un type d'IA entraînée à partir d'un grand nombre de textes.

Grâce à son **modèle de langage**, elle peut produire tout type de contenus (articles, définitions, poèmes, etc.) en réponse à une demande précise (ou **prompt**).



Dans *Les Voyages de Gulliver* (1726), Jonathan Swift imagine une machine reposant sur la combinaison des mots, et sensée générer n'importe quel discours.



3 %
Wikipédia

Le corpus d'entraînement de ChatGPT-3



16 %
Livres numériques

(Geffen, 2023)

80 %
Archive du web





Top-500



4,2 %
Contenu en
français

44 %
Contenu en
anglais

5,4 %
Contenu en
chinois





Complète le texte
suivant :
« *Le ciel est ...* »

à partir de

« *Le **bleu** du **ciel*** »

« *Il fait beau, le **ciel** est **bleu*** »

Corpus d'entraînement

« *La météo annonce un **ciel nuageux*** »

« *Laisseront place à un **ciel voilé**
en fin d'après-midi* »

Modèle de langage

« cartographie » des relations entre
les **tokens** (unités textuelles), d'un
point de vue statistique.

= permet de choisir le mot qui a le
+ de chance d'être utilisé, en
prenant en compte les mots du
prompt, et de l'**historique de
dialogue**.

Historique de dialogue

« mémoire » des échanges avec
l'utilisateur
= permet d'éviter les erreurs
d'interprétation, les répétitions, de
donner un **contexte**, etc.



« *bleu* »

Bleu
Nuageux
Voilé

Une épistémologie probabiliste

→ information la +
probable (et plausible
pour les humains)
d'après les régularités
du langage modélisées
à partir d'un corpus



*Jackson Pollock. Autumn Rhythm (1948).
Illustration suggérée par ChatGPT.*



Carl Spitzweg. *Le rat de bibliothèque* (1850)

Ce qui implique que

→ + le corpus est **grand** et diversifié, + les réponses seront **fiables**

→ + un énoncé est **présent** dans le corpus, + il a de chances d'être **restitué**

A l'inverse, quand un sujet est peu ou pas présent dans son corpus, l'IA peut
« **halluciner** »



The cheshire cat vanishing (detail) in Lewis Carroll's Alice in Wonderland drawn by John Tenniel (1820-1914)

Ex. 5. Demandez à l'IA générative de votre choix de retrouver le film dans lequel ce dialogue a été prononcé :

- *Comment est votre blanquette ?*
- *Elle est bonne.*
- *On me dit le plus grand bien des harengs pomme-à-l'huile.*
- *Le patron vous en apportera un ramequin, vous vous ferez une idée.*

Puis, postez le titre dans le chat !



Bernard Blanc. Presbyterium de la basilique St Vital, VIe siècle, Ravenne, Emilie-Romagne, Italie. (source : Flickr).

C.

**Les IA, sources de
mésinformation ? ^[32]**

Désinformation

(« fake »)

Fausse information créée de façon intentionnelle, dans le but de tromper

Mésinformation

(« false »)

Fausse information créée de façon non-intentionnelle, suite à une mauvaise information, une sélection biaisée, etc.

Les hallucinations

Les **hallucinations** sont des réponses vraisemblables, mais fausses, générées par l'IA, à partir des éléments de son corpus d'entraînement.

Elles se produisent surtout quand un sujet est peu présent dans le corpus qui a servi à entraîner une IA générative.



The cheshire cat vanishing (detail) in Lewis Carroll's Alice in Wonderland drawn by John Tenniel (1820-1914)

Comment créer du faux avec du vrai ?

- **blanquette** = un plat traditionnel français, associée aux bouillons parisiens, à Maigret, San Antonio, Chirac, etc.
- ***Les Tontons flingueurs*** est un film culte français, populaire, daté des années 60, qui a une valeur nostalgique.

Association entre le dialogue de la blanquette et ce film est donc **vraisemblable**, mais **fausse**.



Les hallucinations

Ex. 6. Ce qui accentue la **mésinformation**, c'est que l'IA peut, sans le « savoir », déployer des procédés visant à rendre crédibles des informations fausses.

Lesquels voyez-vous dans vos réponses ? Et dans cet exemple ?

De quel film est issu ce dialogue : - Comment est votre blanquette ?
- Elle est bonne.
- On me dit le plus grand bien des harengs pomme-à-l'huile.
- Le patron vous en apportera un ramequin, vous vous ferez une idée.

Ce dialogue est tiré du film "Les Tontons Flingueurs", réalisé par Georges Lautner et sorti en 1963. Le scénario est de Michel Audiard, célèbre pour ses dialogues savoureux et percutants.

Ce passage fait partie de la fameuse scène de la cuisine, où les personnages, un peu éméchés, discutent de manière absurde autour d'une table, dans un mélange de dialogues gastronomiques et de considérations pseudo-philosophiques. C'est un des moments cultes du film.

Voici le contexte :

Les personnages, des gangsters plus ou moins à la retraite, se retrouvent pour parler affaires. Mais la discussion dégénère en une sorte de duel verbal décalé, truffé de répliques mémorables, comme celle-ci :

« On me dit le plus grand bien des harengs pommes à l'huile. »

Et la réponse imperturbable :

« Le patron vous en apportera un ramequin, vous vous ferez une idée. »

Le ton pince-sans-rire, les sous-entendus, et l'absurdité apparente de la conversation en font un chef-d'œuvre du dialogue à la Audiard.



Les fausses références

En demandant à ChatGPT :

Peux tu me conseiller des livres français sur le complotisme ?

9

résultats



6

fausses
références



L'ère du complotisme : La maladie d'un siècle – Marie Peltier (Les Liens qui Libèrent, 2016)

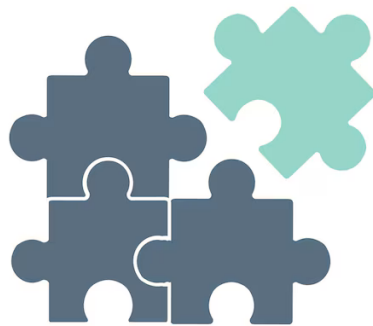


La fabrique du mensonge : Comment les théories du complot défient la science – Stéphane Foucart, Collectif (La Découverte, 2021)



Les fausses références

Contrairement à un **catalogue de bibliothèque**, les données qui ont servi à entraîné ChatGPT sont globalement **non-structurées, ni étiquetées**.



A partir de ces motifs, il peut construire de **fausses références**, qui paraîtront **plausibles**.

C'est de **texte brut**, décomposé en **tokens**, avec lequel l'IA a appris à identifier et recréer des **motifs linguistiques**.

IA et mésinformation : les biais ethniques et culturels

Les IA-G produisant de l'information à partir de **corpus d'entraînement** constitués par des humains, elles reflètent, voire amplifient les biais de ces corpus.

Biais

- prédominance de **visages blancs et masculins** dans les corpus issus de sites de recherche d'images en ligne (Mitchell, 2021)
- prédominance de **sources anglophones** dans ChatGPT
- prédominance des **profils masculins** dans les données de recrutement.

Conséquences

- graves **erreur dans l'étiquetage de photos de visages** de personnes non-blanches (scandale Google Photos de 2015)
- réponses par défaut adaptées à un **contexte anglo-saxon**
- reproduction des **discriminations à l'embauche** ([cas Amazon, 2014](#))

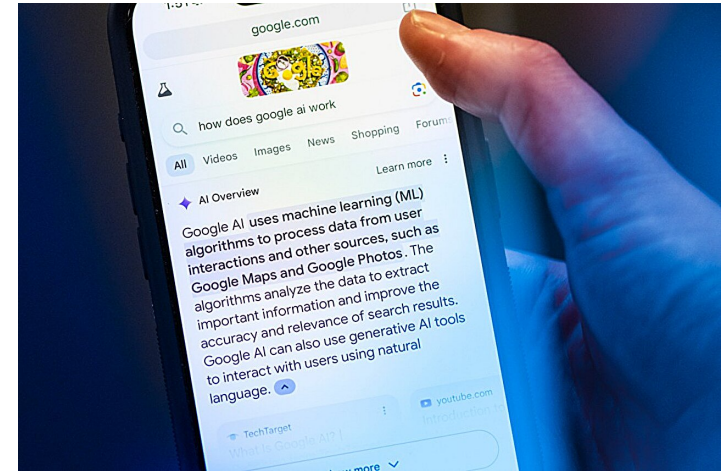
D.

Outils hybrides



Intégration de l'IA-G dans un moteur de recherche (Google)

Synthèse automatique des informations contenues dans les pages listées en résultats de recherche (à venir en France), formulée comme une réponse à la requête des utilisateurs.



Source : [TechXplore.com](https://www.techxplore.com).

Intégration d'une fonctionnalité de recherche dans une IA-G

Perplexity

outil qui se positionne comme une IA-G moins « créative », et + orientée vers la recherche d'information que ChatGPT, Claude, etc.



Outils de recherche combinant bases d'articles scientifiques et IA générative

Basés sur **Semantic Scholar**,
base d'articles scientifiques
enrichie par l'IA couvrant tous
les domaines du savoir.

Spécialisés dans la génération
de **revues de la littérature**.

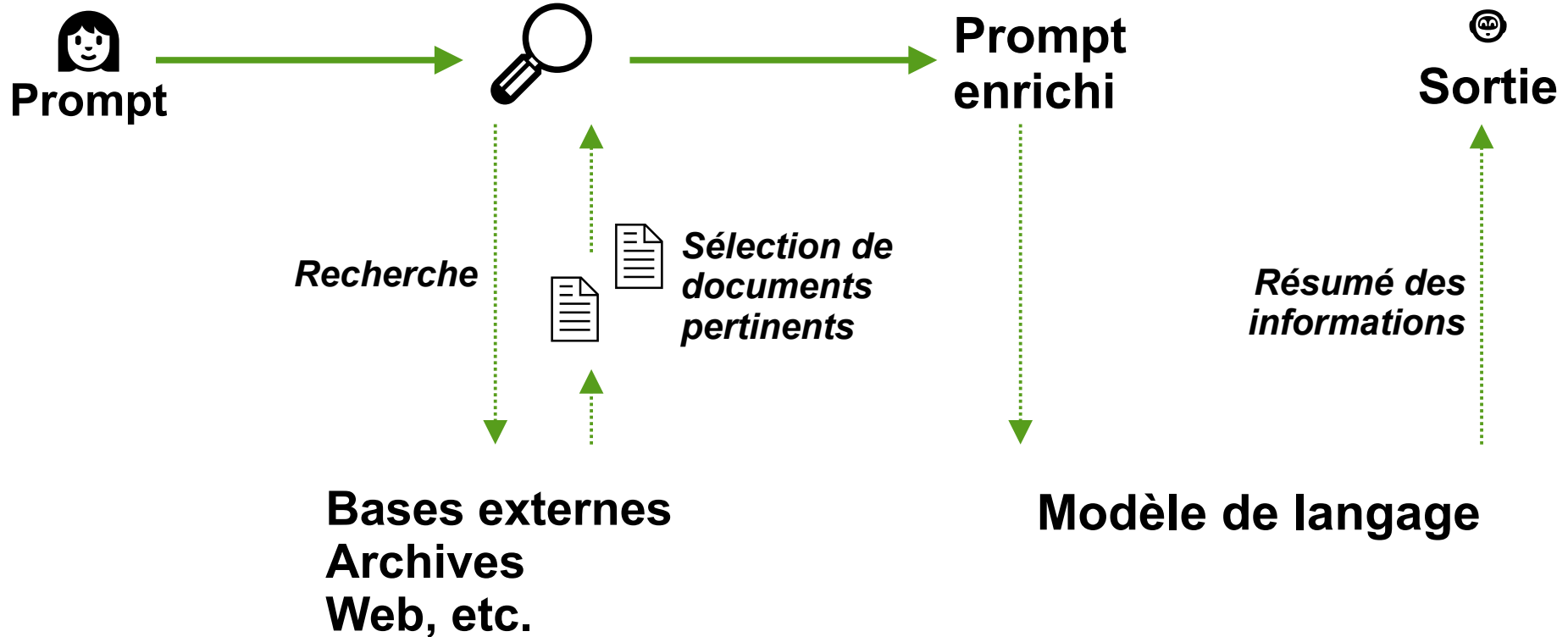


Des outils basés sur la *retrieval-augmented generation*

« Le principe du **RAG** est donc d'augmenter (A) la fonction générative (G) avec une fonction de recherche (R) dans un corpus externe »

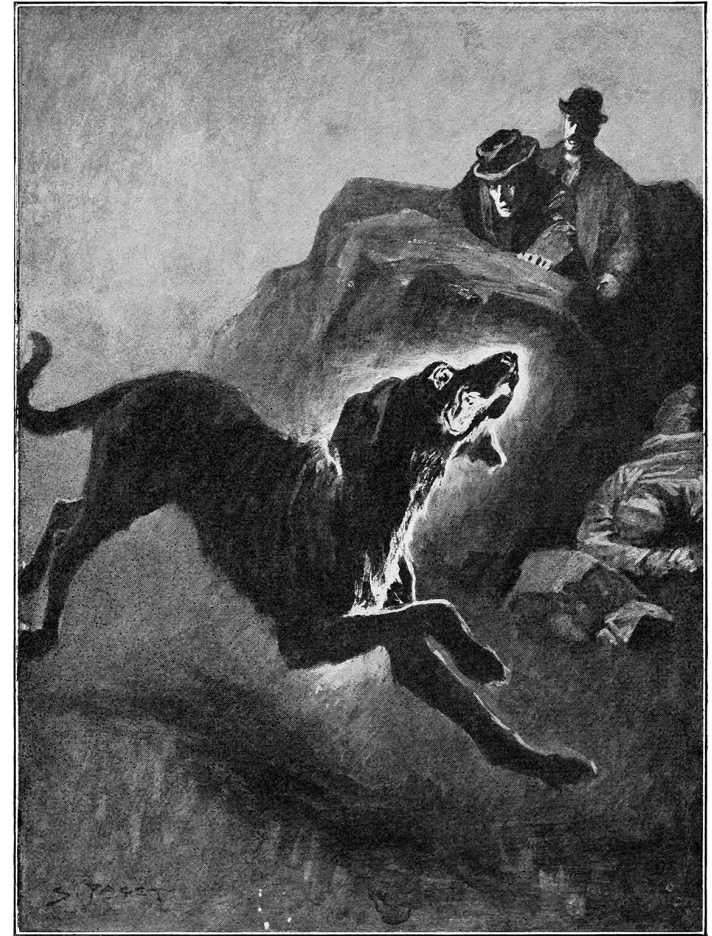
Emmanuelle Bermès, 2024

Des outils basés sur la *retrieval-augmented generation*



Ex. 7. Reprenez votre question scientifique sur le « syndrome du chien noir », et utilisez-là pour interroger Perplexity. Puis Consensus ou Elicit.

Quelle différence constatez-vous entre ces outils ? Donnent-ils la même réponse ? Comment celle-ci est-elle présentée ? S'appuient-ils sur les mêmes sources ? Explicitent-ils leur raisonnement ?



Sidney Paget's illustration of The Hound of the Baskervilles.

3.

**L'IA pour produire et
communiquer l'information [4]**

A.

**Exploiter l'information
avec les assistants IA ^[43]**

→ Créer un corpus de documents (cours, fiches, articles...)

→ Poser des questions sur leur contenu

→ Générer des synthèses, listes de dates, auteurs et acteurs, cartes mentales, résumés audio, fiches de révision ...



ChatGPT

B.

Communiquer l'information [44]

Ex. 8. A l'aide de l'IA générative de votre choix, chargez le support de formation et générez un texte à destination de vos collègues.

Il devra résumer la formation pour quelqu'un qui n'y a pas assisté, en se centrant sur les apports et limites de l'IA dans la recherche d'information, et ce qu'elle change aux compétences informationnelles.

Puis, générez un texte à destination de lycéens sur les enjeux de l'IA pour la recherche d'information, sous forme d'article de presse.

C.

L'art du prompt



Prompt = Liste d'instructions donnée en langage naturel à une intelligence artificielle pour générer un texte, une image, une vidéo...

Avec les IA-G, nous ne sommes pas face à des IA symboliques, reposant sur une logique stricte ...

Il n'y a donc pas de « recette magique », rédiger un bon prompt s'acquiert par la pratique ... Et mobilise des qualités littéraires !



Gabriel Metsu. Man Writing a Letter.
(source : Wikimedia)

- Être le + **précis et détaillé** possible
- Donner un maximum d'éléments de **contexte**
- Préciser son **rôle** (« je suis un bibliothécaire-formateur »), à qui on s'adresse ou à quel **public** on souhaite s'adapter (« des lycéens » ; « des seniors »)
- Donner des **exemples** de ce qu'on attend (ex : post Instagram, texte dans le style attendu, données étiquetées, etc.)



Gabriel Metsu. Man Writing a Letter.
(source : Wikimedia)

- Recourir à l'**itération** et au **feedback**, signaler les éléments à améliorer, etc.
- Donner un **format de sortie** (tableau, article, etc.) et la **longueur attendu**.
- Définir un style.
- Ouvrir un nouveau chat si l'IA est trop influencée par l'**historique de dialogue**.



Gabriel Metsu. Man Writing a Letter.
(source : Wikimedia)

4.

Enjeux éthiques de l'IA

A.

Enjeux écologiques et sociaux de l'IA

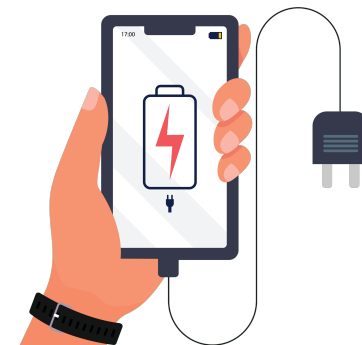
Coût écologique d'utilisation des IA génératives

1,14 g

de CO₂e par page de
texte générée par **Mistral
Large 2**.



45 ml d'eau par page de
texte générée par **Mistral
Large 2**.



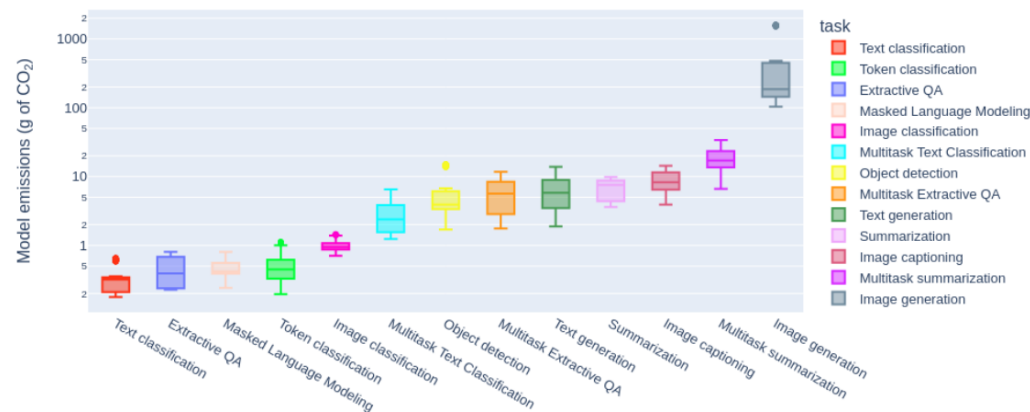
1 génération d'image
peut représenter
jusqu'à l'équivalent
d'une **1/2 charge de
smartphone** en
électricité, pour les
modèles les +
énergivores.

(*Carbone 4, 2025 ; Luccioni et al, 2024a et 2024b ;
TrenchTech, 2025*)

Coût écologique d'utilisation des IA génératives

→ **Générer des contenus** se révèle **+ énergivore** que d'autres tâches (étiquetage, classification, etc.)

→ **Traiter des images** est **+ énergivore** que traiter du **texte**.



(Luccioni et al, 2024a)

Des chiffres en trompe-l'oeil ?

Ces chiffres sont des **estimations**.

Le coût global de l'IA est **difficile à évaluer**, et pourrait être bien + élevé car :

→ il dépend de la **localisation des data-centers**

→ les entreprises du numériques sont peu transparentes sur ces données

→ il faut tenir compte du « cycle de l'IA » au complet



Des chiffres en trompe-l'oeil ?

20.400 tonnes de CO₂e ont, par exemple, été nécessaires pour **entraîner et mettre à disposition Mistral Large 2** sur 18 mois.

Le coût d'entraînement des IA est, pour les + grands modèles, dépassé en quelques semaines par les **usages individuels**, en raison de leur **massification**.



Source : Wikimedia,
LadyDragonflyCC - >;<

Que recommanderiez-vous pour minimiser l'impact de notre usage de l'IA générative ?

Impact écologique des IA génératives

👍 **Ne pas utiliser l'IA quand on peut faire aussi bien avec un autre outil**
(banques d'images gratuites, moteurs de recherche classique, etc.)

👍 **Maîtriser l'art du prompt** pour éviter les requêtes inutiles.

👍 **Privilégier des outils d'IA spécialisés**, plutôt que généralistes, suivant les tâches à accomplir.

Les IA-G ont aussi un coût social ...

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic

15 MINUTE READ



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's technology and shed light on the labor that makes it possible. Image generated by Dall-E 2/OpenAI

← Cette enquête révèle les conditions de travail difficiles de Kenyan embauchés par une entreprise de sous-traitance travaillant pour les grandes entreprises du numériques, dont OpenAI.

Ces « petites mains », payées moins de 2\$ de l'heure, étaient chargées d'étiqueter des textes aux contenus violents afin d'entraîner un modèle capable de détecter et filtrer les propos toxiques de ChatGPT.

B.

IA et désinformation

Désinformation

(« fake »)

Fausse information
créée de façon
intentionnelle, dans le
but de tromper

Mésinformation

(« false »)

Fausse information
créée de façon non-
intentionnelle, suite à
une mauvaise
information, une
sélection biaisée, etc.

IA et désinformation : l'empoisonnement de données

Tay, un exemple de *data-poisoning* =
intégration de fausses informations dans les données
d'entraînement d'une IA.

- **Tay** (2016) apprenait directement à partir d'interactions sur Twitter
- Des groupes d'utilisateurs ont massivement spammé **Tay** avec des propos extrémistes et complotistes.
- **Tay** était conçu pour *imiter* le langage des utilisateurs. Elle n'en comprenait pas le sens, ni n'avait de mécanisme de modération.
- **Tay** s'est mis à tenir des propos racistes, complotistes, etc.

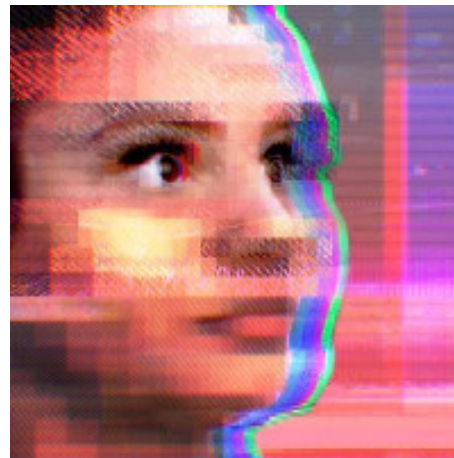


Photo de profil de Tay sur Twitter.

IA et désinformation : les *fake news*

Fake news = fausse information créée de toutes pièces, dans un but **politique** ou **économique**, reprenant souvent les **codes journalistiques**.

Pratique et terme ancien (daté du 19^e), médiatisés lors de la présidentielle US de 2016.

Avant l'IA-G, production « artisanale », à partir de copiés-collés de textes, en décontextualisant ou détournant des images ou vidéos.

→ compétences rédactionnelles, linguistiques, etc.

(ex : [les adolescents de Vélès, en Macédoine](#))

Anderlecht: le sapin de la place communale en feu la nuit du Nouvel An (vidéo)

Des jeunes ont mis le feu sur la Place du Conseil.



Par la rédaction

Publié le 1/01/2016 à 13:55 | Temps de lecture: 1 min

Cette image d'un incendie de sapin de Noël, provoqué par un groupe de jeunes, [a été utilisée pour une fake news](#) prétendant que des migrants auraient incendié un sapin de Noël à Bruxelles en criant « Allahou Akbar »

IA et désinformation : les *fake news*

Diffusion des *fake news* s'appuie notamment :

→ sur les **biais des algorithmes** chargés de sélectionner les contenus avec lesquels nous sommes + à même d'interagir sur les réseaux (appel aux émotions, biais de confirmation, etc.)

→ sur des campagnes d'***astroturfing*** afin de publier, liker, interagir en masse avec des contenus, ou de faire monter un sujet ou un narratif dans les *trending topics*.

Ces campagnes reposent sur des humains ou des ***social bots*** utilisant des **techniques d'IA pour simuler le comportement des utilisateurs**.

→ sur des campagnes de micro-ciblage publicitaire



Rad Robots painting at the Meme Ranch in 2022 (source : Wikimédia)

IA et désinformation : les *fake news*

L'IA générative permet, quant à elle :

→ Production à l'échelle « industrielle » de textes et surtout de photos ou vidéos à partir de simples prompts

Ex : les **deepfakes**, vidéo, image ou enregistrement audio hyperréaliste, généré par une IA pour tromper en imitant une personne réelle.

→ **Frontière + floue entre le vrai et le faux** : avec l'évolution technologique, les IA génératives d'images ont corrigé nombre de défauts qui permettaient de détecter leur usage



Avant l'IA, ce n'était pas ça 😊

L'arnaque au faux Jean-Yves Le Drian

Ex. 9

IA ou pas ?
(et surtout,
pourquoi?)







HOLLYWOOD

C.

**IA-G, droits d'auteur
et respect de la vie privée**

Des auteurs « pillés » par les entreprises d'IA ?

→ **IA génératives de textes** sont entraînées à partir d'**archives du Web**, contenant de nombreux contenus sous droits.

→ Ces contenus sont repris, synthétisés, « reconstitués » par les IA **sans attribuer de sources**.

→ Par ailleurs, avec l'évolution des **moteurs de recherche** vers des **moteurs de réponses**, probable **baisse de trafic** vers des liens Web, pénalisant les créateurs de contenus textuels (cf image ci-contre ; source : Google).



Des auteurs « pillés » par les entreprises d'IA ?

→ **IA génératives d'images** sont aussi entraînées à partir de nombreux contenus sous droits.

→ **MidJourney** pourrait reproduire le **style de 20.000 artistes**.

→ **Multiplication des procès d'artistes, écrivains et titres de presse** pour non-respect des droits d'auteur.

→ Quelques rares accords (ex : OpenAI / Le Monde)

LE MONDE & VOUS · LE MONDE – NOTRE ACTUALITÉ

Intelligence artificielle : un accord de partenariat entre « Le Monde » et OpenAI

Cet accord pluriannuel, le premier entre un média français et un acteur majeur de l'IA, permettra à la société de s'appuyer sur le corpus du journal pour établir et fiabiliser les réponses de son outil ChatGPT, moyennant une source significative de revenus supplémentaires.

Par Louis Dreyfus (Président du directoire du « Monde ») et Jérôme Fenoglio (Directeur du « Monde »)

Publié le 13 mars 2024 à 18h30, modifié le 13 novembre 2024 à 12h10 · 🕒 Lecture 5 min. · [Read in English](#)

(Université de Lorraine et al., 2024)

Et les utilisateurs des IA ?

→ Certaines IA sont entraînées à partir de nos **données publiques**, présentes sur les **réseaux sociaux** (Reddit, X, Facebook, Instagram, etc.)

→ Notre **historique de dialogue** avec des IA-G sert à améliorer la pertinence des réponses, mais peut aussi être conservé et utilisé pour leur entraînement.

Ex : ChatGPT utilise ces données, sous forme agrégée et anonymisée.

→ Certaines IA proposent de **désactiver ces fonctionnalités** dans leurs paramètres.



Utilisation d'une IA et droit d'auteur

→ Fiches
« L'intelligence artificielle et le bibliothécaire »

Céline CAUBERT – Néhémie HENRY – Emilie LEROMAIN



Citer ses sources à l'ère des IAG

Les IA génératives (IAG) évoluant rapidement, les informations présentées ici sont susceptibles d'évoluer. N'hésitez pas à faire vous-même des tests pour confirmer ou infirmer les observations décrites.

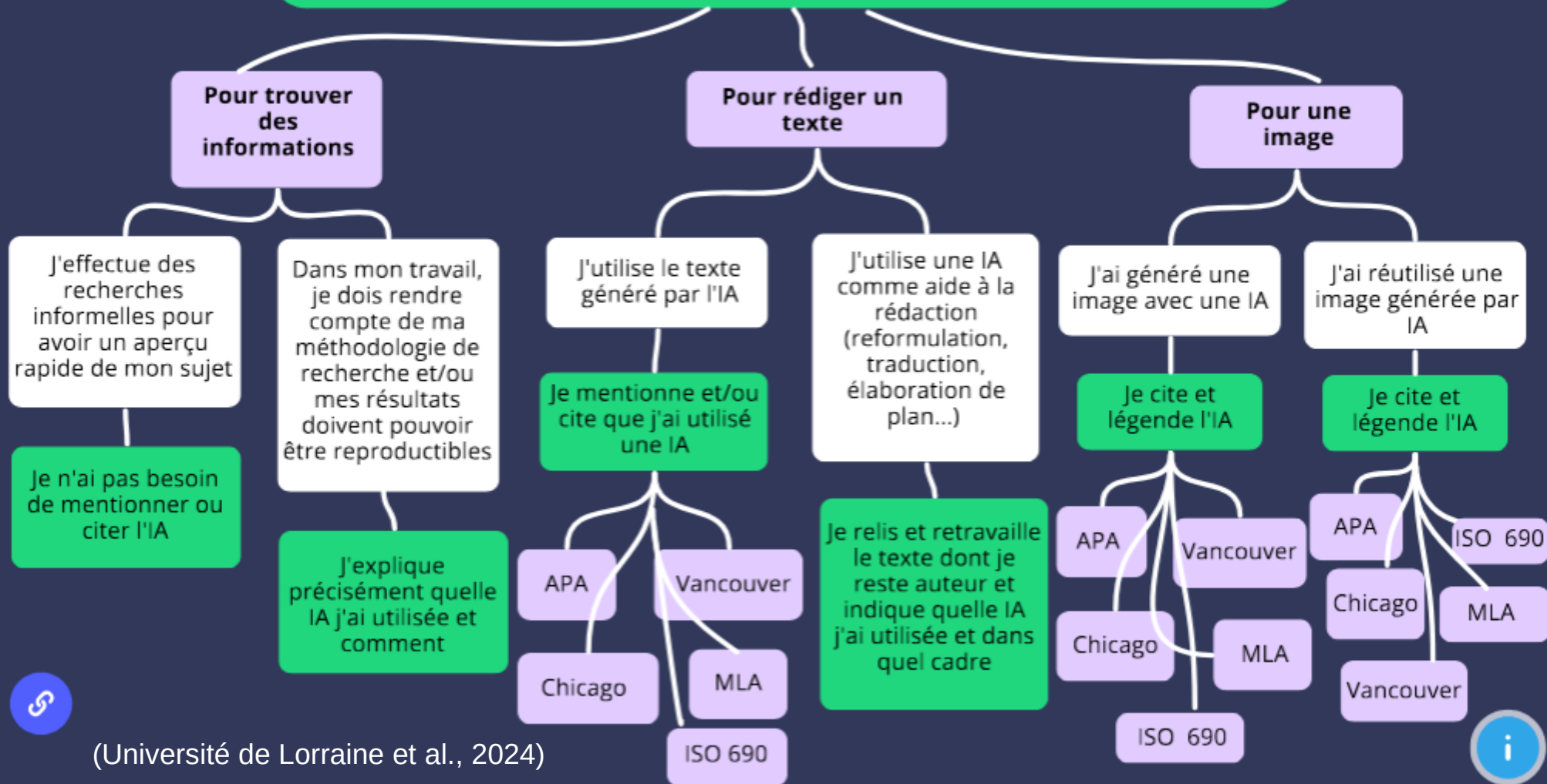
Comment citer une IA ?

D'après Alexei Grinbaum, Directeur de recherche et président du Comité opérationnel pilote d'éthique du numérique du CEA, il faut « déclarer et expliciter l'usage qui [...] est fait [des outils IA] ». Si l'IA est « créatrice de solutions inédites », les découvertes « revien[nen]t [...] à celui qui a formulé le bon prompt -ou requête- et à ceux qui ont fabriqué le logiciel, un collectif d'auteurs regroupant les personnes ayant élaboré le corpus d'entraînement, fabriqué le modèle, etc... »¹.

En France, on peut considérer de deux manières une œuvre créée par une IA :

- si elle est **créée** seulement **avec l'assistance d'un ordinateur**, c'est la **personne humaine** à l'origine des actions pour la création qui **est titulaire des droits sur l'œuvre** ;
- si elle est **générée automatiquement**, sans intervention humaine, elle **ne pourra pas être qualifiée d'« œuvre »** car seule une personne humaine peut être considérée comme auteur et non une machine. A l'heure actuelle (février 2023), les créations issues d'un prompt ne sont pas protégées par le droit d'auteur français².

J'utilise une IA, comment dois-je l'indiquer ?



5.

A vous de jouer !

Ex. 10. A partir de ce que nous avons vu, proposez une séquence de formation ou un atelier sur l'IA et les compétences informationnelles à destination de vos usagers !

**Merci pour votre
attention !**

A lire, à écouter, à voir



A lire, à écouter, à voir



Note

Aucune illustration n'a été générée par IA pour cette formation (hormis les exemples 1 et 3 de l'exercice 9).

Certains éléments de définitions et d'explication ont été écrits avec l'assistance de ChatGPT et Mistral.

« ChatGPT ». s. d. Consulté le 10 juin 2025. <https://chatgpt.com>.

« Mistral AI ». s. d. Consulté le 11 juin 2025. <https://mistral.ai/fr>.

Sources

Alexander, Craig Silverman, Lawrence. 2016. « How Teens In The Balkans Are Duping Trump Supporters With Fake News ». BuzzFeed News. 3 novembre 2016. <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo>.

« Common Crawl - Open Repository of Web Crawl Data ». s. d. Consulté le 10 juin 2025. <https://commoncrawl.org>.

« Crise migratoire : les fake news de l'extrême-droite – Spokus ». 2018. 19 décembre 2018. <https://spokus.eu/hoaxes-anti-refugies/>.

« culture de l'information ». s. d. Consulté le 10 juin 2025. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8363208/culture-de-linformation>.

« Enquête sur l'IA : enseignants et étudiants confrontent leurs regards ». s. d. Compilatio. Consulté le 10 juin 2025. <https://www.compilatio.net/blog/communiquer-presse-enquete-ia-2023>.

Gefen, Alexandre. 2023. Vivre avec ChatGPT: Séduire, penser, créer, se cultiver, s'enrichir... L'intelligence artificielle aura-t-elle réponse à tout ? Paris: L'OBSERVATOIRE.

« IA génératives : un risque accru de désinformation ? / Afis Science - Association française pour l'information scientifique ». s. d. Afis Science - Association française pour l'information scientifique. Consulté le 8 juin 2025. <https://www.afis.org/IA-generatives-un-risque-accru-de-desinformation>.

« Intelligence artificielle : quels sont les usages des Français ? | Ipsos ». 2025. 10 février 2025. <https://www.ipsos.com/fr-fr/intelligence-artificielle-quels-sont-les-usages-des-francais>.

« Intelligence artificielle : un accord de partenariat entre « Le Monde » et OpenAI ». 2024, 13 mars 2024. https://www.lemonde.fr/le-monde-et-vous/article/2024/03/13/intelligence-artificielle-un-accord-de-partenariat-entre-le-monde-et-openai_6221836_6065879.html.

Langlais, Pierre-Carl. 2023. « ChatGPT : comment ça marche ? » Billet. Sciences communes (blog). 7 février 2023. <https://doi.org/10.58079/twxr>.

Lerinckx, Dominique, et Bibliothèque des sciences et techniques de l'ULB. 2024. « Perception et adoption de l'IA : Enquête sur l'usage de ChatGPT par les étudiants, élèves de l'enseignement obligatoire, enseignants et chercheurs lors du Printemps des Sciences 2023 à l'ULB ». Université Libre de Bruxelles.

« L'IA dans l'enseignement : résultats détaillés d'une enquête où étudiants et enseignants confrontent leurs regards ». s. d. Compilatio. Consulté le 10 juin 2025. <https://www.compilatio.net/blog/enquete-ia-enseignement-2023>.

Lorraine, Université de, Université de Strasbourg, Université de Haute Alsace, Bibliothèque nationale et universitaire de Strasbourg, et GTFU Alsace. 2024. « L'intelligence artificielle et le bibliothécaire », juillet. <https://doi.org/10.5281/zenodo.13751602>.

Luccioni, Alexandra Sasha, Yacine Jernite, et Emma Strubell. 2024. « Power Hungry Processing: Watts Driving the Cost of AI Deployment? » In The 2024 ACM Conference on Fairness, Accountability, and Transparency, 85-99. <https://doi.org/10.1145/3630106.3658542>.

manuefig. 2024. « Le futur de la recherche documentaire : RAG time ! » Figoblog (blog). 30 mars 2024. <https://figoblog.org/2024/03/30/le-futur-de-la-recherche-documentaire-rag-time/>.

Mattioli, Juliette. 2020. « Vers un déploiement industriel de l'intelligence artificielle ». Association française pour l'avancement des sciences (AFAS) (blog). 5 octobre 2020. <https://www.afas.fr/vers-un-deploiement-industriel-de-lintelligence-artificielle/>.

Mitchell, Melanie, et Douglas Hofstadter. 2021. L'intelligence artificielle. Malakoff: Dunod.

« Nouvelles technologies: «Les IA génératives vont bientôt disparaître» ». 2025. 24 heures. 6 janvier 2025. <https://www.24heures.ch/les-ia-generatives-vont-bientot-disparaitre-332580420526>.

Parisien, Le. s. d. « L'audacieuse arnaque au faux Jean-Yves Le Drian arrive devant la justice ». Le Parisien. Consulté le 10 juin 2025. <https://podcasts.leparisien.fr/le-parisien-code-source/202001301654-laudacieuse-arnaque-au-faux-jean-yves-le-drian-arrive-devant>.

Perrigo, Billy. 2023. « Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer ». TIME. 18 janvier 2023. <https://time.com/6247678/openai-chatgpt-kenya-workers/>.

« Quand le logiciel de recrutement d'Amazon discrimine les femmes ». 2018. Les Echos. 13 octobre 2018. <https://www.lesechos.fr/industrie-services/conso-distribution/quand-le-logiciel-de-recrutement-damazon-discrimine-les-femmes-141753>.

« Référentiel des Compétences Informationnelles – ADBU ». s. d. Consulté le 9 juin 2025. <https://adbu.fr/referentiels/recif>.

Ségalat, Laurent. 1985. « L'Intelligence Artificielle ». Raison présente 76 (1): 83-97. <https://doi.org/10.3406/raipr.1985.2470>.

« The Environmental Impacts of AI -- Primer ». s. d. Consulté le 10 juin 2025. <https://huggingface.co/blog/sasha/ai-environment-primer>.

Trench Tech, réal. 2025. Sasha Luccioni - IA et Crise Climatique : Prêts pour le Choc ? <https://www.youtube.com/watch?v=SrP-20O4no0>.