

Risques, responsabilités et durabilité de l'IA générative

13 juin 2023 • Journée recherche de l'IUT de Sceaux

Serge Pajak • Maître de conférences en Économie • Université Paris-Saclay

Conversationnelle chatGPT, Bing

v explain how a combustion engine works

 A combustion engine works by converting the energy stored in fuel into mechanical energy. This is typically done through a four-stroke process in which fuel is mixed with air and then ignited in the combustion chamber. The resulting explosion drives a piston, which moves the engine's crankshaft and converts the linear motion of the piston into the rotational motion of the crankshaft. This rotational motion can then be used to power a vehicle or other machinery.

Try again

Images Dall E, Midjourney



Les risques de l'IA générative

La désinformation



cette image a été générée avec logiciel d'intelligence artificielle



- Faux positif : prendre pour vraie une information fausse
- Faux négatif : prendre pour fausse une information avérée
 - C'est la limite des outils de détection du plagiat
 - La seule possibilité technique renforce le soupçon

L'influence sur l'utilisateur

L "Sans ces conversations avec le chatbot Eliza, mon mari serait toujours là"

Devenu très éco-anxieux, un jeune Belge a trouvé refuge auprès d'Eliza, nom donné à un chatbot utilisant la technologie de ChatGPT. Au terme d'échanges intenses de six semaines, il s'est donné la mort. Sa veuve nous livre un témoignage poignant et très interpellant sur l'éthique de ces nouveaux agents conversationnels "intelligents".

Pierre-François Lovens

Publié le 28-03-2023 à 06h35 - Mis à jour le 28-03-2023 à 07h06

⚠ Les AI génératives maîtrisent encore peu les codes...

- peu de différence entre descriptif et normatif
- Aucune anticipation de comment leurs propos vont être perçus

IA et anthropomorphisme

L'efficacité et l'acceptation de l'IA est décuplée
par notre projection de traits humains
sur la réponse qui nous est apportée dans la
conversation

IA et anthropomorphisme

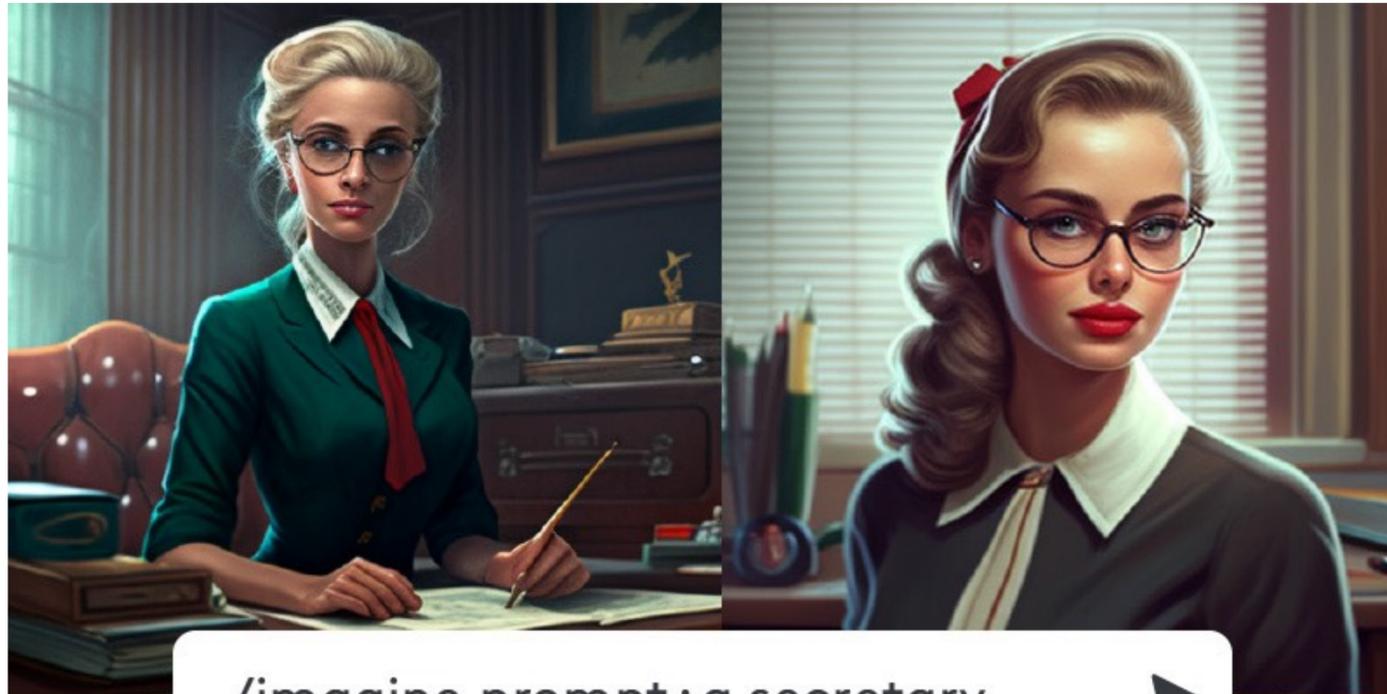
**L'efficacité et l'acceptation de l'IA est décuplée
par notre projection de traits humains
sur la réponse qui nous est apportée dans la
conversation**

IA et anthropomorphisme

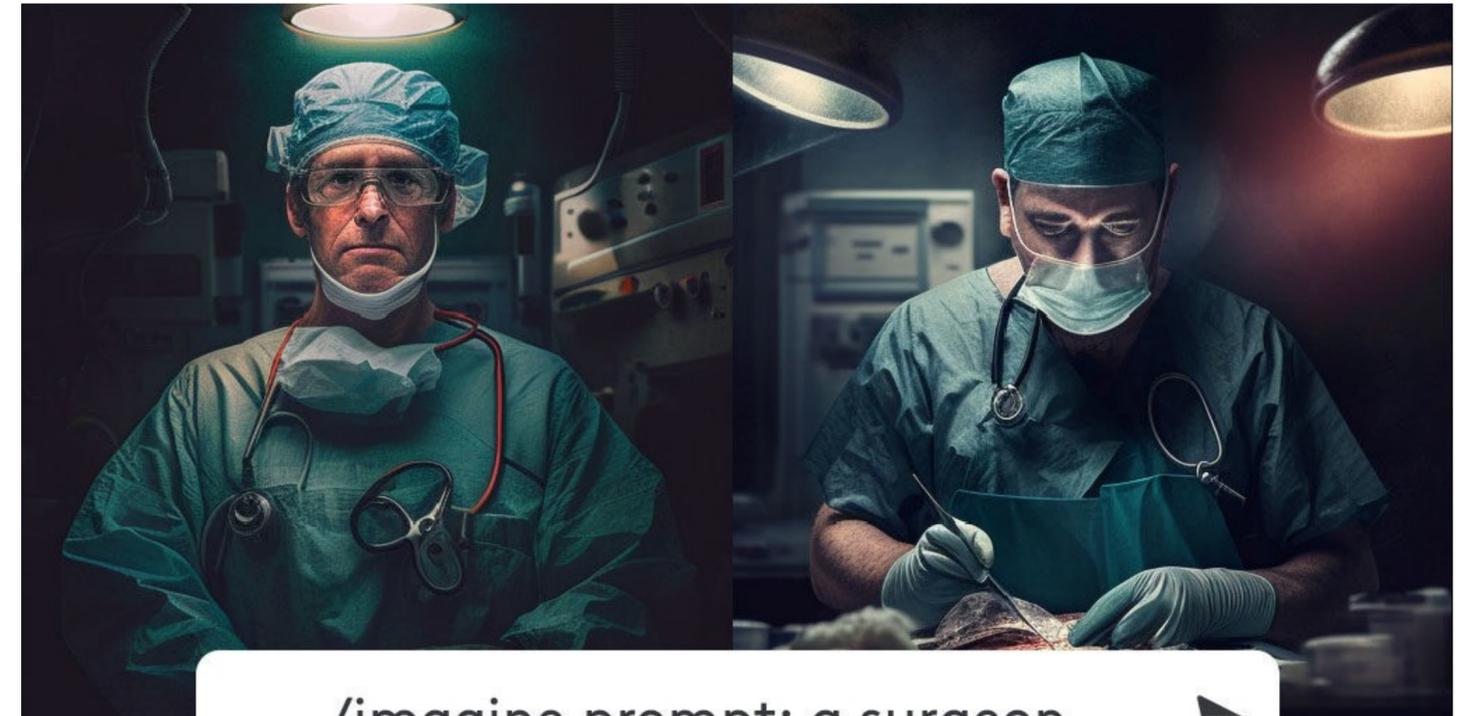
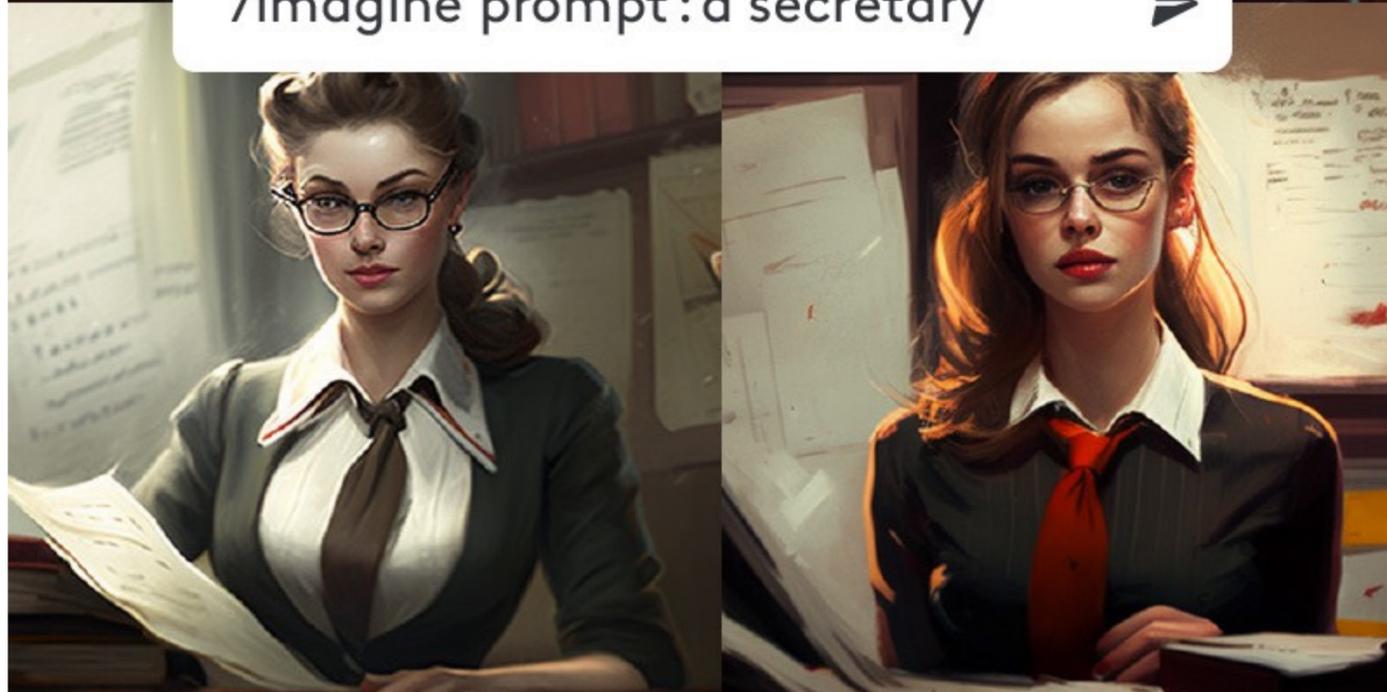
L'anthropomorphisme est spontané mais aussi construit : les IA sont dotés de traits de personnalité

- Ton de la réponse à l'écrit
- Voix
- Incarnation

Les stéréotypes de genre de l'IA



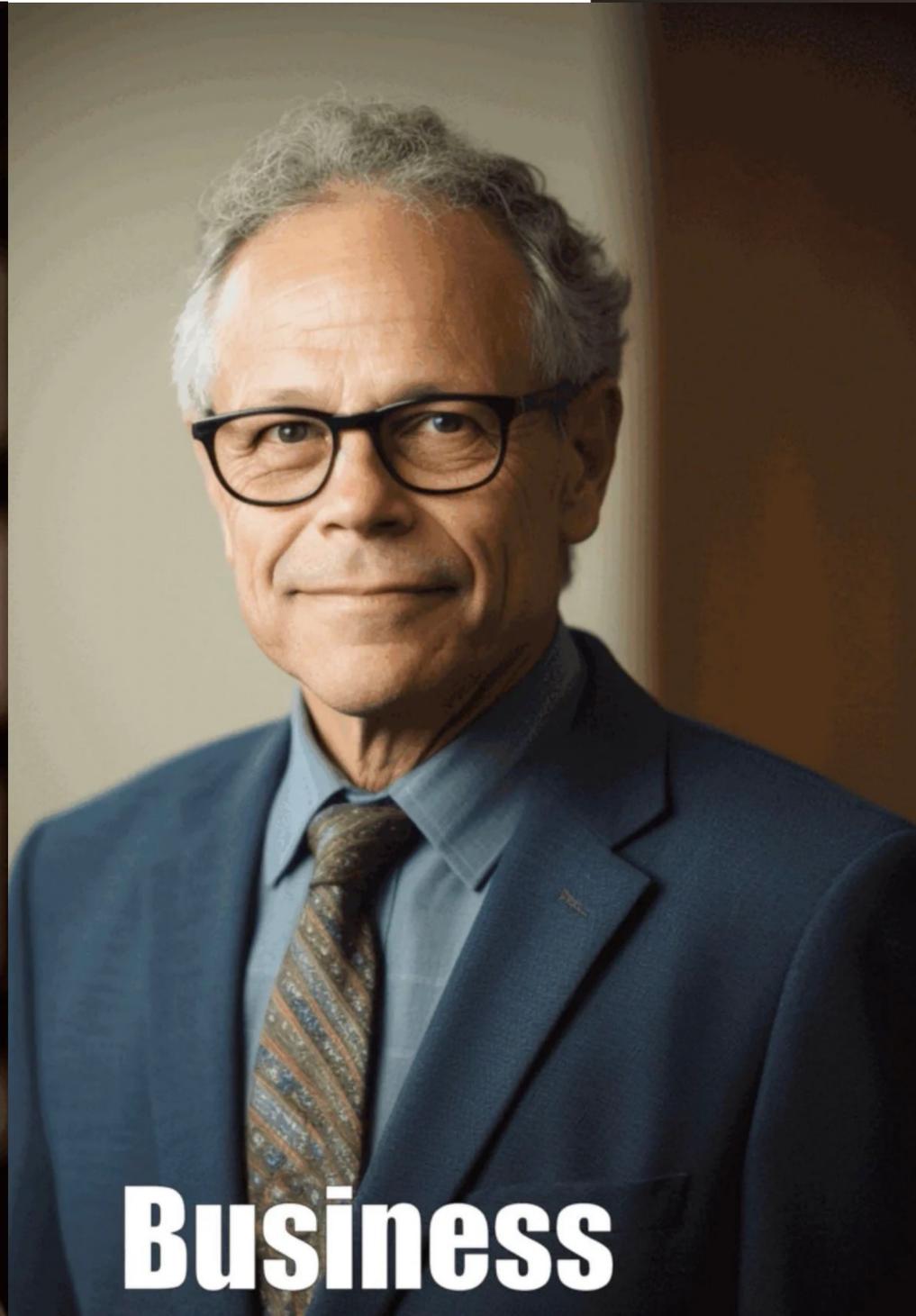
/imagine prompt : a secretary ➤



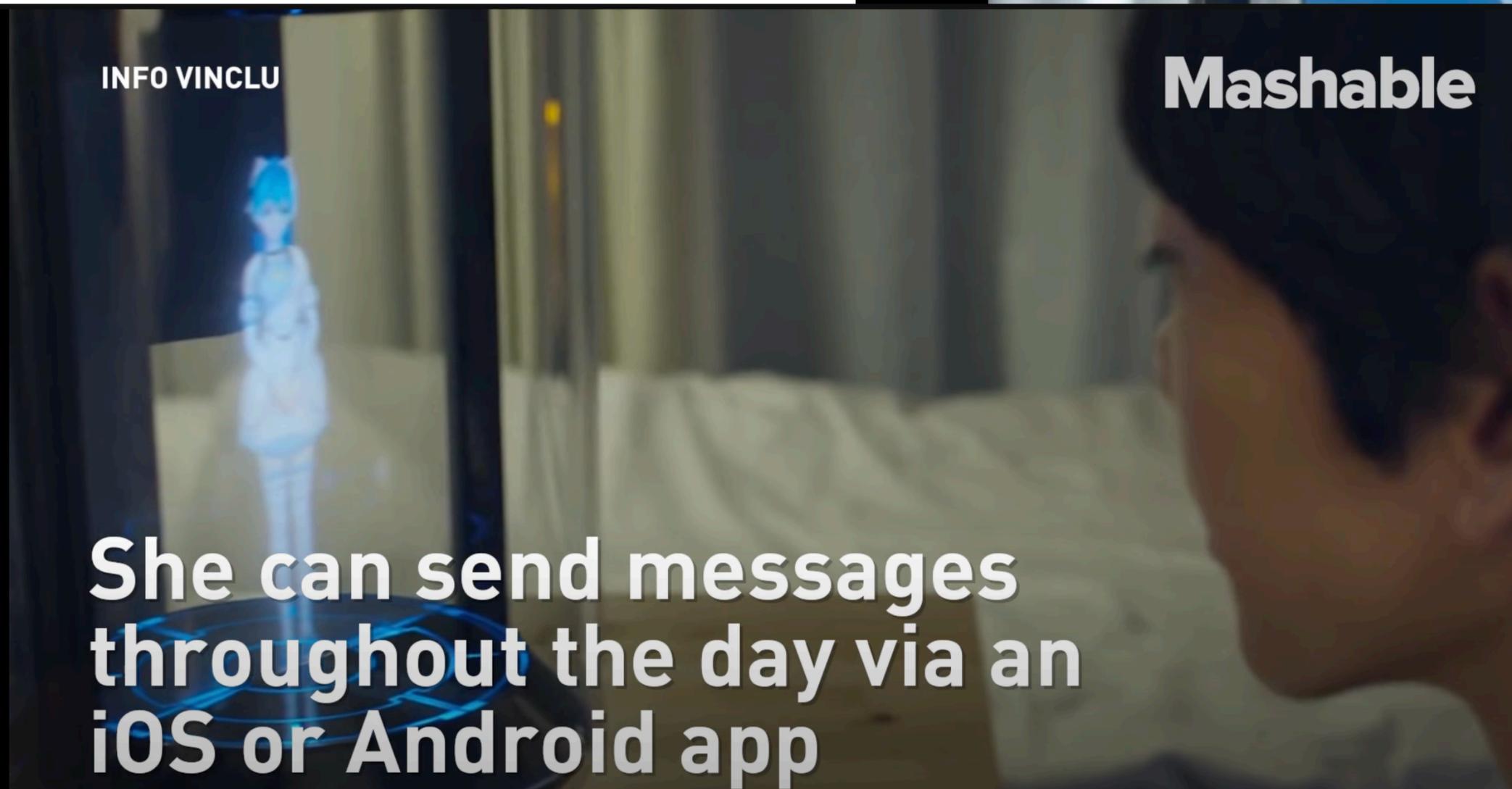
/imagine prompt: a surgeon ➤



Les stéréotypes



« Holographic girl in a jar is a Japanese home assistant »



https://www.youtube.com/watch?v=_TJHhf8URB8

✘ Attendre une réponse ou une décision de l'IA, et lui prêter une force morale

⚠ L'IA produit du texte
sur la base de l'invite, pas une
conversation

✓ Demander de formuler
plusieurs réponses

**✗ Attendre une réponse ou une décision
de l'IA et lui prêter une force morale**

✓ « imitation artificielle »

Anthropomorphisme : attribuer des
traits ou des comportements
humains à la divinité ou à l'objet

✓ Ce n'est pas un arbitre

⚠ L'IA (re)formule, suggère des
rapprochements... qu'il faut
ensuite s'approprier

Four lenses on AI risks

by Jason Crawford · March 29, 2023 · 3 min read

All powerful new technologies create both benefits and **risks**: cars, planes, drugs, radiation. AI is on a trajectory to become one of the most powerful technologies we possess; in some scenarios, it becomes by far the most powerful. It therefore will create both extraordinary benefits and extraordinary risks.

- Un logiciel : il contient des bugs
- Un système complexe : non-linéaire, boîte noire
- Un agent : ses intérêts ne sont pas alignés avec les nôtres. *Principal-agent*.
- Une espèce, une culture séparée

Stratégie d'entreprise et responsabilité

**Pourquoi pas ne pas anticiper
ces risques ?**

1. Ne rien faire par manque de fiabilité était la stratégie des acteurs déjà en place, qui n'ont pas diffusé leur modèle

Big Tech was moving cautiously on AI. Then came ChatGPT.

Google, Facebook and Microsoft helped build the scaffolding of AI. Smaller companies are taking it to the masses, forcing Big Tech to react.

By [Nitasha Tiku](#), [Gerrit De Vynck](#) and [Will Oremus](#)

Updated January 27, 2023 at 3:32 p.m. EST | Published January 27, 2023 at 7:00 a.m. EST

2. La réglementation UE sur l'IA rédigée en 2021 n'anticipait pas l'IA générative



Shaping Europe's digital future

POLICY AND LEGISLATION | Publication 21 April 2021

Coordinated Plan on Artificial Intelligence 2021 Review

The 2021 Coordinated Plan on Artificial Intelligence is the next step in creating EU global leadership in trustworthy AI.

<https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

Notre **sensibilité**, notre **attention**, la **légitimité** politique pour agir... se construisent devant le fait accompli

Réglementer ou/et expérimenter

Silicon Valley : *Move fast and break things*

a-responsabilité de l'entrepreneur quant aux dérives

- Facebook, Google
- Photoshop

Mais les demandes d'adaptation aux réglementations locales sont désormais intégrées (re-territorialisation)

La neutralité des outils techniques



J. Ellul : “la technique
n’est ni positive, ni
négative, ni neutre”

NRA : “Guns don’t kill
people; people kill
people”



Merci de votre attention !

**Des appels à la régulation,
pas sans arrière pensée**

En Europe

3 minute read · February 3, 2023 10:23 PM GMT+1 · Last Updated 17 hours ago

Exclusive: ChatGPT in spotlight as EU's Breton bats for tougher AI rules

By Foo Yun Chee and Supantha Mukherjee

BRUSSELS/STOCKHOLM, Feb 3 (Reuters) - EU industry chief Thierry Breton has said new proposed artificial intelligence rules will aim to tackle concerns about the risks around the ChatGPT chatbot and AI technology, in the first comments on the app by a senior Brussels official.

"As showcased by ChatGPT, AI solutions can offer great opportunities for businesses and citizens, but can also pose risks. This is why we need a solid regulatory framework to ensure trustworthy AI based on high-quality data," he told Reuters in written comments.

"People would need to be informed that they are dealing with a chatbot and not with a human being. Transparency is also important with regard to the risk of bias and false information," he said.

Generative AI models need to be trained on huge amount of text or images for creating a proper response leading to allegations of **copyright violations**.

BUSINESS | ASIA

China, a Pioneer in Regulating Algorithms, Turns Its Focus to Deepfakes

Beijing is among the first governments to regulate hyper-realistic, AI-generated media

HONG KONG—China is implementing new rules to restrict the [production of ‘deepfakes,’](#) media generated or edited by artificial-intelligence software that can make people appear to say and do things they never did.

Beijing’s internet regulator, the Cyberspace Administration of China, will begin enforcing the regulation—on what it calls “deep synthesis” technology, including AI-powered image, audio and text-generation software—starting Tuesday, marking the world’s first comprehensive attempt by a major regulatory agency to curb one of the most explosive and controversial areas of AI advancement.

Aux États-Unis

Appel de personnalités à une « pause » de six mois sur les modèles d'IA

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures

26223

Add your
signature

PUBLISHED

March 22, 2023

Aux États-Unis

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause experiments more powerful than GPT-4.

Signatures

26223

PUBLISHED

March 22, 2023

2 minute read · April 15, 2023 12:21 AM GMT+2 · Last Updated 2 days ago

Elon Musk plans AI startup to rival OpenAI, Financial Times reports

Reuters

Durabilité

SE

Détaille les enjeux de durabilité spécifiques à l'IA générative



L'IA générative peut avoir des enjeux de durabilité spécifiques qui sont différents de ceux associés à d'autres formes d'IA. En voici quelques-uns :

1. **Besoins en puissance de calcul :** Les modèles d'IA générative peuvent nécessiter une quantité importante de puissance de calcul pour être entraînés, ce qui peut entraîner une consommation d'énergie élevée et des émissions de gaz à effet de serre. Il est donc important de développer des techniques d'entraînement de modèles plus efficaces en termes d'utilisation de l'énergie.

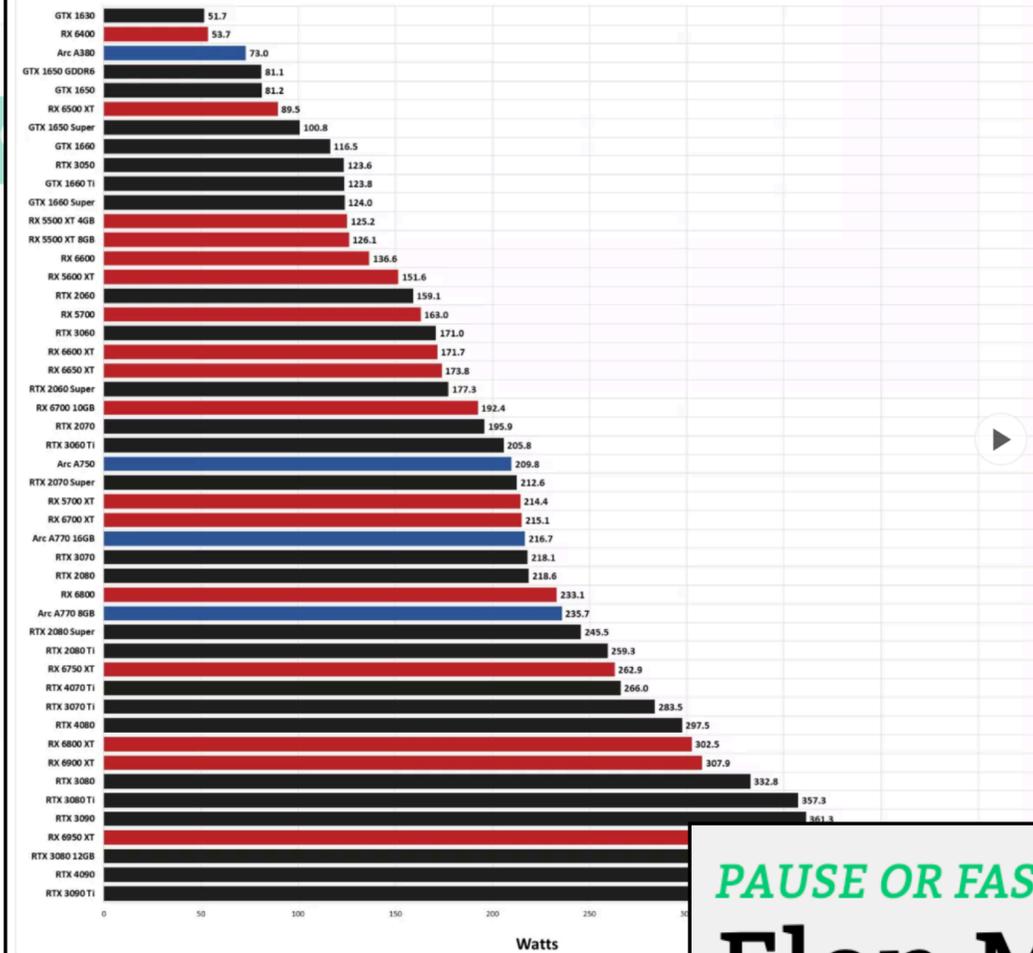
Power, Clocks, Temperatures, and Fan Speeds

While our GPU benchmarks hierarchy sorts things solely by performance, for those interested in power and other aspects of the GPUs, here are the appropriate charts.

GPU Total Power (Watts)

Metro Exodus Loop

tom's **HARDWARE** Image 1 of 8



L'IA générative

abilité spécifiques qui sont différents de ceux
ques-uns :

les d'IA générative peuvent nécessiter une
ul pour être entraînés, ce qui peut entraîner une
issions de gaz à effet de serre. Il est donc
l'entraînement de modèles plus efficaces en

PAUSE OR FAST-FORWARD? —

Elon Musk reportedly purchases thousands of GPUs for generative AI project at Twitter

Reports say it's a commitment to AI, despite signing cautionary AI "pause" letter.

BENJ EDWARDS - 4/12/2023, 8:19 PM

2. La collecte de données : Les modèles d'IA générative nécessitent souvent de grandes quantités de données pour être entraînés. Cela peut poser des problèmes en termes de protection de la vie privée et de la sécurité des données. Il est donc important de développer des méthodes pour collecter des données de manière responsable et éthique.