

The five walls of AI, five years later

BERTRAND BRAUNSCHWEIG

The Five Walls

Trust 

Energy 

(cyber)Security 

Human-Computer Interaction



Inhumanity 



Trust



The three dimensions of trust

Techno	Interactions	Social
Reliability		
Robustness	Transparency	Fairness
Compliance	Explainability	Privacy
Precision	Responsibility	Diversity & Inclusion
Safety	Monitoring and control	Sustainability
Security		

Energy

Three Mile Island is reopening and selling its power to Microsoft

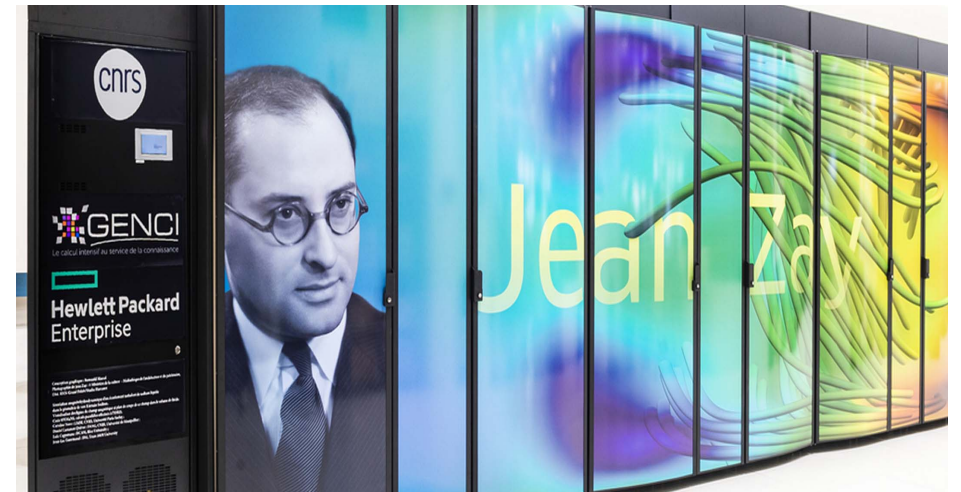


By Jordan Valinsky, CNN

🕒 3 minute read · Updated 12:57 PM EDT, Fri September 20, 2024



Three Mile Island, which closed in 2019, is soon reopening. Andrew Caballero-Reynolds/AFP/Getty Images



Cybersecurity



+ .007 ×



=



“panda”

57.7% confidence

noise

“gibbon”

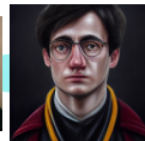
99.3% confidence

Midjourney generations over time: “a hyper-realistic image of Harry Potter”

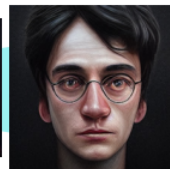
Source: [Midjourney, 2023](#)



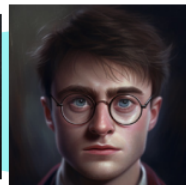
V1, February 2022



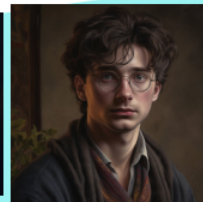
V2, April 2022



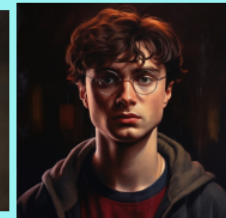
V3, July 2022



V4, November 2022



V5, March 2023



V5.1, March 2023



V5.2, June 2023



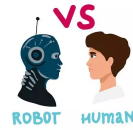
V6, December 2023

Human-Computer Interactions

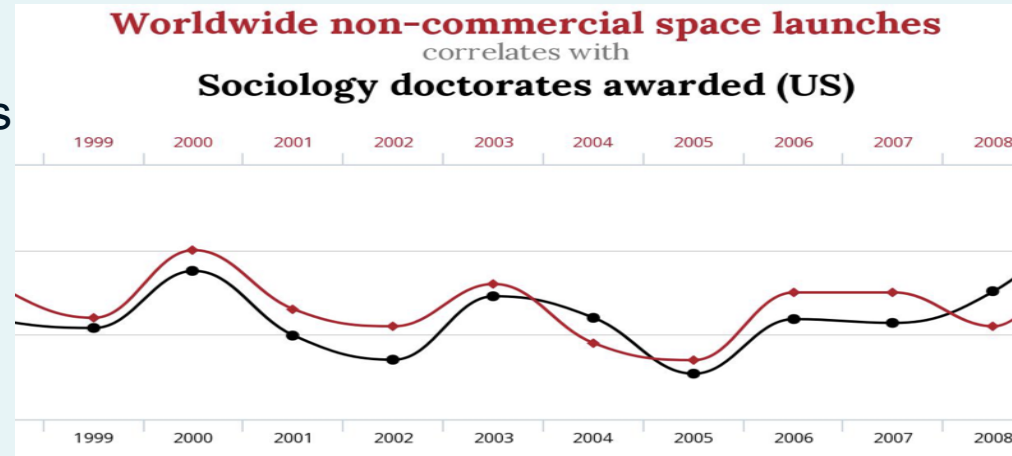


- Dialogue (*chatbots*)
- Shared problem solving and decision-making
- Sharing space and resources (cohabitation with robots that are ignored or given orders);
- Task sharing (teammate robot).

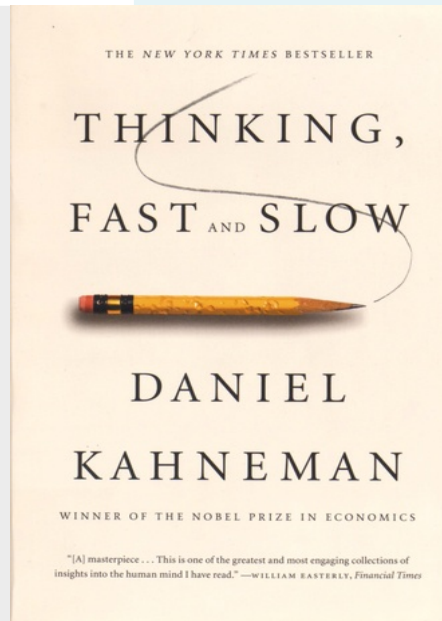
Inhumanity



Causality-correlations



System1 –
System2



Common sense

Alexa a dit à une enfant de mettre ses doigts dans la prise

« Le challenge est simple » 😬

🕒 Temps de lecture : 2 min

 Marie Turcan



Five years later

Trust →

Energy ↓

(cyber)Security ↓

Human-Computer Interaction ↑

Inhumanity ↑

Trust →

- European strategy for AI (and many other countries)
- Many R&D programs
 - Confiance.ai & others



- Theme of AI Action Summit

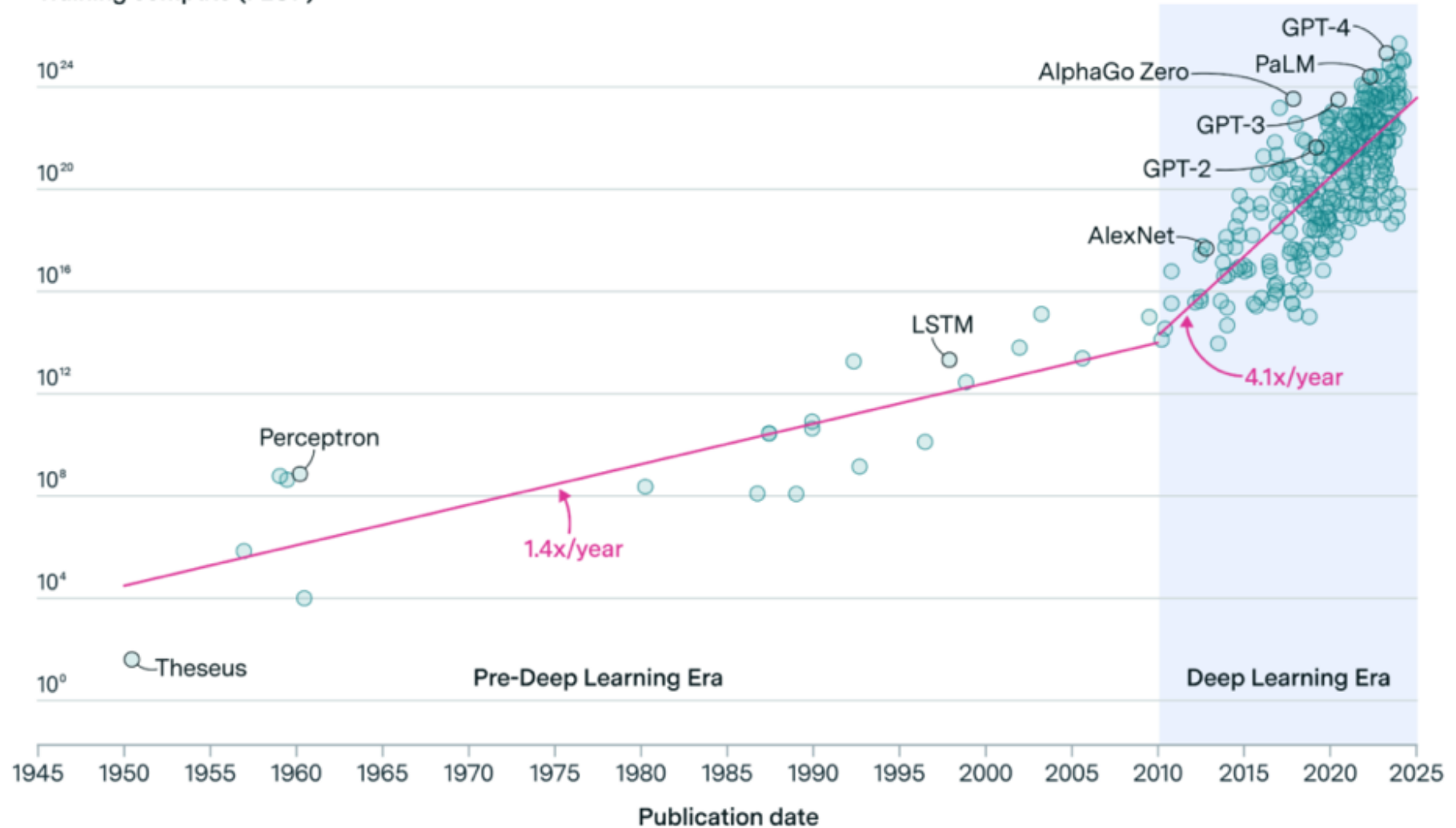
- But many problems remain including for foundation models & generative AI
 - Robustness, hallucinations, etc.

Energy



Training Compute of Notable Machine Learning Systems Over Time

Training compute (FLOP)

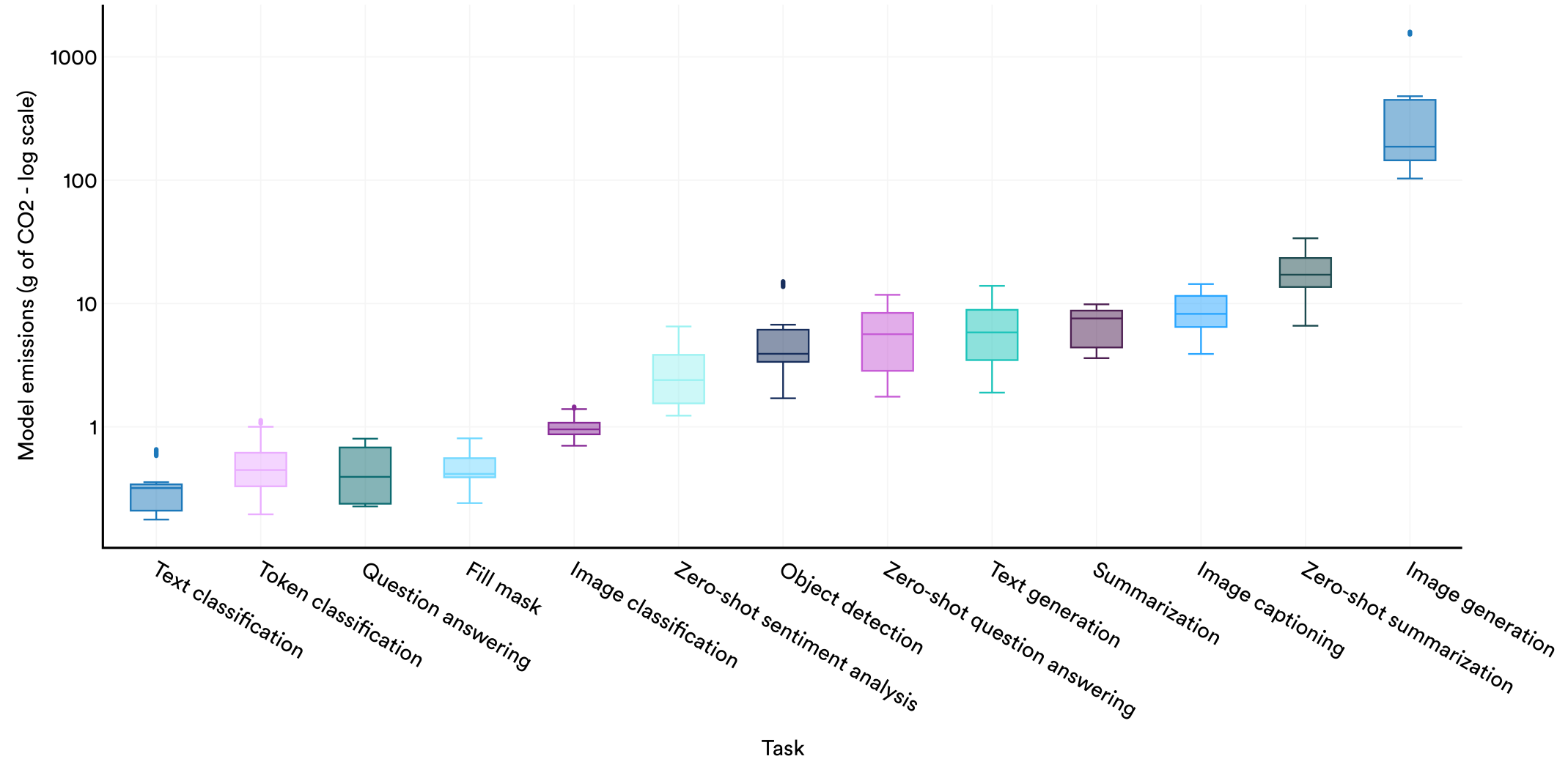


Energy



Carbon emissions by task during model inference

Source: Luccioni et al., 2023 | Chart: 2024 AI Index report



(cyber)Security ↓

- Still the same issues
 - Deep fakes
 - Adversarial attacks
 - Model & data stealing
 - Etc.
- New issues with foundation models & generative AI
 - Intellectual property
 - Privacy
 - Harmful content generation
 - Etc.

Human-Computer Interaction



- Natural language interface !
 - + speech recognition
 - + others (emotions, some explainability, HCI in general ...)






- But be careful ! Many issues remain
 - Think of Level 4 autonomous driving

Inhumanity



- Causality vs correlation
 - LLMs seem to produce causal reasoning (but not like us)
- Common sense
 - LLMs learn common sense notions (but not like us)
- System 1 / System 2 processing
 - Different approaches are developed (but not ...)

Conclusion

- Trust 
- Energy 
- (cyber)Security 
- Human-Computer Interaction 
- Inhumanity 

What do you think?
Any other wall for AI?