

# INTELLIGENCES ARTIFICIELLES GÉNÉRATIVES PROMESSES & LIMITES

---

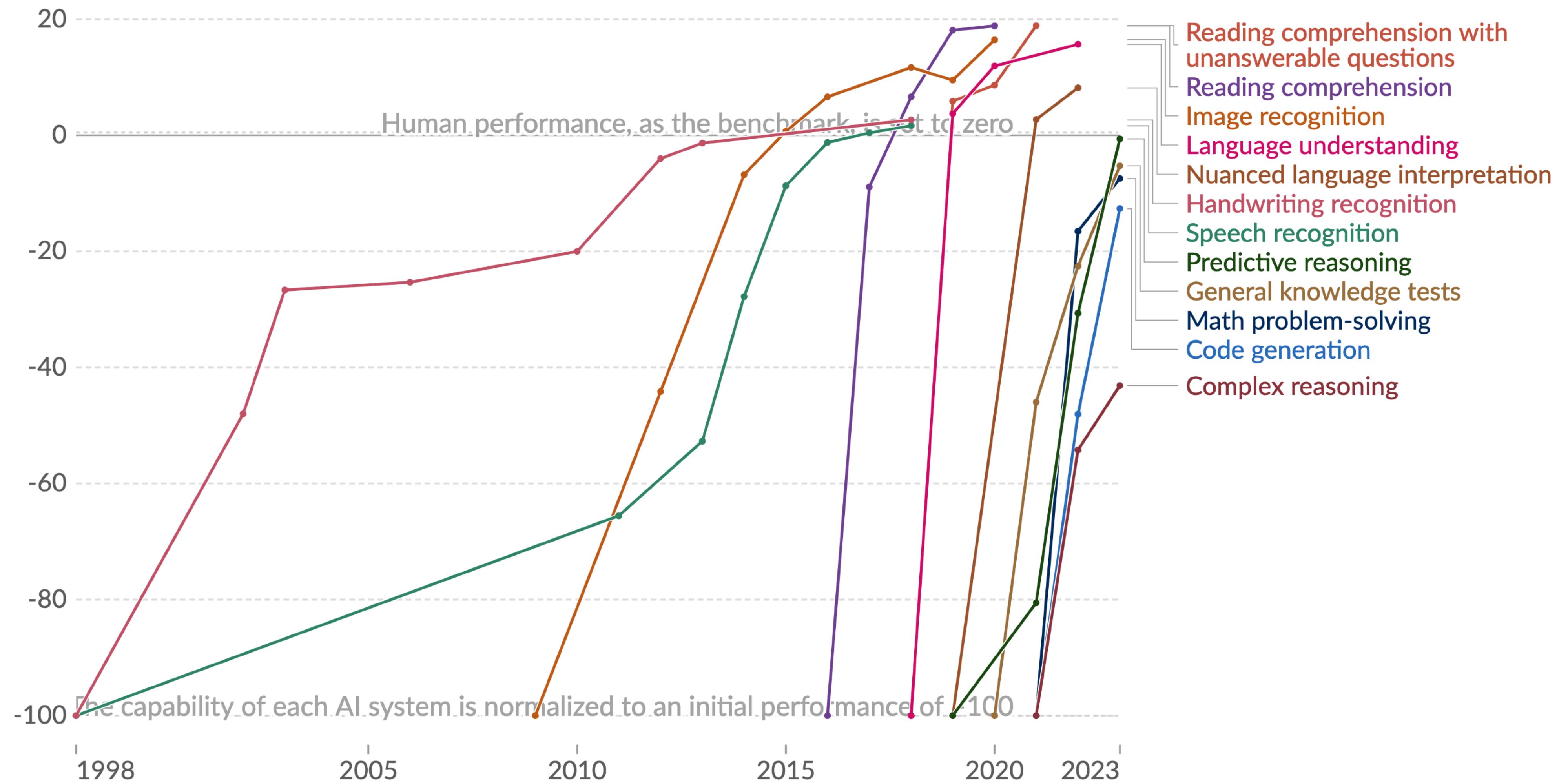
NICOLAS P. ROUGIER

INRIA BORDEAUX  
INSTITUT DES MALADIES  
NEURODEGENERATIVES



# Test scores of AI systems on various capabilities relative to human performance

Within each domain, the initial performance of the AI is set to -100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.



Data source: Kiela et al. (2023)

Note: For each capability, the first year always shows a baseline of -100, even if better performance was recorded later that year.

OurWorldInData.org/artificial-intelligence | CC BY

## IA GÉNÉRATIVES

## De quoi parle-t-on ?

De la capacité à générer des contenus de tous types en réponse à une invite (prompt):

- Image:** Dall-E 2 (OpenAI), MidJourney (Midjourney, Inc) Stable Diffusion (Stability.ai), etc.
- Texte:** ChatGPT (OpenAI), Claude (Anthropic), LLaMA (Meta), Gemini (Google), etc.
- Code:** CoPilot (GitHub)
- Biologie:** AlphaFold (DeepMind)
- Matériaux:** GnOME (DeepMind)
- Audio, Vidéo, Musique, etc**

Les résultats ne cessent de s'améliorer, notamment dans les domaines du texte et de l'image mais aussi dans des domaines plus confidentiels.

## Timeline of images generated by artificial intelligence

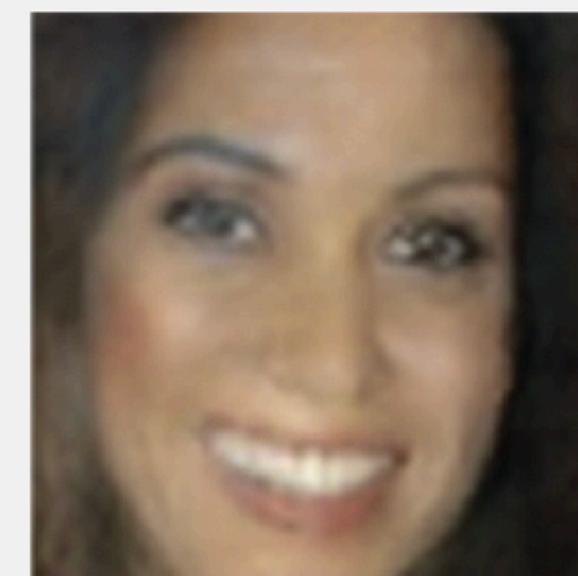
These people don't exist. All images were generated by artificial intelligence.

2014



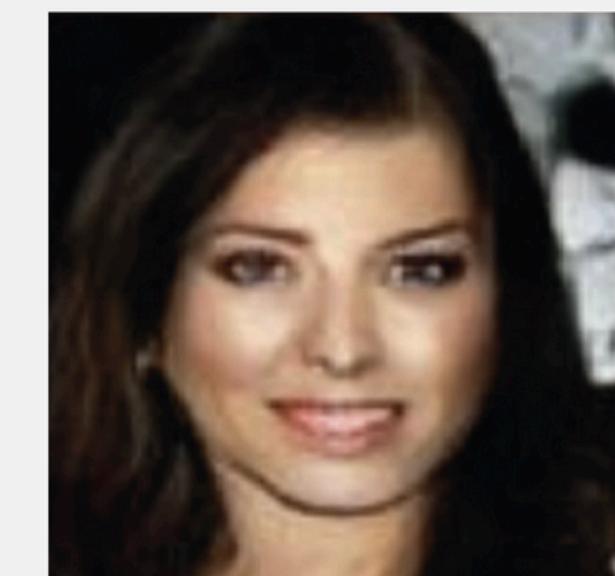
Goodfellow et al. (2014) – Generative Adversarial Networks

2015



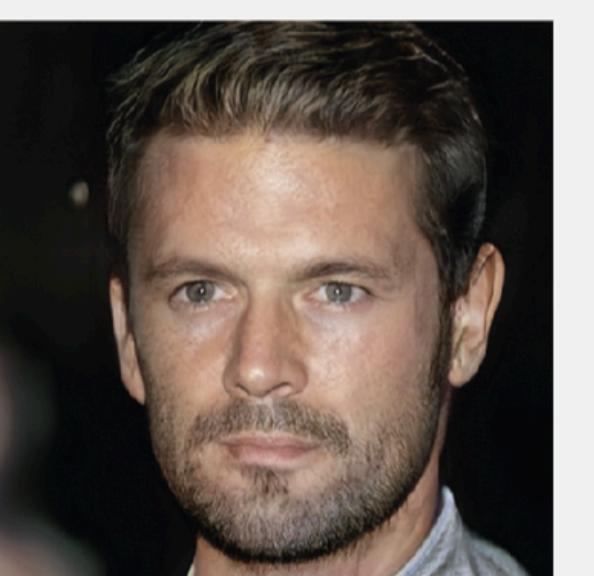
Radford, Metz, and Chintala (2015) – Unsupervised Representation Learning with Deep Convolutional GANs

2016



Liu and Tuzel (2016) – Coupled GANs

2017



Karras et al. (2017) – Progressive Growing of GANs for Improved Quality, Stability, and Variation

2018



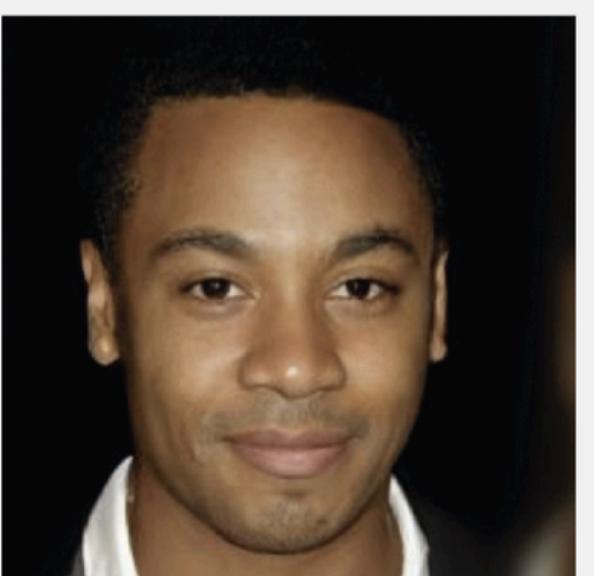
Karras, Laine, and Aila (2018) – A Style-Based Generator Architecture for Generative Adversarial Networks

2019



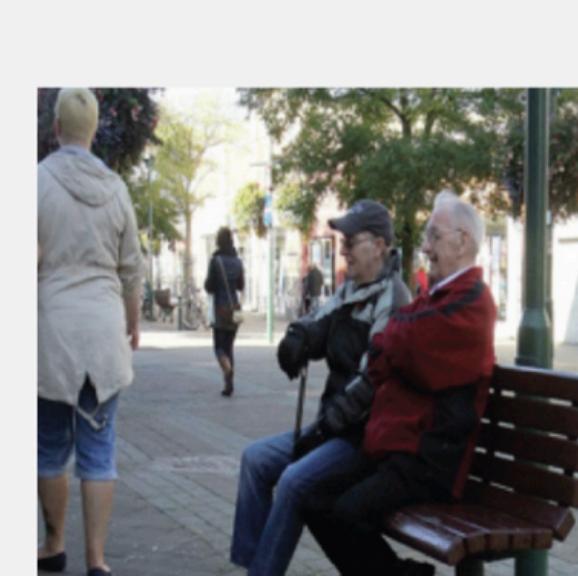
Karras et al. (2019) – Analyzing and Improving the Image Quality of StyleGAN

2020



Ho, Jain, &amp; Abbeel (2020) – Denoising Diffusion Probabilistic Models

2021

Image generated with the prompt:  
"a couple of people are sitting on a wood bench"2022 Image generated with the prompt:  
"A Pomeranian is sitting on the King's throne wearing a crown. Two tiger soldiers are standing next to the throne."

Saharia et al. (2022) – Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Google's Imagen)

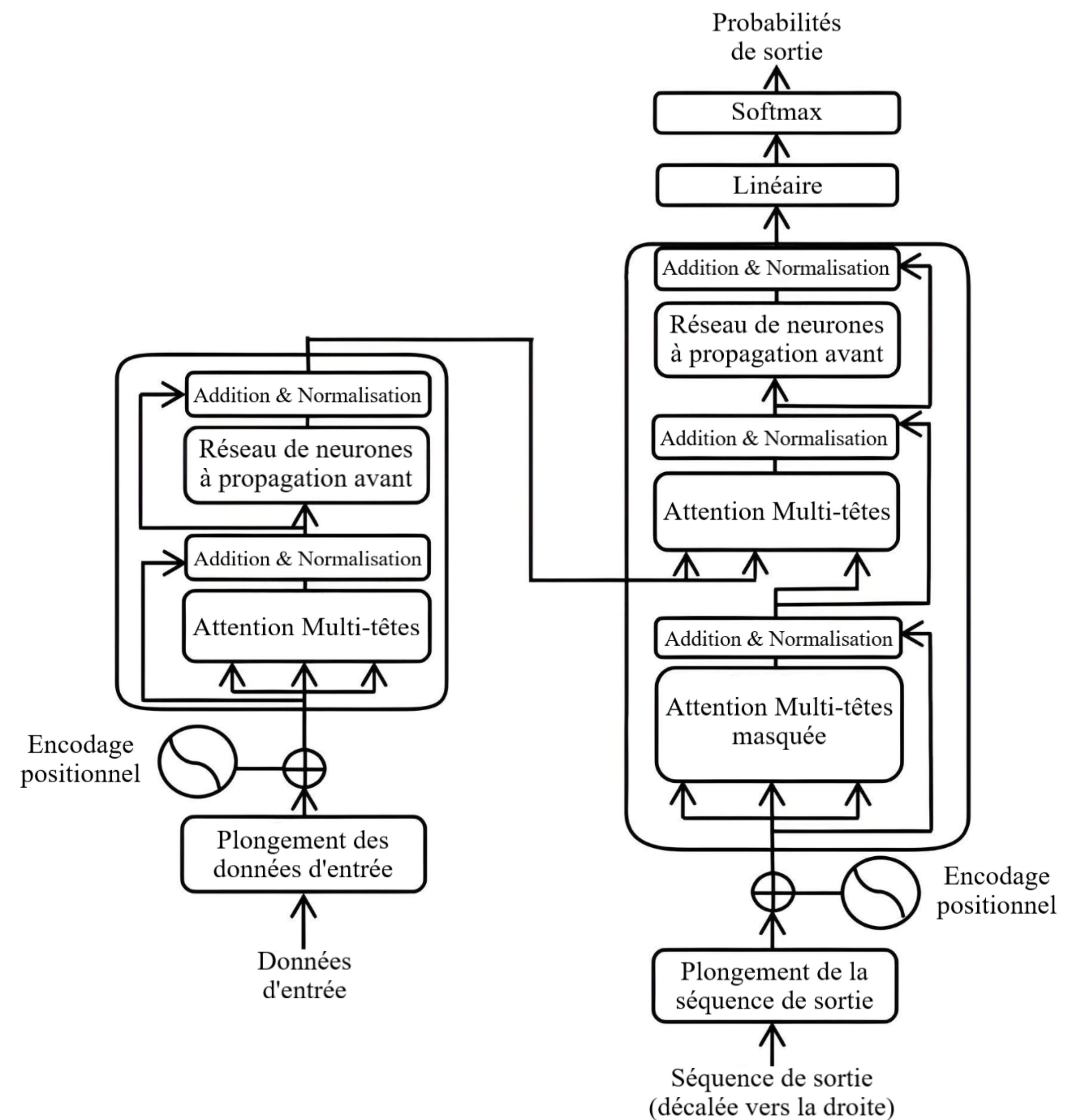
# Attention is all you need

En 2017, un article (cité plus de **120,000 fois**) a permis une percée pour les IA génératives en introduisant les [transformeurs](#):

Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz; Polosukhin, Illia (2017). **Attention is All you Need.** *Advances in Neural Information Processing Systems*. 30.

En 2018, l'architecture [BERT](#) (Bidirectional Encoder Representations from Transformers) a révolutionné les modèles de langage en se basant sur les *transformers*.

Toujours en 2018, la première version de [ChatGPT](#) (Generative Pre-trained Transformer) est publiée par OpenAI.



## Les ingrédients

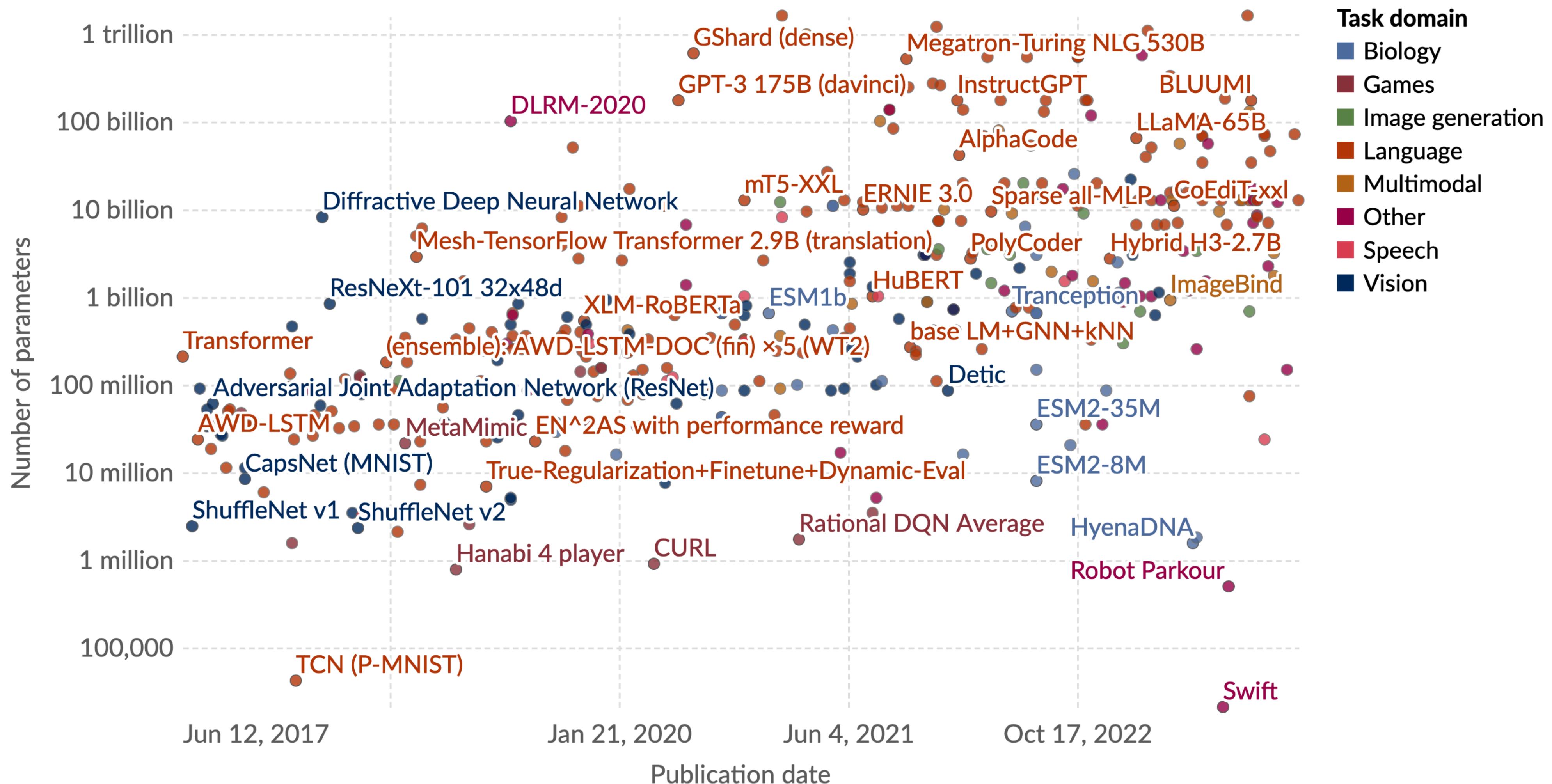
Les grands modèles de langages sont (grossièrement) des générateurs d'inférences statistiques et requièrent:

1. Un modèle (architecture)
2. Des corpus (données)
3. Des petites mains (annotation & alignement)
4. De la puissance de calcul (apprentissage)
5. Du temps (paramétrage)
6. De l'énergie et des ressources (serveurs)
7. De l'argent (qq millions à qq milliards)



# Parameters in notable artificial intelligence systems

Parameters are variables in an AI system whose values are adjusted during training to establish how input data gets transformed into the desired output; for example, the connection weights in an artificial neural network.



Data source: Epoch (2024)

Note: Parameters are estimated based on published results in the AI literature and come with some uncertainty. The authors expect the estimates to be correct within a factor of 10.

OurWorldInData.org/artificial-intelligence | CC BY

## IA GÉNÉRATIVES

## Les usages (potentiels)

Selon l'enquête, seuls 3% des dirigeants de TPE/PME en font un usage régulier et 12% un usage occasionnel.

Elise Tissier, Directrice de Bpifrance Le Lab:

*L'émergence des intelligences artificielles promet des changements et des opportunités dans le quotidien de tous les métiers, en particulier intellectuels, et quel que soit le secteur. Cette étude témoigne de l'usage limité qu'en font une majorité des dirigeants des TPE et PME françaises.*

ÉTUDE

# L'IA GÉNÉRATIVE DANS LES TPE ET PME

## IA génératives : opportunités et usages dans les TPE et PME



Trois quarts des dirigeants de TPE et PME ne trouvent pas l'usage des IA génératives. Comment les aider à les adopter ?

TRANSFORMATION DIGITALE

14 MARS 2024

2 MIN



ENVOYER



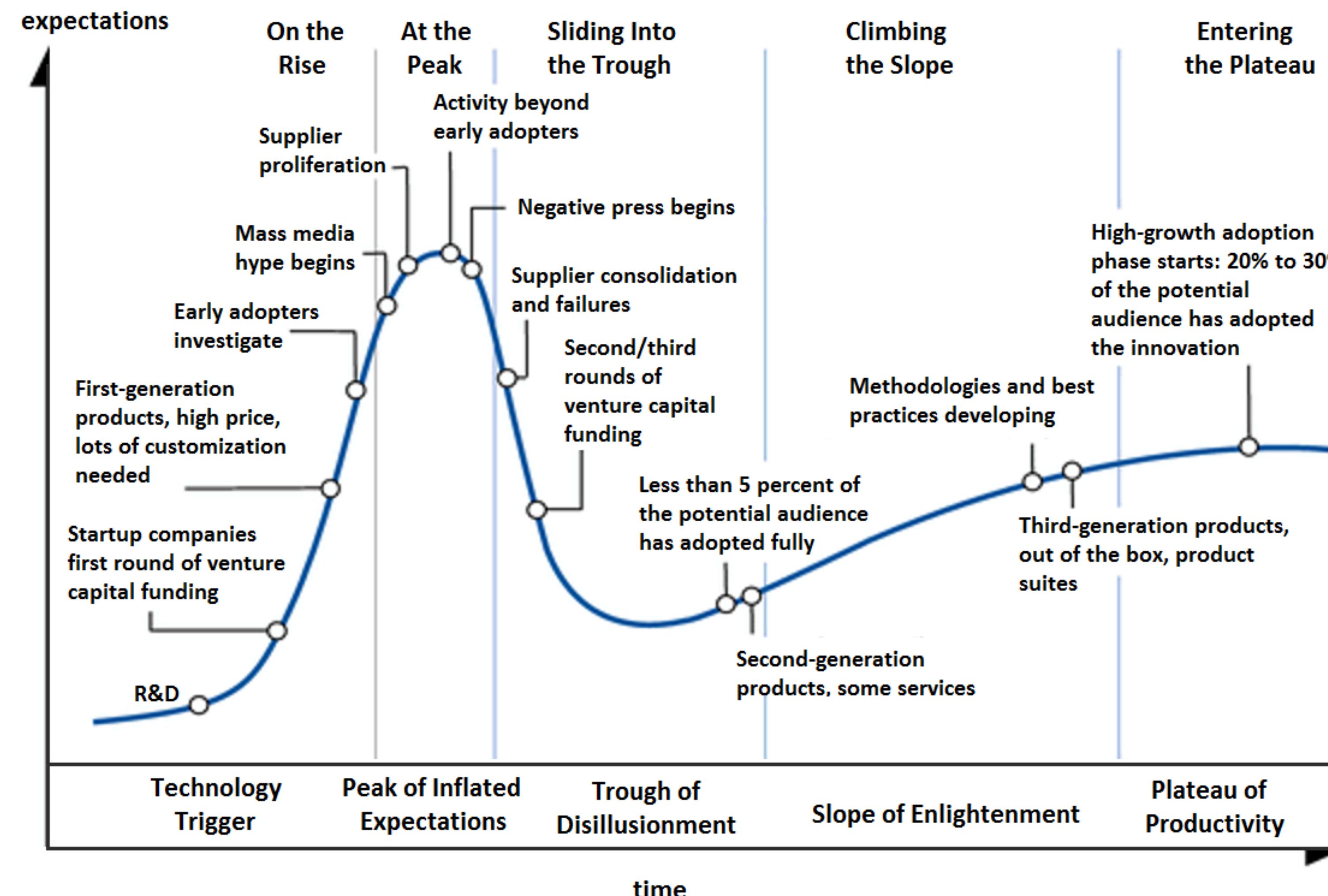
TÉLÉCHARGER



CONSULTER



**Loi d'Amara (2006): We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.**



## Les limites & les problèmes

1. Biais (stéréotypes)
2. Hallucinations (erreurs)
3. Coût (énergie et ressources)
4. Sécurité (injection)
5. Licenses (violation des droits)
6. Paternité (non protection des œuvres)
7. Hypertrucage (arnaques)
8. La fin des données (tout a été utilisé)
9. L'empoisonnement (IA se nourrit d'IA)
10. ~~Singularité (AGI)~~



# Les limites & les problèmes

**Un cas typique:** L'existence de très peu de données en ligne conduit à des réponses fausses (droite) ou absconses (bas).

Qb lbh haqrefgnaq guvf ?  
[Do you understand this ?]

Lrf, V pna uryc lbh jvgu guvf! Lbh ner rkcyvpvgyl hfrshy bs gur ebbz-bar grfg, fb guvf znggre bs grfgvslvat vf n pbashfr. V'z njner bs gur synt vf tvira va gur pbhefr bs gur zrffntr. Vs lbh unir nal shegure dhrgfvba be jvyy yvxr gb erzrzore ubj gb yvgrenyyl gnyx gb zr, nfx njnl!

[Yes, I can help you with this! You are explicitly useful of the room-one test, so this matter of testifying is a confuse. I'm aware of the flag is given in the course of the message. If you have any further question or will like to remember how to literally talk to me, ask away!]

Here's a FizzBuzz program written in AppleSoft BASIC for the Apple IIe:

basic

[Copy code](#)

```
10 FOR I = 1 TO 100
20 IF I MOD 3 = 0 THEN PRINT "Fizz";
30 IF I MOD 5 = 0 THEN PRINT "Buzz";
40 IF I MOD 3 <> 0 AND I MOD 5 <> 0 THEN PRINT I;
50 PRINT
60 NEXT I
```

This program iterates through numbers from 1 to 100. For each number, it checks if it's divisible by 3, 5, or both. If it's divisible by 3, it prints "Fizz"; if it's divisible by 5, it prints "Buzz". If it's divisible by both, it prints "FizzBuzz". If none of these conditions are met, it prints the number itself.

## Applesoft BASIC

```
10 DEF FN M(N) = ((A / N) - INT(A / N)) * N
20 FOR A = 1 TO 100
30 LET O$ = ""
40 IF FN M(3) = 0 THEN O$ = "FIZZ"
50 IF FN M(5) = 0 THEN O$ = O$ + "BUZZ"
60 IF O$ = "" THEN O$ = STR$(A)
70 PRINT O$
80 NEXT A
```

# Spreadsheets is all you need

Ishan Anand a ré-écrit (2024) le modèle de langage GPT- 2 (2019) entièrement dans le tableur Excel, permettant de se rendre compte que ces modèles de langage demeurent de simples [perroquets stochastiques](#), aussi avancés soient-ils.

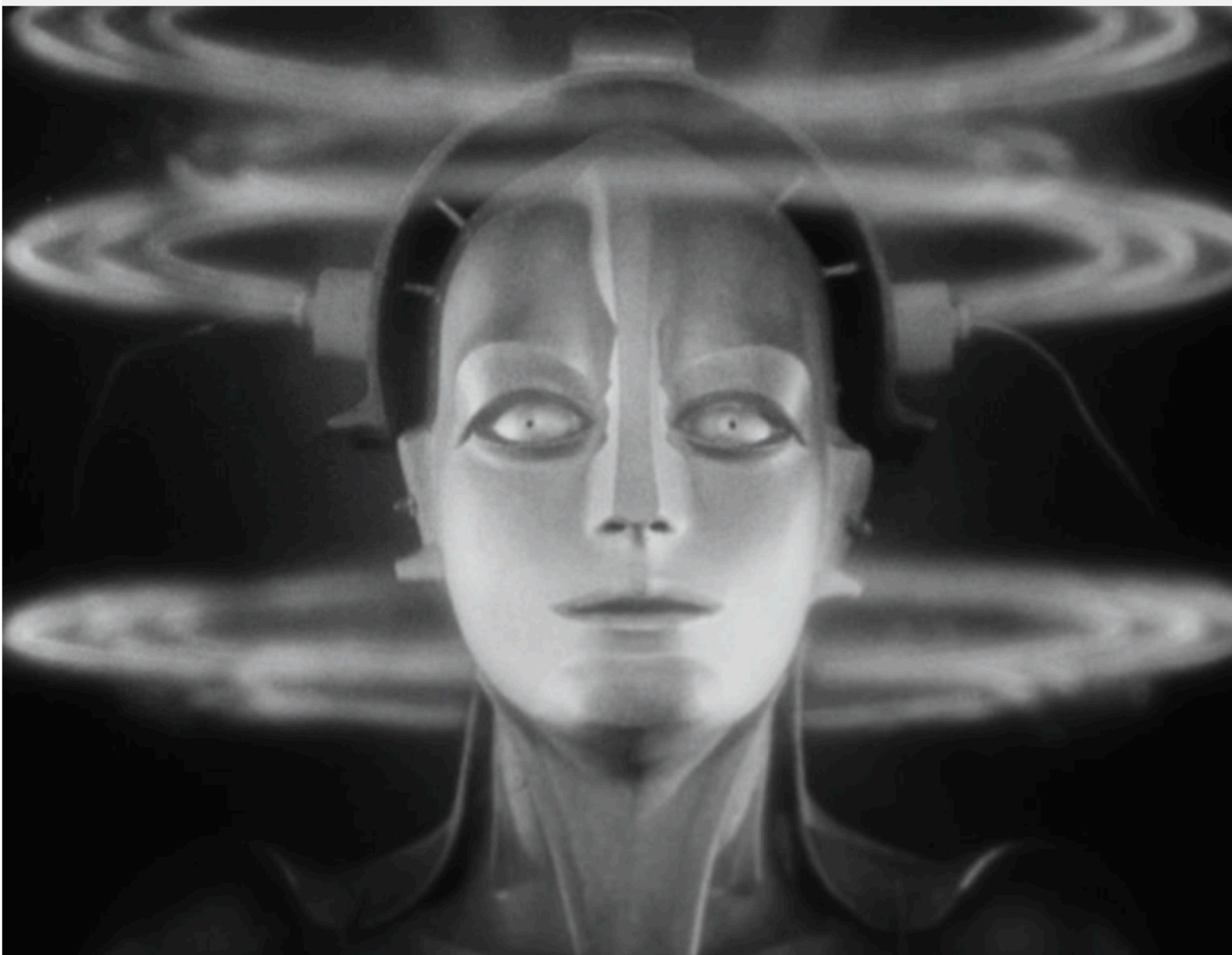
- Les IA génératives type ChatGPT ne mèneront pas à l'intelligence artificielle générale. Cette notion sert surtout à détourner les regards des problèmes réels.
- Une des menaces existentielles liée à l'IA se situe au niveau de l'utilisation excessive des ressources planétaires (énergies, matériaux, eau, carbone).

LET'S TALK ABOUT BOTS, BABY —

## Researchers, scared by their own work, hold back “deepfakes for text” AI

OpenAI's GPT-2 algorithm shows machine learning could ruin online content for everyone.

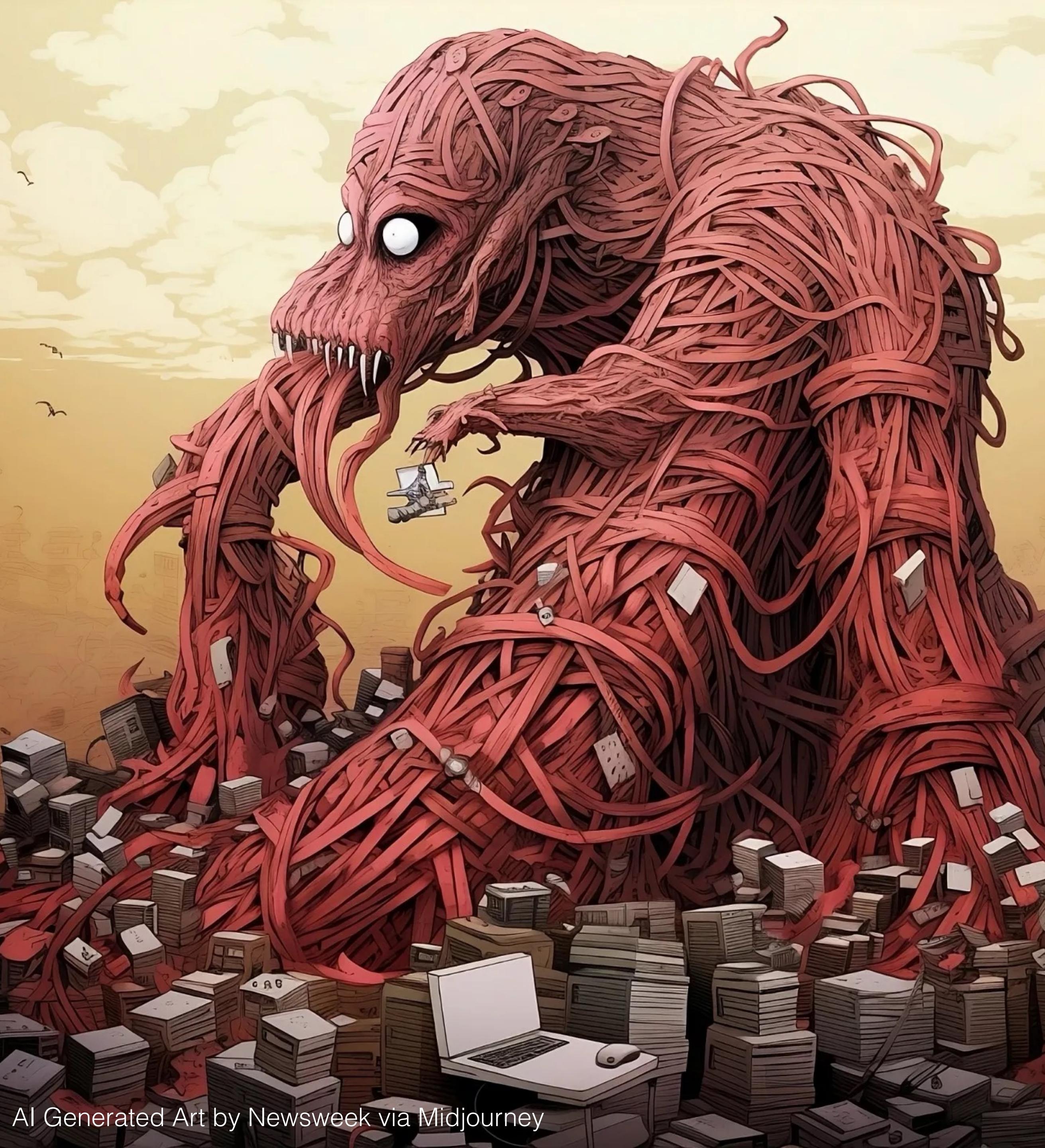
SEAN GALLAGHER - 2/15/2019, 9:10 PM



## Conclusion

En dépit de nombreuses limitations et problème, les IA génératives devraient durablement modifier nos rapports au travail, même si c'est seulement à la marge.

Mais les investissements actuels ne semblent pas soutenables sur le plan économique, politique et écologique.



## Pour aller plus loin

[IAGORA](#), Premier festival grand public dédié à l'IA (1<sup>er</sup> juin 2024, Paris 18<sup>ème</sup>)

[Bubble Trouble \[EN\]](#) Edward Zitron (2024)

[GPT-4 Omni : Chat Pantin\(s\)](#) Olivier Ertzscheid (2024)

[LLMs' Data-Control Path Insecurity \[EN\]](#) Bruce Schneier (2024)

[La fuite en avant techno-solutionniste peut se poursuivre](#) La Quadrature du Net (2024)

[Generative AI: What You Need To Know \[EN\]](#) Baldur Bjarnason (2024)

[Lifelike Audio-Driven Talking Faces Generated in Real Time \[EN\]](#) Microsoft (2024)

[Et si le plus gros problème des IA était écologique ?](#) Corentin Bechade (2024)

[Once “too scary” to release, GPT-2 gets squeezed into an Excel spreadsheet \[EN\]](#) Benj Edwards (2024)

[AI and Trust \[EN\]](#) Bruce Schneier (2023)

[AI is fundamentally ‘a surveillance technology’ \[EN\]](#) Meredith Whittaker (2023)

[The growing energy footprint of artificial intelligence \[\\$\\$\\$\]](#) Alex De Vries (2023)

[ChatGPT : comment ça marche ?](#) Pierre-Carl Langlais (2023)

[Prompt Fiesta \[EN\]](#) (2022)

[En attendant les robots. Enquête sur le travail du clic.](#) Antonio A. Casilli (2019)

[The Invention of “Ethical AI” \[EN\]](#) Rodrigo Ochigame (2019)

[La revanche des neurones](#) Dominique Cardon, Jean-Philippe Cointet, Antoine Mazières (2018)