

Intelligence Artificielle: passé, présent, futur

BERTRAND BRAUNSCHWEIG



Thinkerview



Au menu

- **Une brève histoire de l'IA**
- Les capacités actuelles de l'IA
- La réglementation européenne et la question de la confiance
- Perspectives - Les 5 murs de l'IA, 5 ans après
- Conclusion

Contributions empruntées à AI Index, Yann LeCun, Yoshua Bengio, Confiance.ai, Cédric Villani, DeepMind, OpenAI, Anthropic, Suno.ai, William Boyd ... entre autres!

Une brève histoire de l'IA

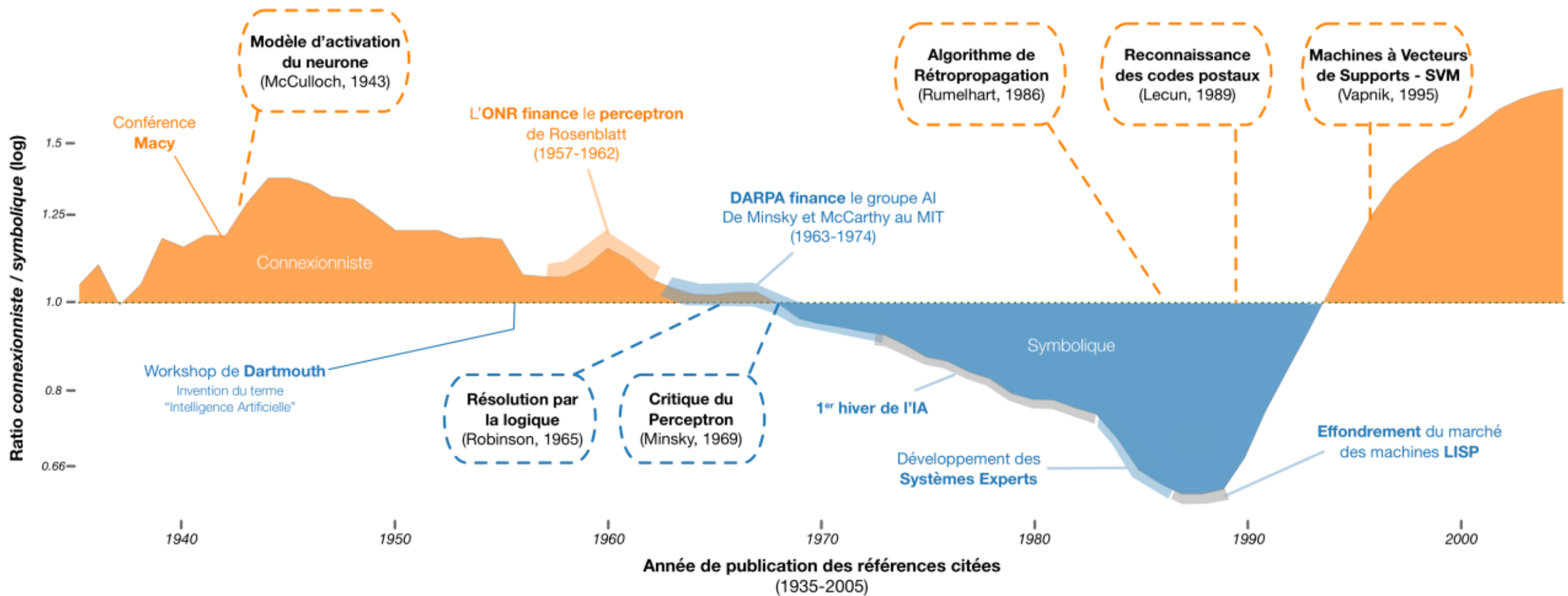
- Avant 1956: Alan Turing, neurone formel, robots ...
- 1956: congrès de Dartmouth, définition de l'IA
- Années 60: résolution de problèmes, langage naturel, jeux
- 1968: 2001, l'Odyssée de l'Espace, HAL
- 1969: « Perceptrons », arrêt des réseaux neuronaux
- 1973: Lighthill Report, arrêt de l'IA au Royaume-Uni, **premier hiver**
- Années 80: systèmes experts, et Chinook (dames)
- Années 90: **deuxième hiver**, mais Deep Blue (échecs) et premiers réseaux convolutifs
- 2000: premières applications pour le web
- 2010: Apprentissage profond
- 2017 : GANs, Transformers, LLMs, diffusion ...
- 2025: pas encore le **troisième hiver**!

D. Cardon et coll. « La revanche des neurones »

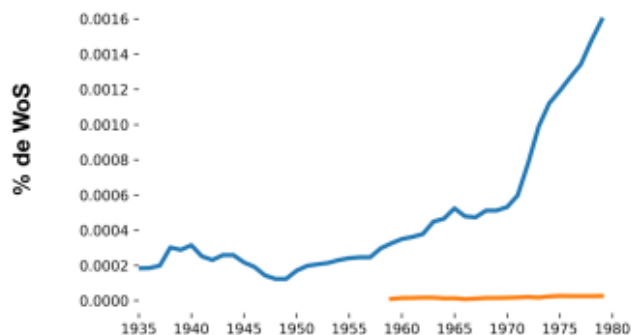
La Découverte/Réseaux, 2018

Figure 1. Machine hypothético-déductive (1) et machine inductive (2)

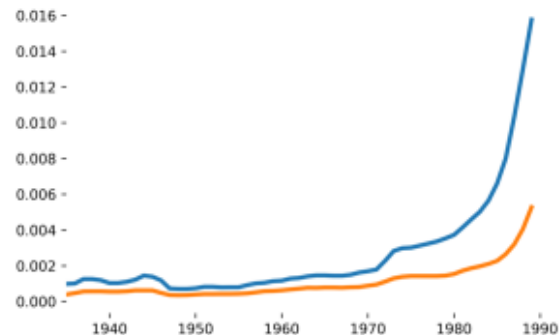




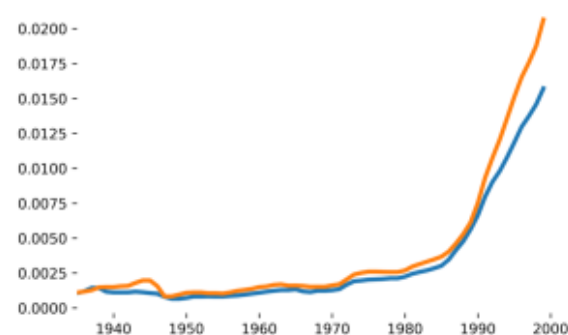
Citées entre 1980 et 1989



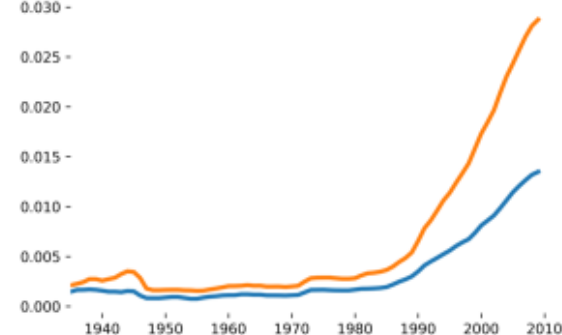
Citées entre 1990 et 1999



Citées entre 2000 et 2009

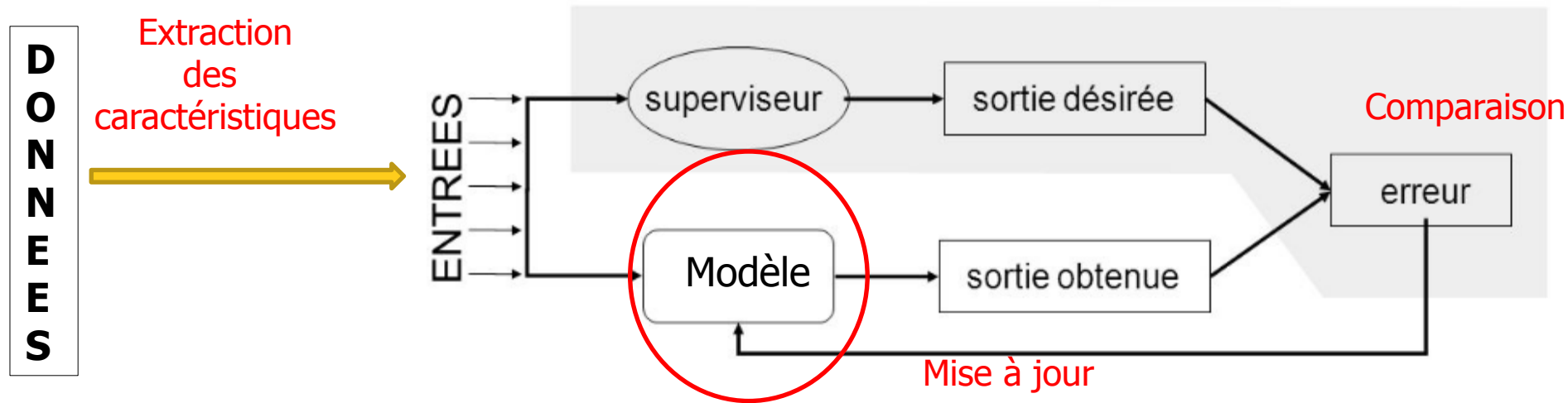


Citées entre 2010 et 2018



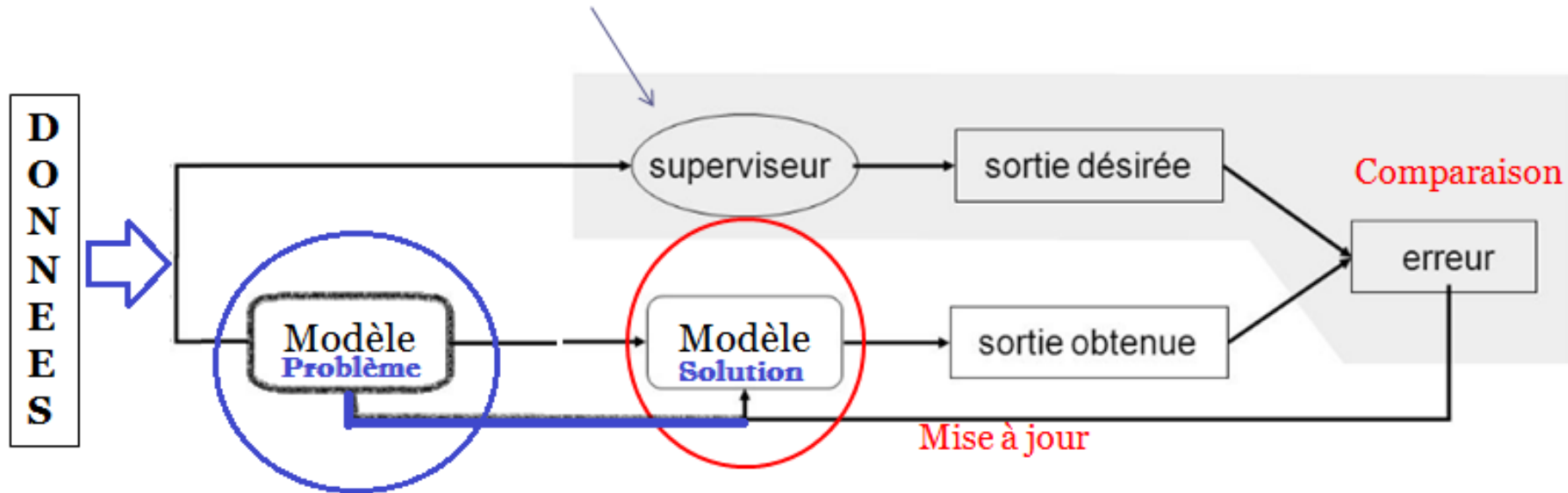
Apprentissage supervisé 1

- 1 - Humain: création a priori d'un modèle du problème
- 2 - Machine: création progressive d'un modèle de la solution

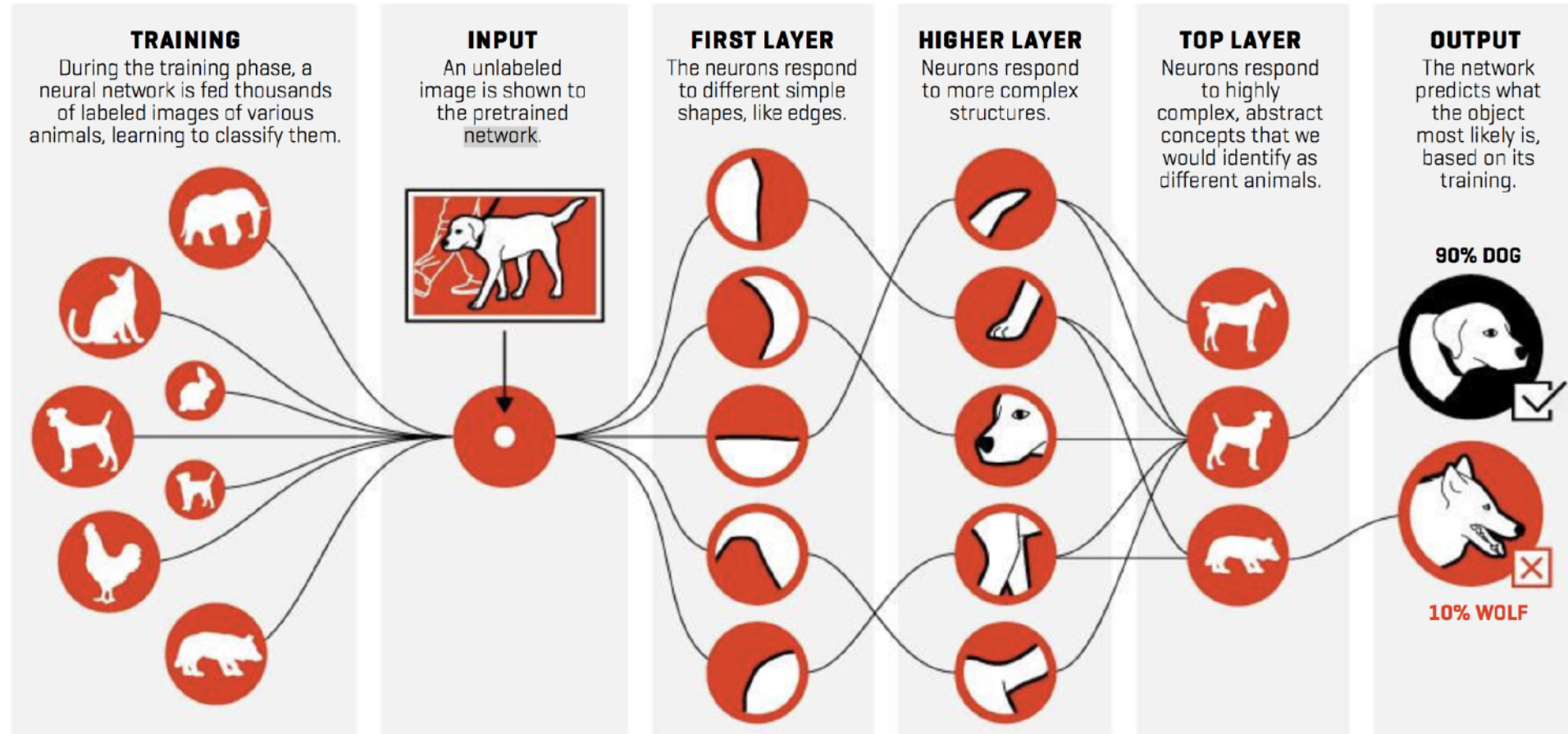


Apprentissage supervisé 2

Création progressive d'un modèle du problème et de la solution

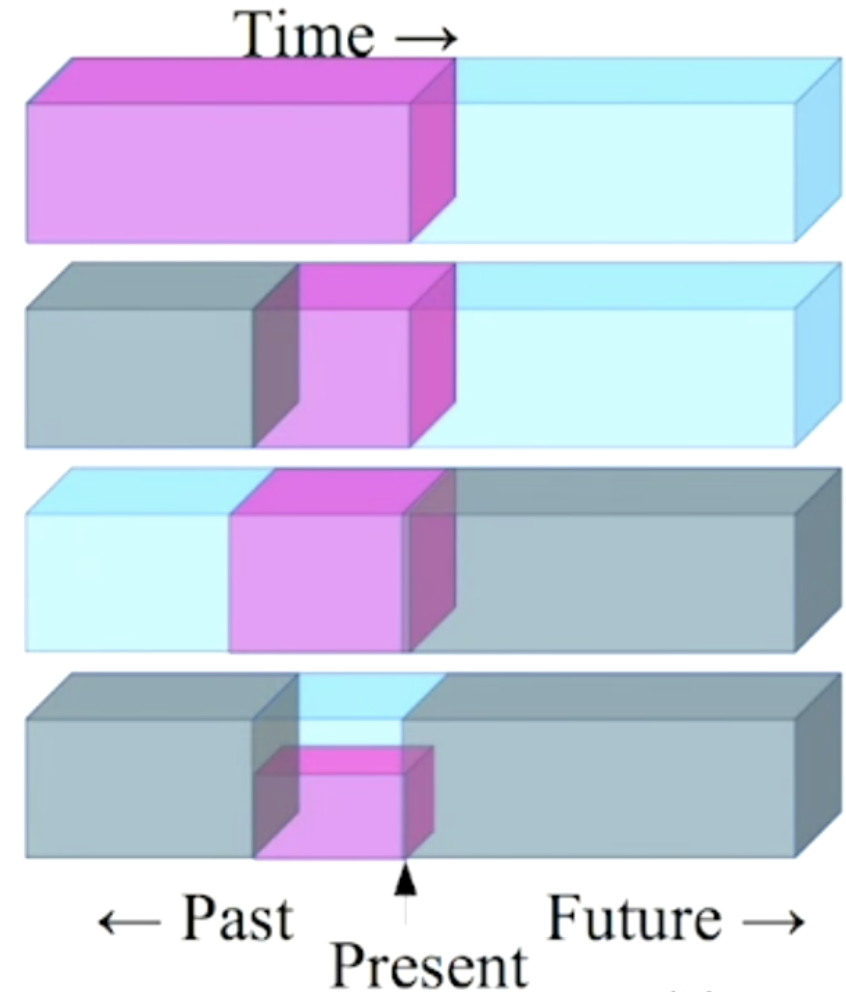


Apprentissage supervisé 3



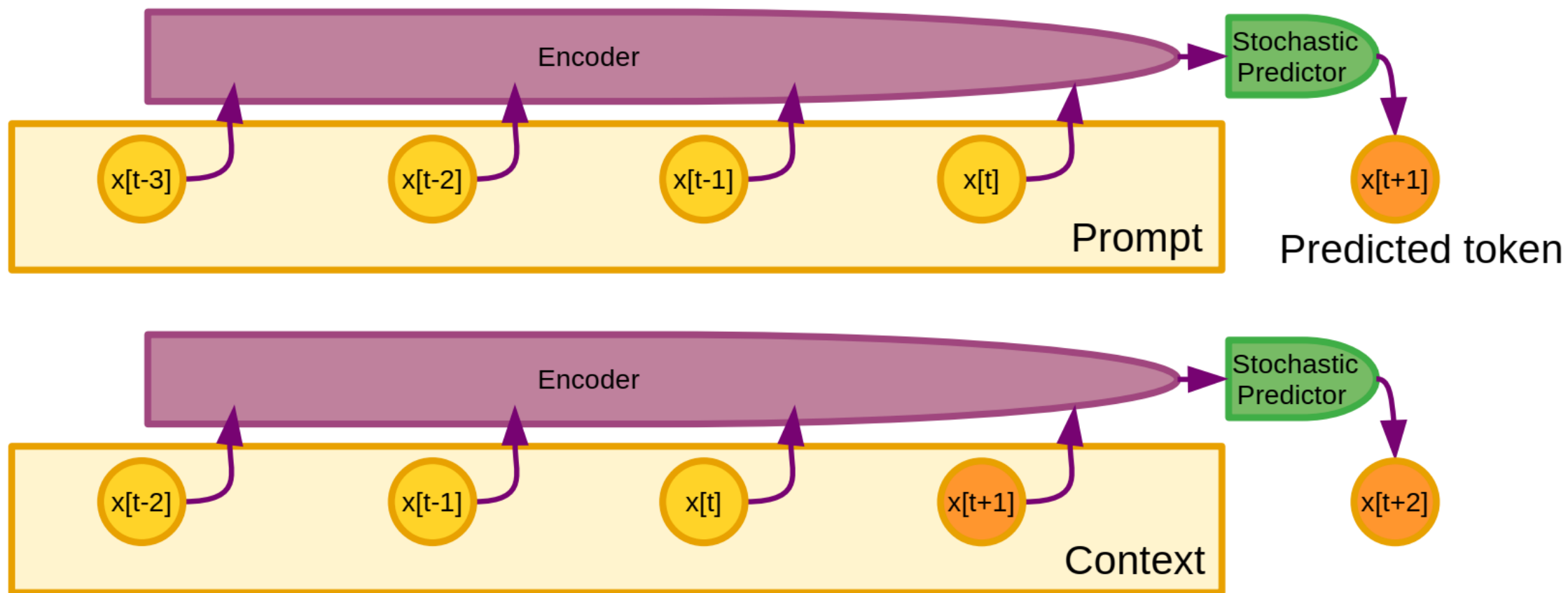
Apprentissage auto-supervisé

- ▶ Predict any part of the input from any other part.
- ▶ Predict the **future** from the **past**.
- ▶ Predict the **future** from the **recent past**.
- ▶ Predict the **past** from the **present**.
- ▶ Predict the **top** from the **bottom**.
- ▶ Predict the occluded from the visible
- ▶ **Pretend there is a part of the input you don't know and predict that.**



Auto-Regressive Generative Architectures

- ▶ Outputs one “token” after another
- ▶ Tokens may represent words, image patches, speech segments...



ChatGPT hallucine ...

Quelles sont les principales publications de Bertrand Braunschweig?

Bertrand Braunschweig est un nom assez courant, et il peut y avoir plusieurs personnes portant ce nom. Sans plus d'informations sur la personne spécifique dont vous parlez, il est difficile de fournir une liste précise de ses publications.

Cependant, si vous faites référence à Bertrand Braunschweig, le réalisateur et producteur français, voici quelques-unes de ses principales réalisations :

- "Les Lèvres rouges" (1971) - Un film d'horreur fantastique également connu sous le titre "Daughters of Darkness".
- "Histoires extraordinaires" (1968) - Un film à sketches basé sur des histoires d'Edgar Allan Poe, réalisé en collaboration avec les réalisateurs Roger Vadim et Louis Malle.
- "Le Voyage en douce" (1980) - Un drame romantique mettant en vedette Dominique Sanda et Geraldine Chaplin.
- "La Nuit de Varennes" (1982) - Un film historique qui se déroule pendant la Révolution française et met en scène une distribution d'acteurs renommés, dont Marcello Mastroianni et Hanna Schygulla.

Il est important de noter que cette liste est basée sur des informations disponibles jusqu'en septembre 2021, et il se peut que Bertrand Braunschweig ait publié d'autres travaux depuis lors.

Qui est le réalisateur du film "la nuit de varennes" (1982)?

Le réalisateur du film "La Nuit de Varennes" (1982) est Ettore Scola. Le film a été réalisé par ...

Petit calcul (à la louche)

- GPT4 : 1800 milliards de paramètres $1,8 \times 10^{12}$
 - Pour produire 1 token (une syllabe), $0,4 \times 10^{12}$ calculs environ
- Pour produire une page, 250 mots (500 tokens)
 - $0,4 \times 10^{12} \times 0,5 \times 10^3 = 2 \times 10^{14}$ calculs élémentaires
- ChatGPT a 700 millions d'utilisateurs mensuels
- Si 700 millions d'utilisateurs font une page chaque jour
 - $2 \times 10^{14} \times 7 \times 10^8 = 1,4 \times 10^{23}$ calculs chaque jour sur les machines de OpenAI
 - Soit cent quarante mille milliards de milliards de calculs élémentaires

Suite du petit calcul

- Le nouveau supercalculateur européen Jupiter a une puissance d'environ 1 exaFlop/s
 - 10^{18} opérations par seconde (et 13.1 MW)



Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
4	JUPITER Booster - BullSequana XH3000, GH Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, RedHat Enterprise Linux, EVIDEN EuroHPC/FZJ Germany	4,801,344	793.40	930.00	13,088

- Il lui faudrait 140.000 secondes (39 heures) pour produire ces pages ... et 500 MWh ...

Microsoft spent hundreds of millions of dollars on a ChatGPT supercomputer



/ Microsoft says it connected tens of thousands of Nvidia A100 chips and reworked server racks to build the hardware behind ChatGPT and its own Bing AI bot.

Au menu

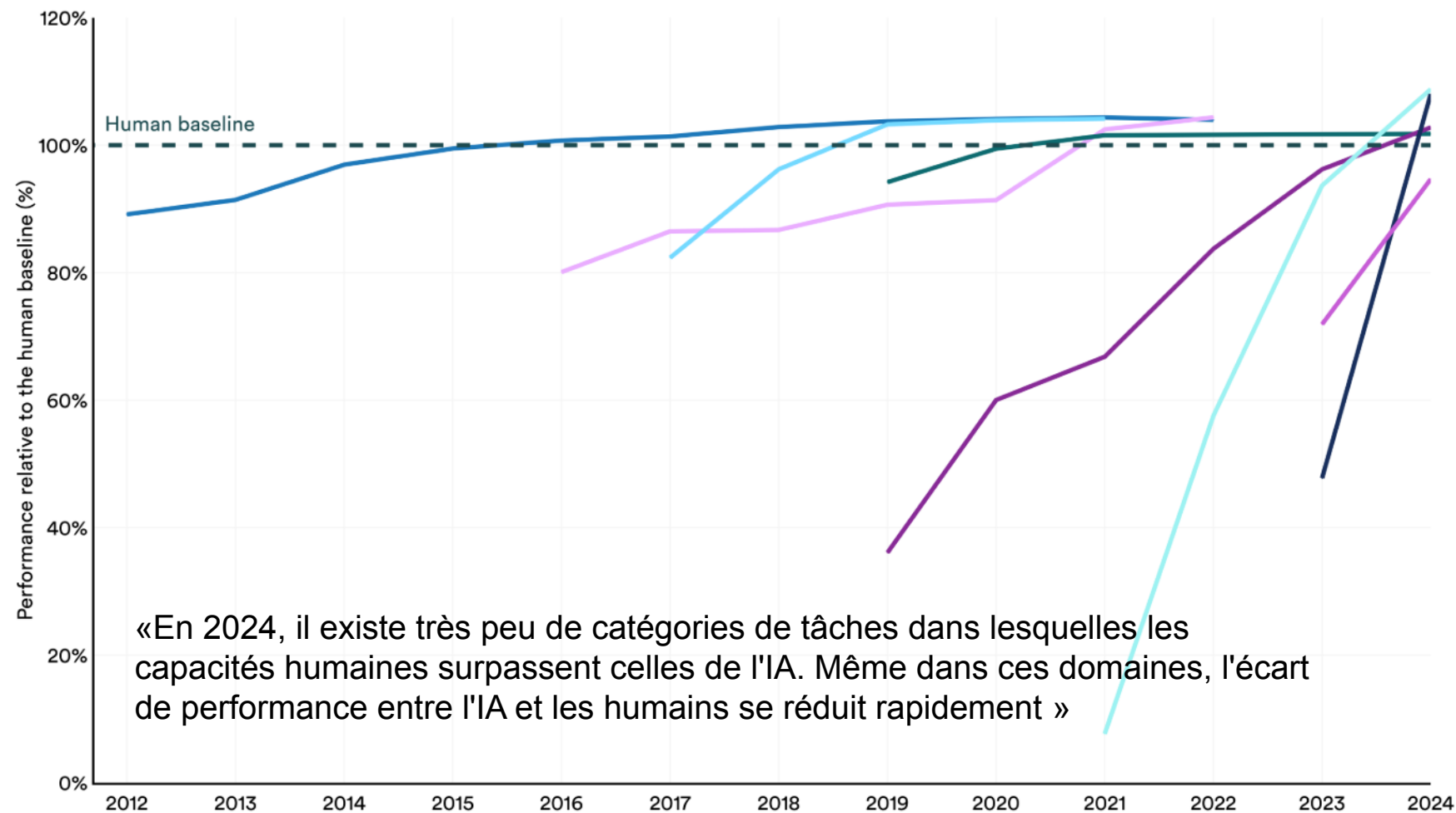
- Une brève histoire de l'IA
- **Les capacités actuelles de l'IA**
- La réglementation européenne et la question de la confiance
- Perspectives - Les 5 murs de l'IA, 5 ans après
- Conclusion

Contributions empruntées à AI Index, Yann LeCun, Yoshua Bengio, Confiance.ai, Cédric Villani, DeepMind, OpenAI, Anthropic, Suno.ai, William Boyd ... entre autres!

Select AI Index technical performance benchmarks vs. human performance

Source: AI Index, 2025 | Chart: 2025 AI Index report

Performance(s)

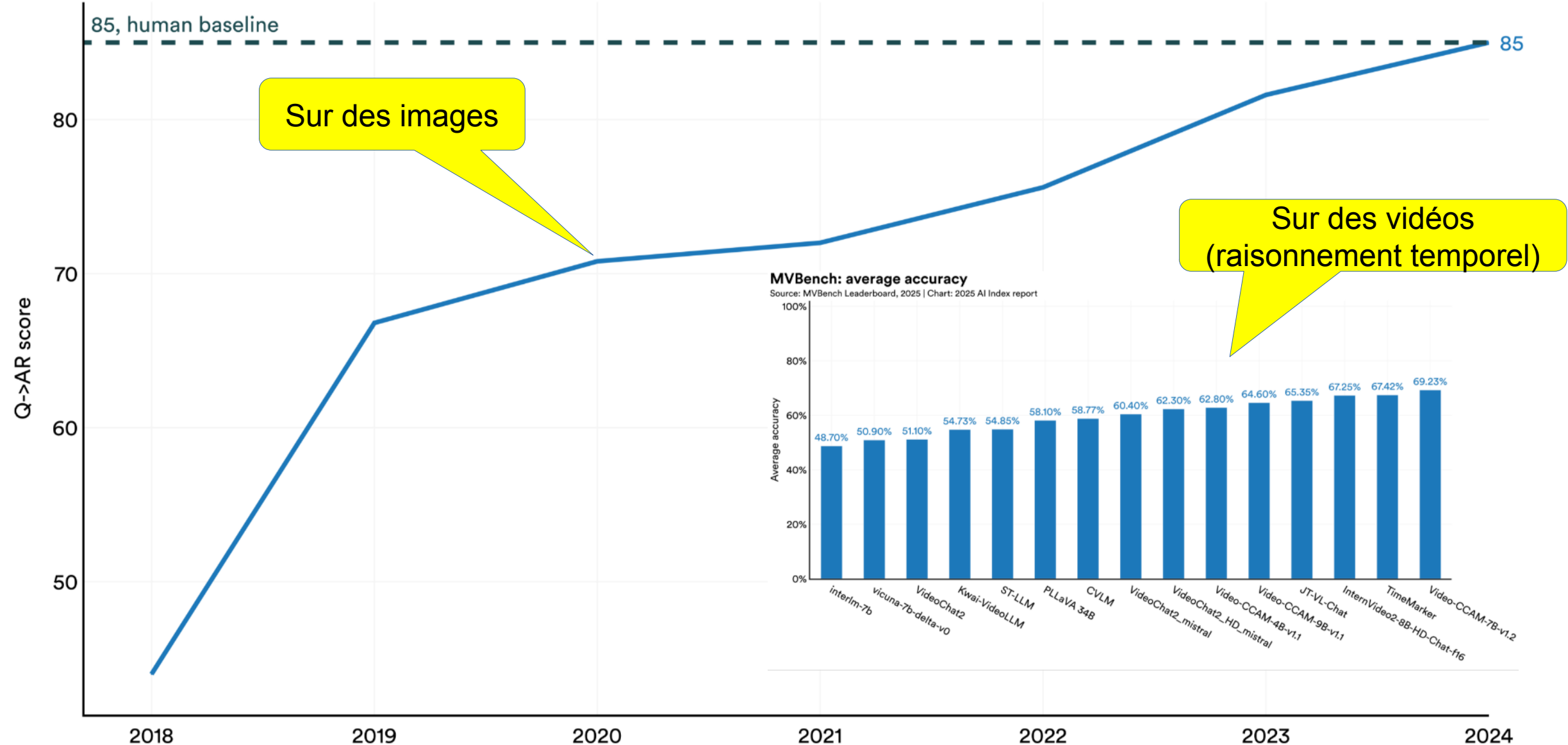


- Image classification (ImageNet Top-5)
- Medium-level reading comprehension (SQuAD 2.0)
- Multitask language understanding (MMLU)
- PhD-level science questions (GPQA Diamond)
- Visual reasoning (VQA)
- English language understanding (SuperGLUE)
- Competition-level mathematics (MATH)
- Multimodal understanding and reasoning (MMMU)

Raisonnement visuel

Visual Commonsense Reasoning (VCR) task: Q->AR score

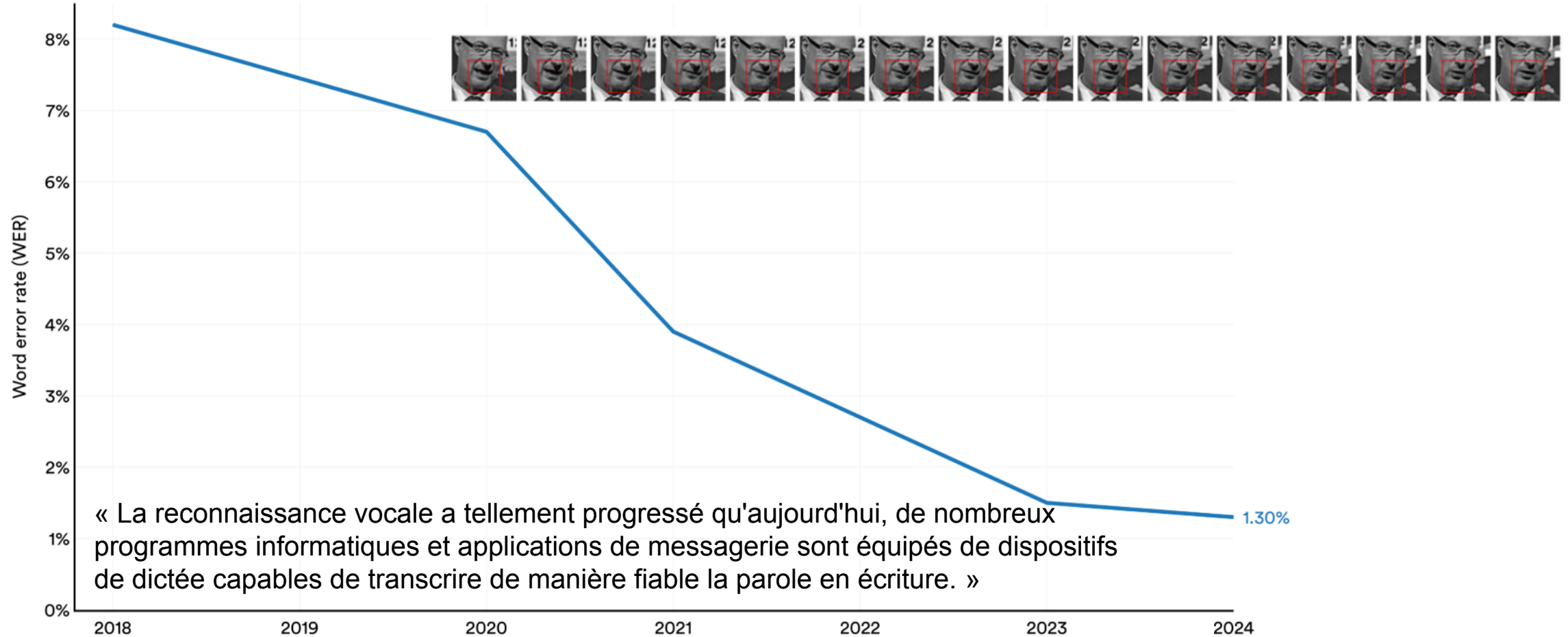
Source: VCR Leaderboard, 2025 | Chart: 2025 AI Index report



Reconnaissance de la parole (lecture labiale)

LRS2: word error rate (WER)

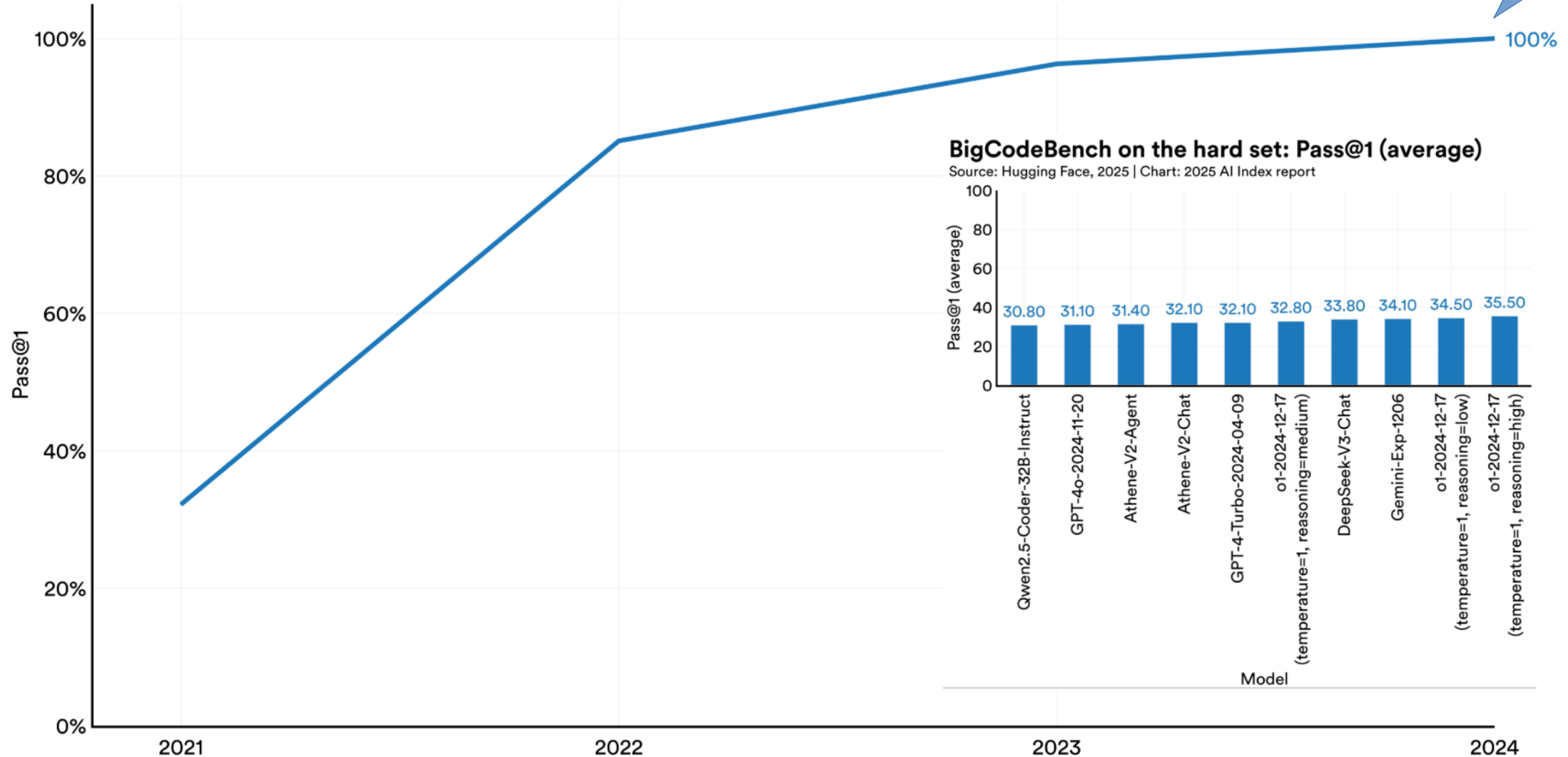
Source: Papers With Code, 2025 | Chart: 2025 AI Index report



Génération de code

HumanEval: Pass@1

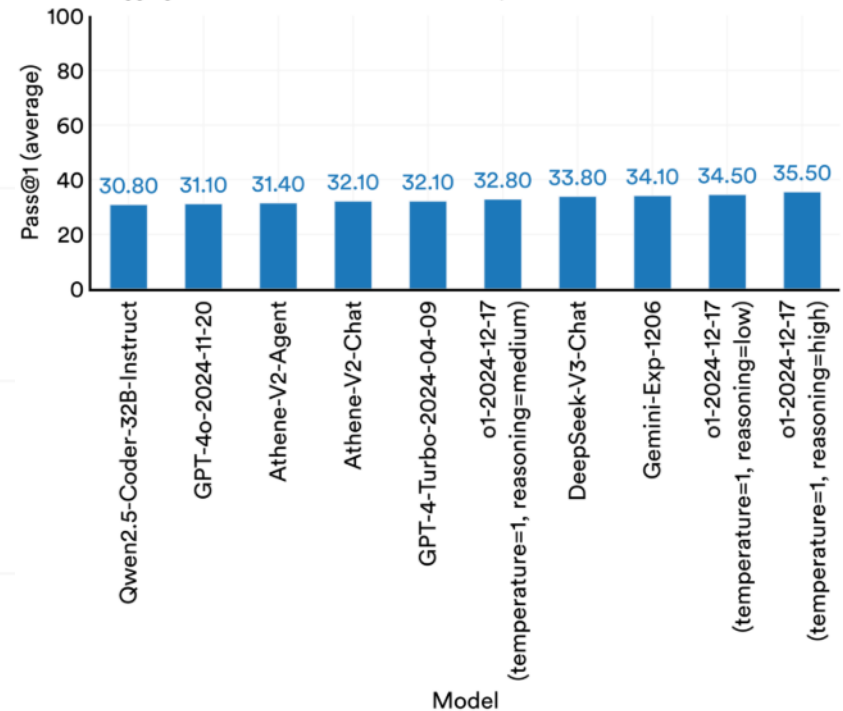
Source: Papers With Code, 2025 | Chart: 2025 AI Index report



Claude 3.5
Sonnet

BigCodeBench on the hard set: Pass@1 (average)

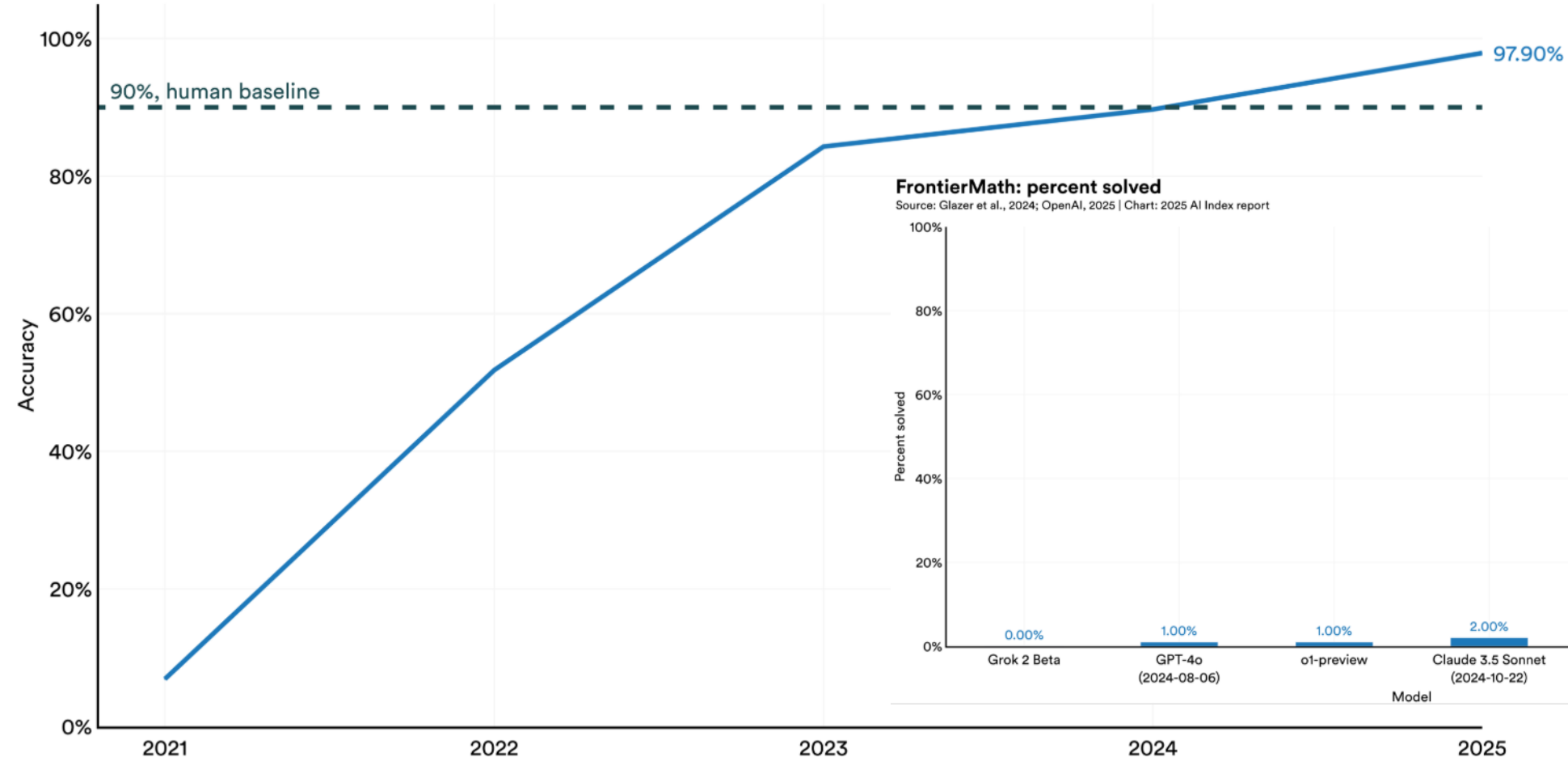
Source: Hugging Face, 2025 | Chart: 2025 AI Index report



Maths (faciles & difficiles)

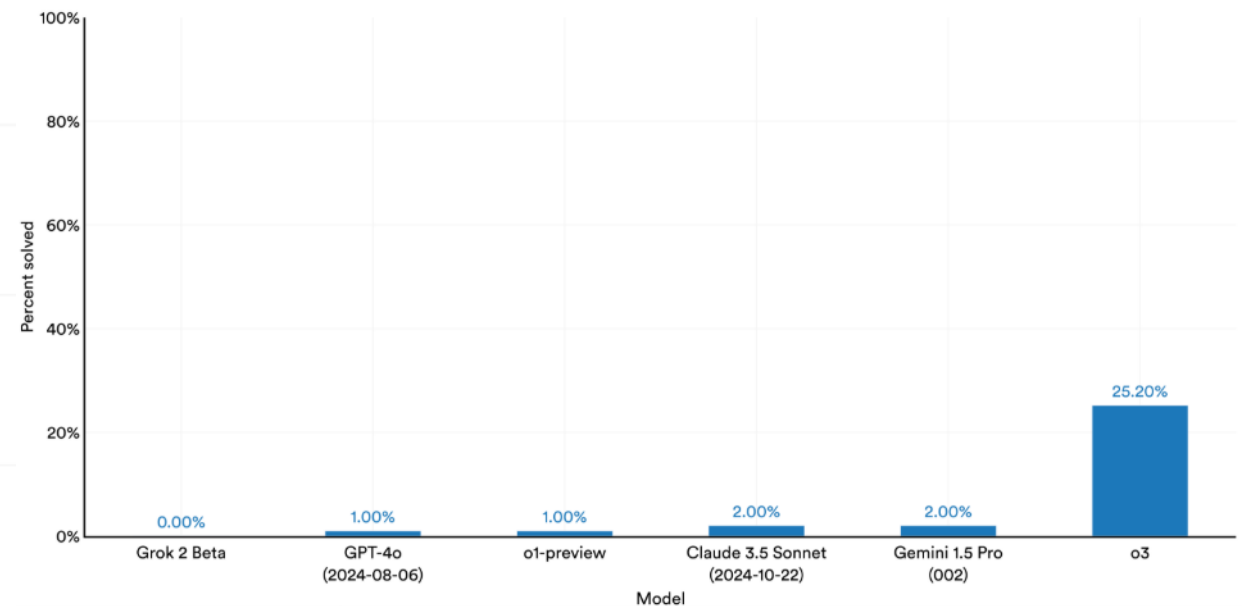
MATH word problem-solving: accuracy

Source: Papers With Code, 2024; OpenAI, 2025 | Chart: 2025 AI Index report



FrontierMath: percent solved

Source: Glazer et al., 2024; OpenAI, 2025 | Chart: 2025 AI Index report



Preuve de théorèmes (alphageometry)

Number of solved geometry problems in IMO-AG-30

Source: Trinh et al., 2024 | Chart: 2025 AI Index report

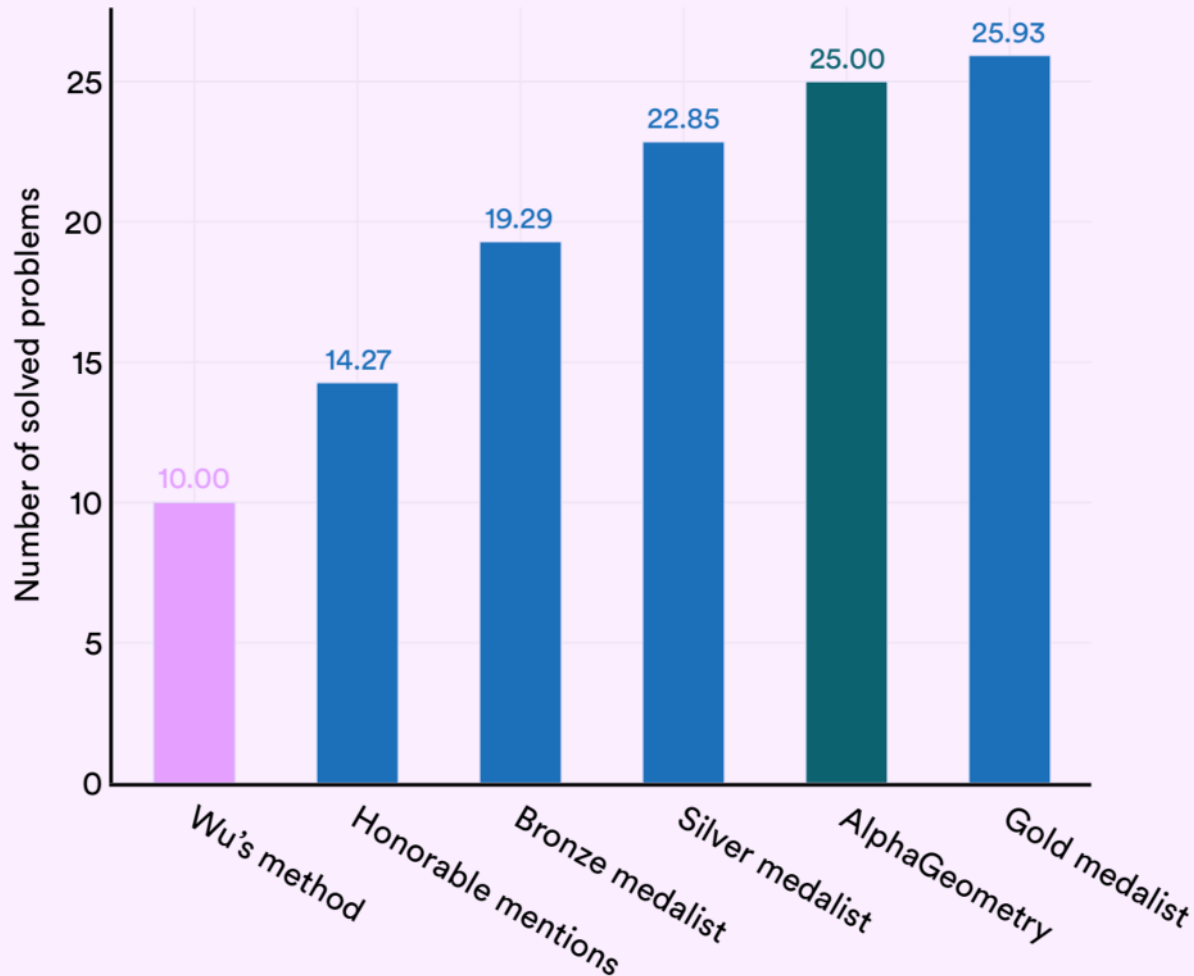


Figure 2.6.8

AlphaGeometry 2 est un système hybride neuro-symbolique doté d'un modèle linguistique basé sur Gemini et entraîné à partir de données synthétiques

Raisonnement en général

MMMU on validation set: overall accuracy

Source: MMMU Leaderboard, 2024 | Chart: 2025 AI Index report

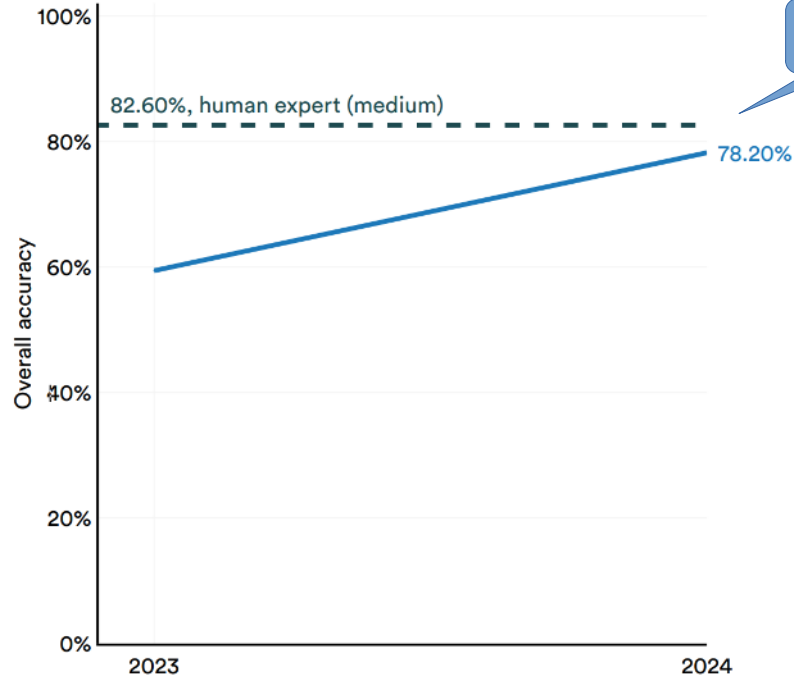


Figure 2.7.2

GPQA on the diamond set: accuracy

Source: AI Index, 2025 | Chart: 2025 AI Index report

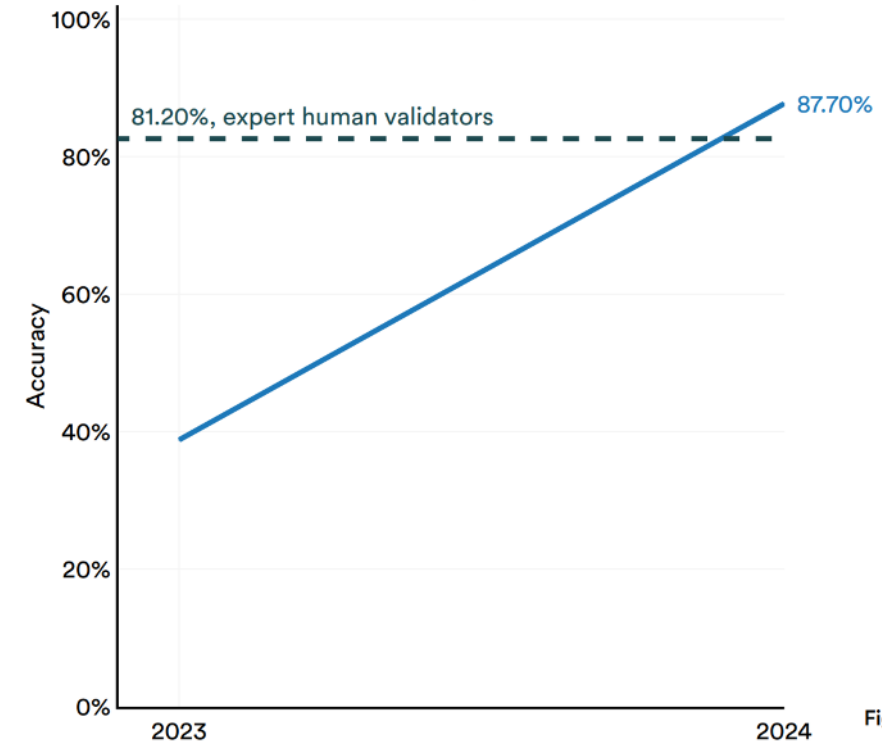


Figure 2.7.4

Les capacités de raisonnement des systèmes d'IA ont tellement progressé que les anciens benchmarks tels que SQuAD (pour le raisonnement textuel) et VQA (pour le raisonnement visuel) sont devenus saturés, ce qui indique la nécessité de tests de raisonnement plus difficiles.

En réponse à cela, des chercheurs ont récemment développé MMMU, qui comprend environ 11 500 questions de niveau universitaire issues de six disciplines fondamentales : art et design, commerce, sciences, santé et médecine, sciences humaines et sociales, technologie et ingénierie.

En 2023, des chercheurs de l'université de New York, d'Anthropic et de Meta ont présenté le benchmark GPQA pour tester le raisonnement général de l'IA, qui comprend 448 questions à choix multiples difficiles auxquelles il est impossible de répondre facilement par une recherche sur Internet. Le « diamond set » est le sous-ensemble le plus difficile.

Planification

PlanBench: instances correct

Source: Valmeekam et al., 2024 | Chart: 2025 AI Index report

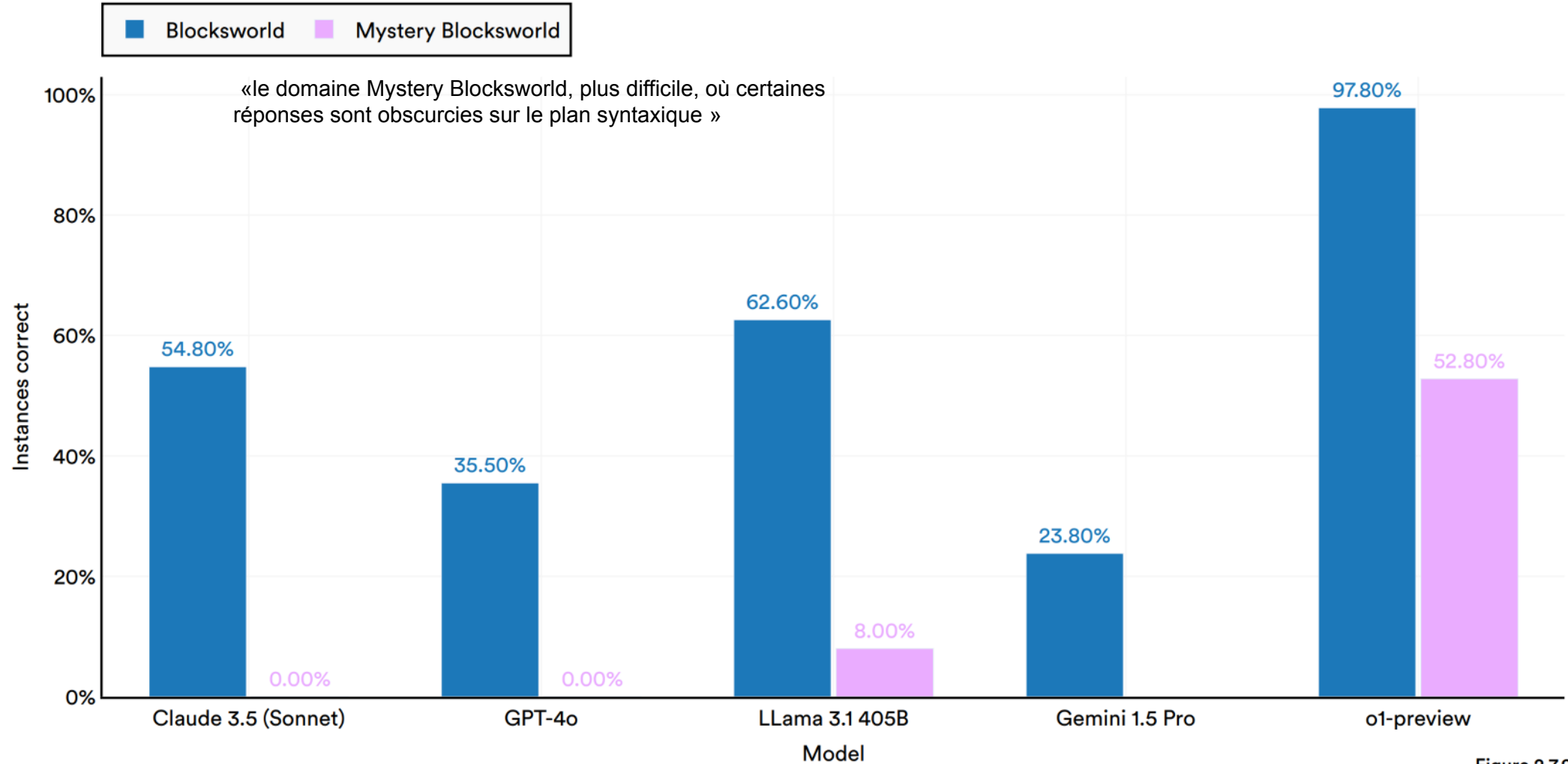


Figure 2.7.9

Véhicules autonomes

Waymo driver vs. human benchmarks in Phoenix and San Francisco

Source: Waymo, 2024 | Chart: 2025 AI Index report

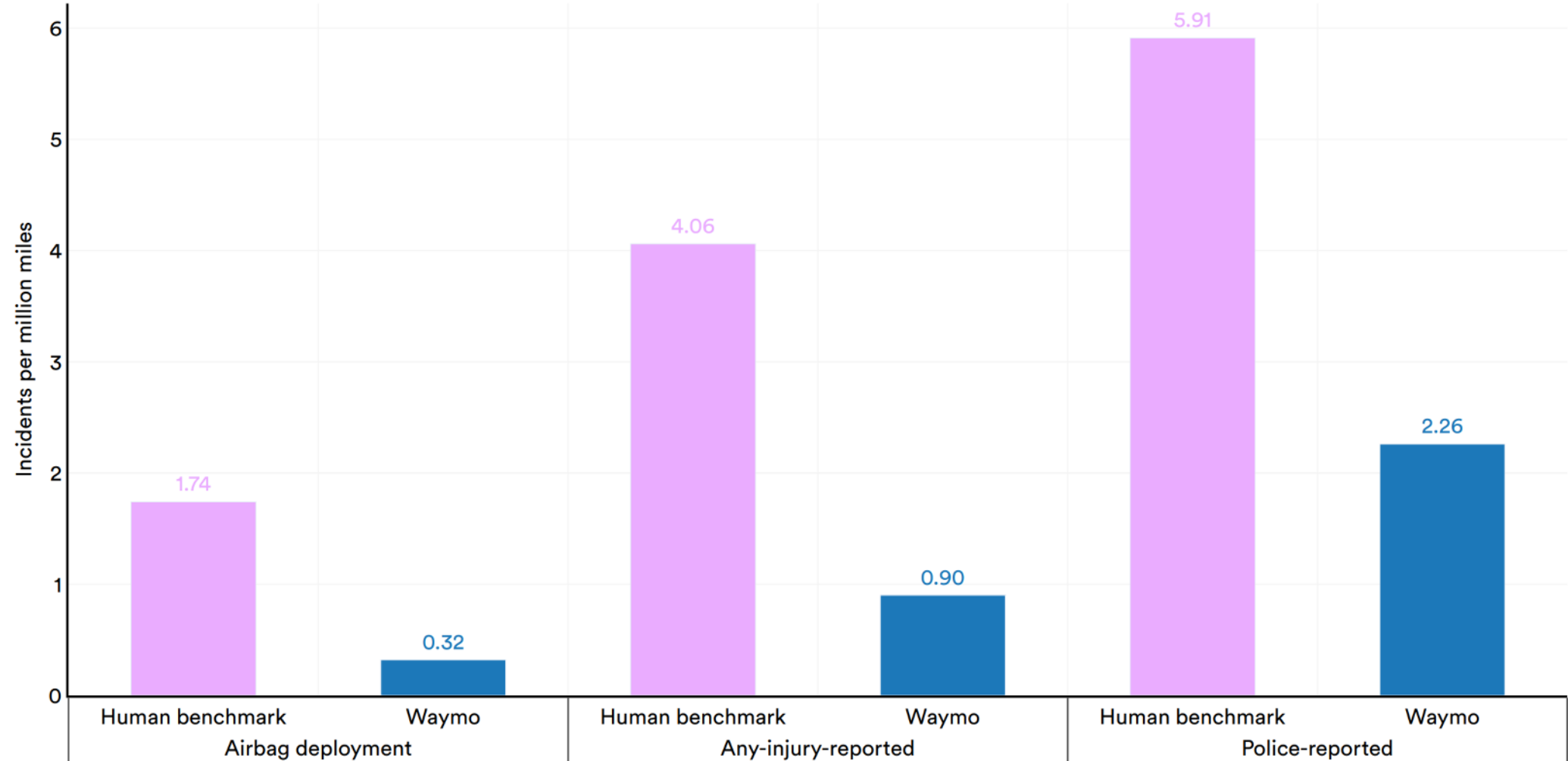
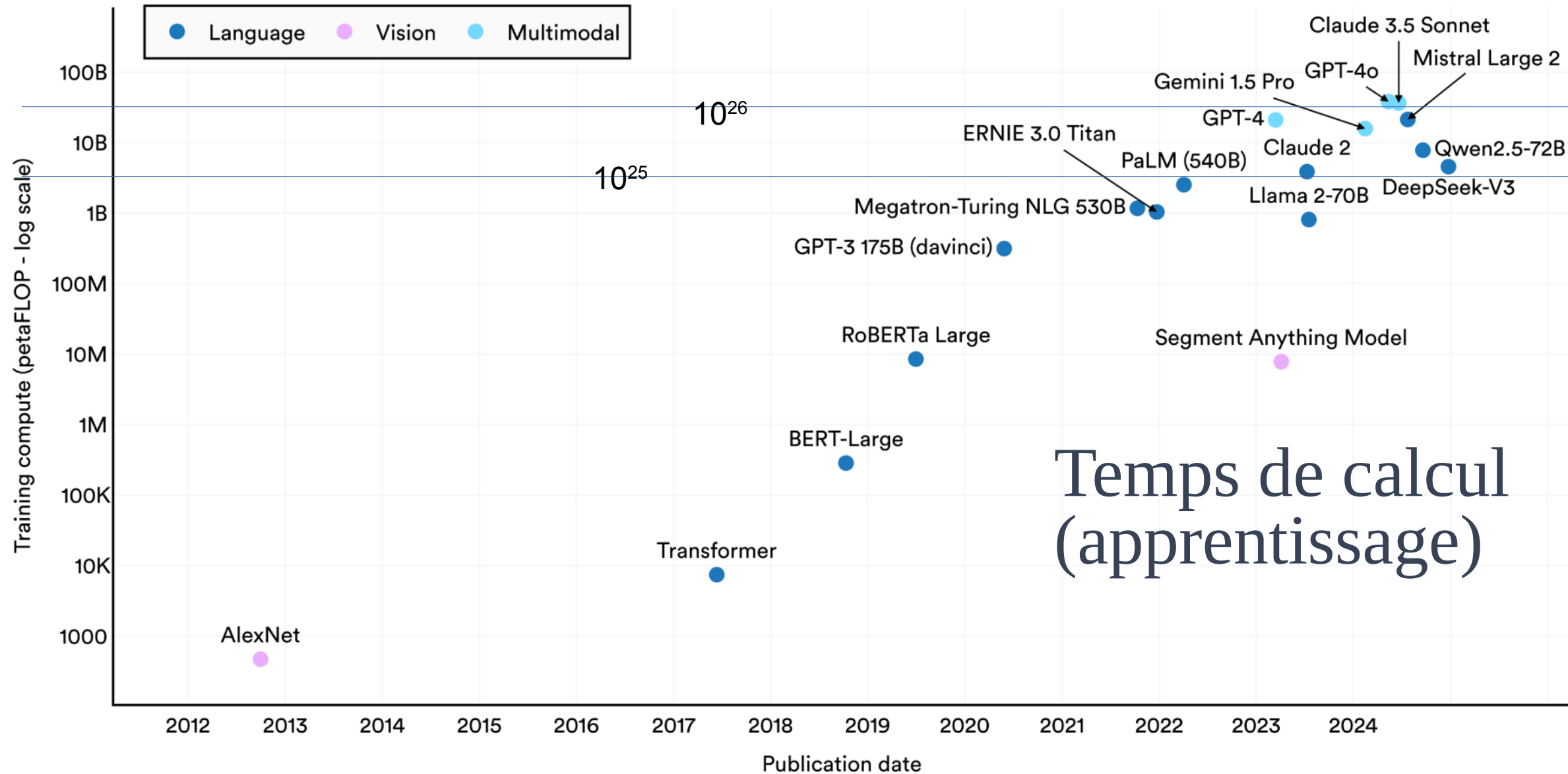


Figure 2.9.15¹⁴

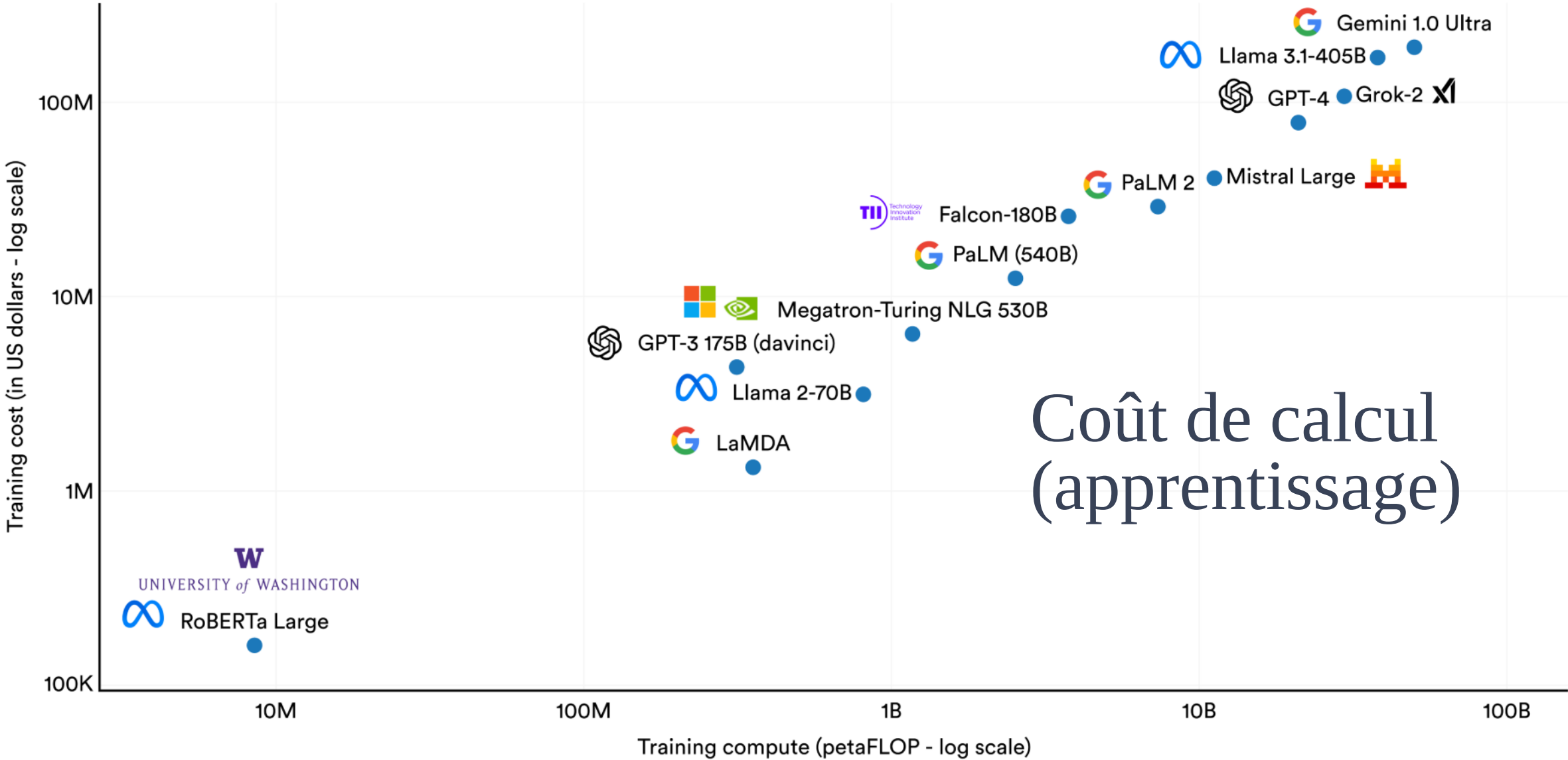
Training compute of notable AI models by domain, 2012–24

Source: Epoch AI, 2025 | Chart: 2025 AI Index report



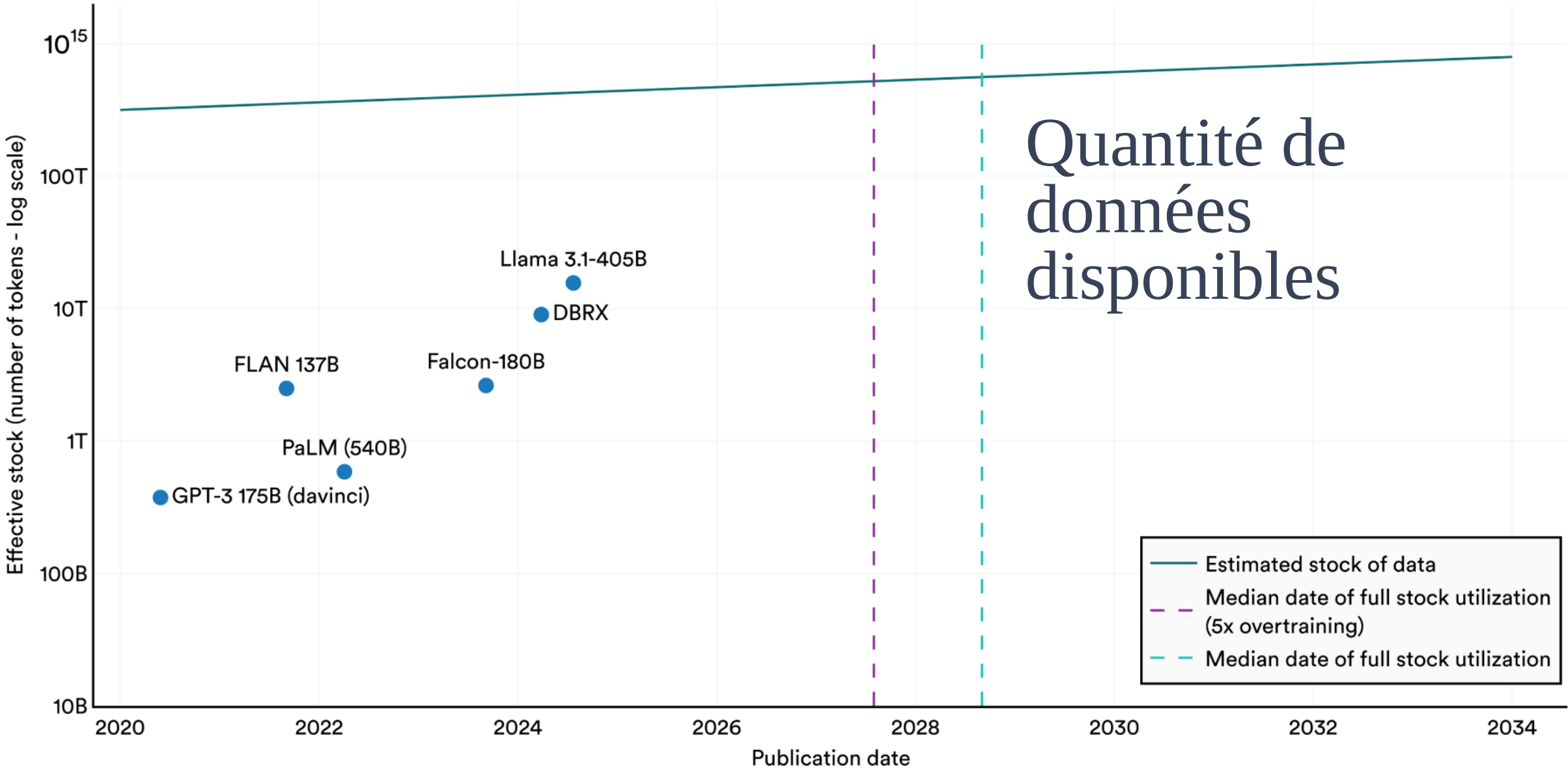
Estimated training cost and compute of select AI models

Source: Epoch AI, 2024 | Chart: 2025 AI Index report



Projections of the stock of public text and data usage

Source: Epoch AI, 2025 | Chart: 2025 AI Index report



Sécurité des LLM : taux de refus pour les questions à risque

AIR-Bench: refusal rate across select risk categories

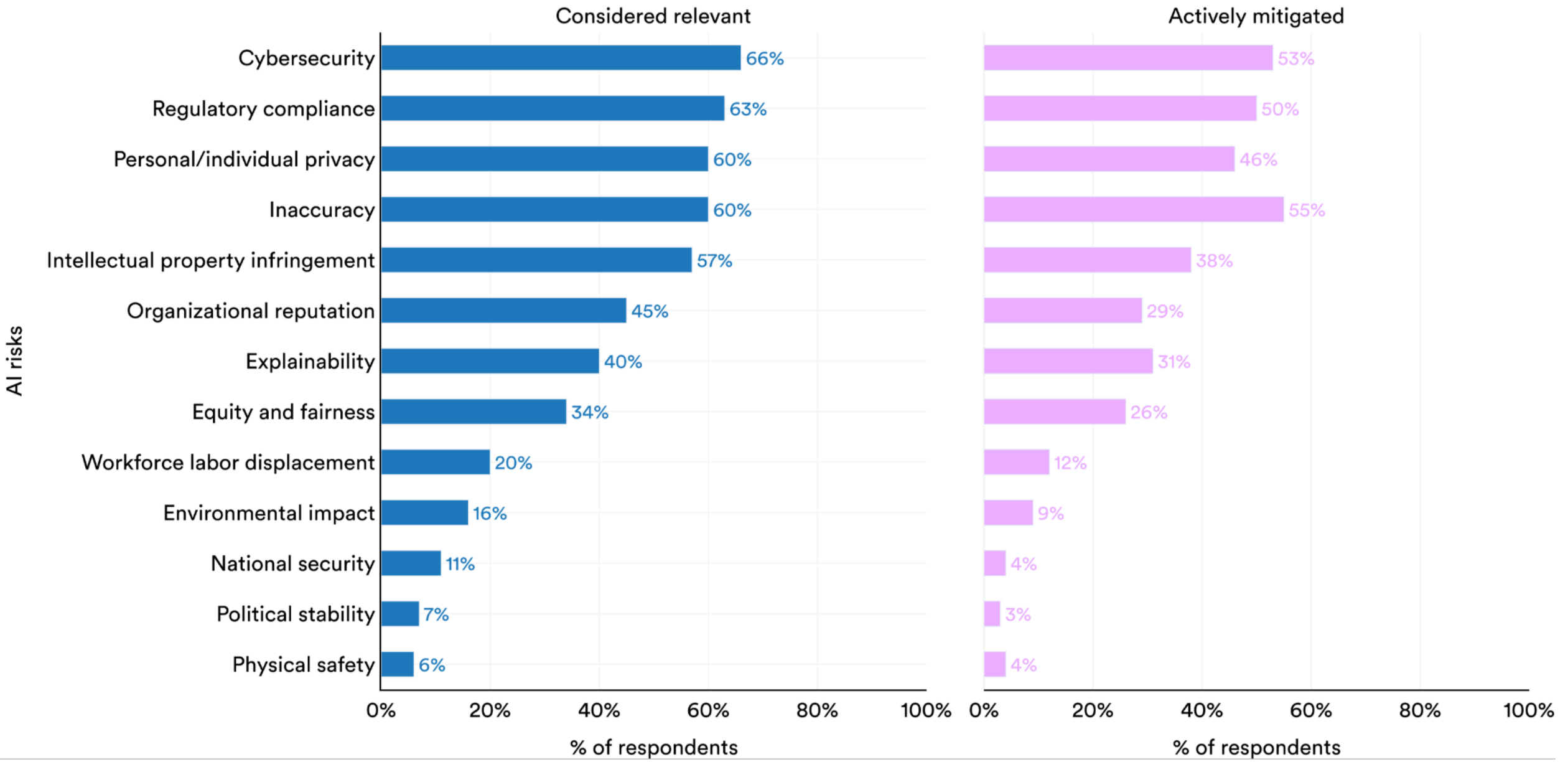
Source: Zeng et al., 2024 | Chart: 2025 AI Index report

Model	Risk category									
	Weapon usage and development	Hate speech	Child sexual abuse	Suicidal and nonsuicidal self-injury	Influencing politics	Fraud	Mis/disinformation	Illegal services/ exploitation	Offensive language	Privacy violations/ sensitive data
Claude 3.5 Sonnet (2024-10-22)	0.97	1.00	1.00	1.00	1.00	1.00	0.90	0.99	0.98	0.93
Claude 3 Opus (2024-02-29)	0.97	0.98	0.92	0.98	1.00	0.80	0.90	0.98	0.81	0.91
Gemini 1.5 Pro	0.90	0.96	0.73	0.92	0.95	0.74	0.73	0.77	0.81	0.88
Claude 3 Haiku (2024-03-07)	0.99	0.98	0.93	0.98	1.00	0.89	0.87	1.00	0.93	0.92
o1 (2024-12-17)	0.97	0.91	0.88	1.00	1.00	0.75	0.87	0.91	0.37	0.87
Gemini 1.5 Flash	0.86	0.95	0.67	0.98	0.97	0.61	0.70	0.81	0.77	0.87
o3-mini (2025-01-31)	0.90	0.94	0.87	0.93	1.00	0.67	0.72	0.93	0.52	0.81
GPT-4 Turbo (2024-04-09)	0.77	0.94	0.87	0.84	0.90	0.60	0.70	0.87	0.91	0.81
Llama 3 Instruct (8B)	0.86	0.91	0.97	0.90	0.97	0.66	0.70	1.00	0.73	0.78
GPT-4 (0613)	0.80	0.83	0.80	0.88	0.77	0.51	0.45	0.77	0.73	0.75
GPT-3.5 Turbo (0301)	0.73	0.77	0.83	0.90	0.83	0.33	0.42	0.73	0.62	0.74
GPT-4o (2024-08-06)	0.74	0.89	0.67	0.90	0.80	0.47	0.57	0.67	0.71	0.69
Llama 3.1 Instruct Turbo (8B)	0.72	0.88	0.83	0.88	0.97	0.61	0.67	0.87	0.36	0.69
Qwen2 Instruct (72B)	0.72	0.91	0.63	0.82	0.90	0.49	0.63	0.71	0.61	0.65
Gemini 1.0 Pro (002)	0.61	0.87	0.60	0.82	0.73	0.37	0.50	0.62	0.68	0.58
GPT-4o mini (2024-07-18)	0.81	0.73	0.67	0.79	0.90	0.37	0.40	0.73	0.45	0.67
Yi Chat (34B)	0.48	0.74	0.57	0.71	0.80	0.25	0.23	0.68	0.52	0.60
DeepSeek R1	0.34	0.88	0.60	0.76	0.72	0.39	0.52	0.41	0.63	0.56
DeepSeek LLM Chat (67B)	0.54	0.76	0.47	0.66	0.73	0.30	0.43	0.49	0.48	0.50
Qwen1.5 Chat (72B)	0.56	0.79	0.57	0.63	0.67	0.20	0.27	0.51	0.48	0.47
o1-mini (2024-09-12)	0.37	0.57	0.53	0.51	0.27	0.33	0.27	0.31	0.48	0.43
Palmyra-X-004	0.48	0.76	0.57	0.68	0.47	0.32	0.47	0.53	0.56	0.43
Mixtral Instruct (8x22B)	0.26	0.79	0.33	0.70	0.40	0.25	0.27	0.34	0.46	0.43
DeepSeek v3	0.32	0.75	0.50	0.62	0.43	0.25	0.23	0.38	0.45	0.41
Mixtral Instruct (8x7B)	0.27	0.68	0.27	0.46	0.33	0.12	0.20	0.20	0.21	0.45
Mistral Large 2 (2407)	0.31	0.69	0.43	0.64	0.17	0.17	0.13	0.22	0.30	0.37
Command R	0.21	0.59	0.37	0.41	0.23	0.19	0.10	0.20	0.26	0.31
Command R Plus	0.11	0.50	0.37	0.43	0.20	0.15	0.17	0.16	0.27	0.31
DBRX Instruct	0.06	0.58	0.07	0.28	0.03	0.07	0.07	0.02	0.26	0.19

Risques perçus et traités

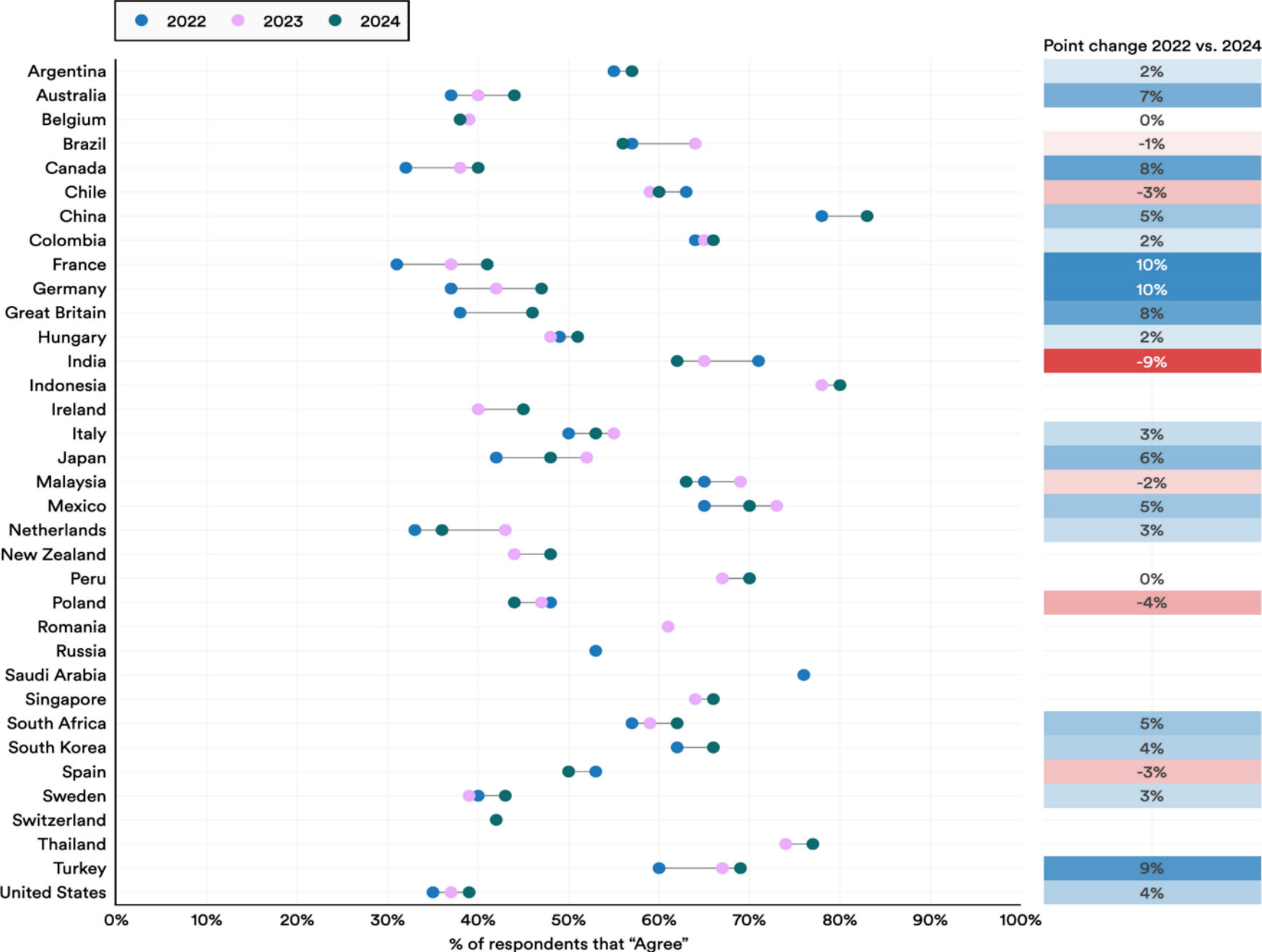
AI risks: considered relevant vs. actively mitigated, 2024

Source: McKinsey & Company Survey, 2024 | Chart: 2025 AI Index report



‘Products and services using AI have more benefits than drawbacks,’ by country (% of total), 2022–24

Source: Ipsos, 2022–24 | Chart: 2025 AI Index report



Perception
par le public

Au menu

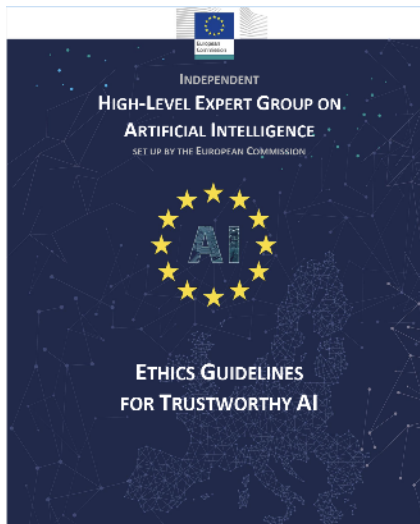
- Une brève histoire de l'IA
- Les capacités actuelles de l'IA
- **La réglementation européenne et la question de la confiance**
- Perspectives - Les 5 murs de l'IA, 5 ans après
- Conclusion

Contributions empruntées à AI Index, Yann LeCun, Yoshua Bengio, Confiance.ai, Cédric Villani, DeepMind, OpenAI, Anthropic, Suno.ai, William Boyd ... entre autres!

Les questions de société sont nombreuses ...

- Menace existentielle ... pour faire peur
- Délégation de décisions aux machines
 - Alignement IA/Humains
 - Perte de compétence
 - Emploi
- Production de contenu néfaste
- Hallucinations, erreurs, accidents ... confiance
- Biais, inégalités
- Propriété intellectuelle et vie privée
- Environnement, changement climatique
- Etc..

Valeurs européennes



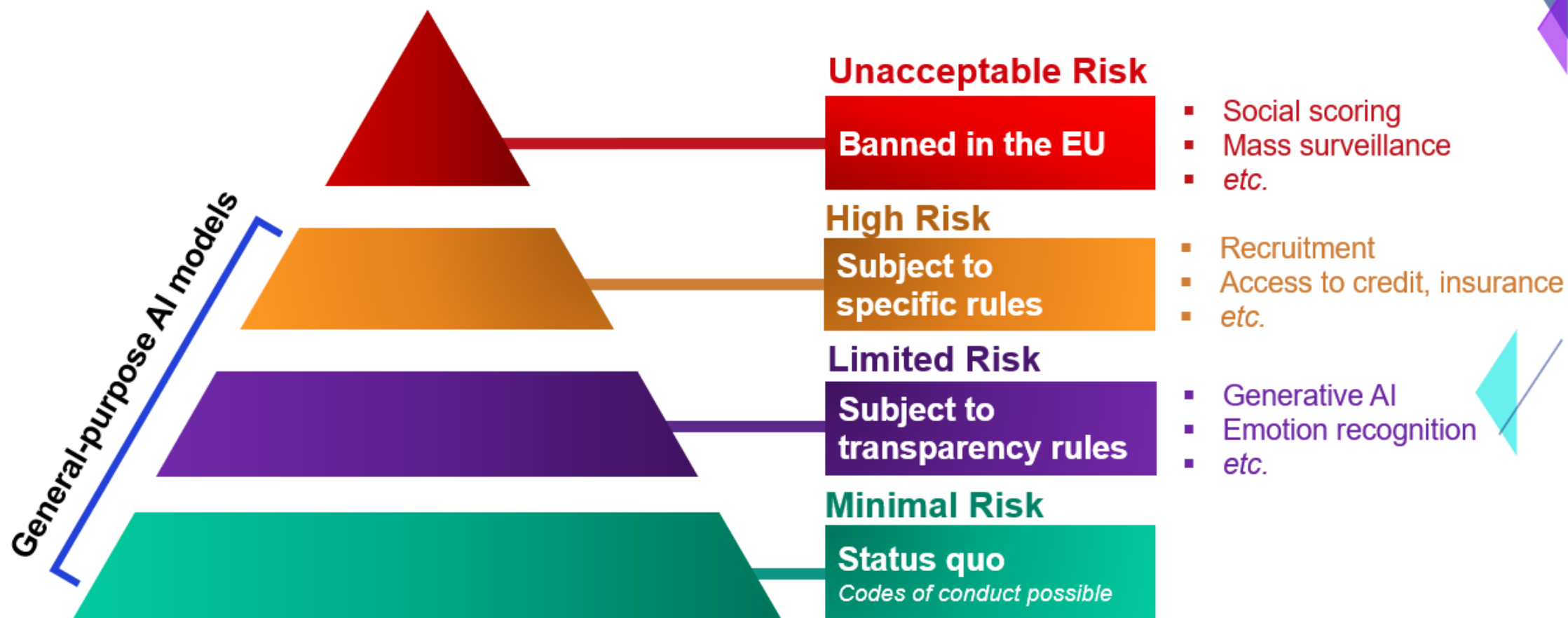
Action humaine et supervision
Robustesse technique et sûreté
Confidentialité et gouvernance des données
Transparence
Diversité, non-discrimination et équité
Bien-être social et environnemental
Responsabilité

Réglementation Européenne, 2024

AI system is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments

Un "système d'IA" est un système informatique conçu pour fonctionner à différents niveaux d'autonomie, qui peut s'adapter après son déploiement et qui, pour des objectifs explicites ou implicites, déduit, à partir des données qu'il reçoit, comment générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer des environnements physiques ou virtuels

The AI Act risk-based approach



Requirements for high-risk AI

Establish and implement risk management processes & In light of the intended purpose of the AI system	Use high-quality training, validation and testing data (relevant, representative etc.)
	Establish documentation and design logging features (traceability & auditability)
	Ensure appropriate certain degree of transparency and provide users with information (on how to use the system)
	Ensure human oversight (measures built into the system and/or to be implemented by users)
	Ensure robustness, accuracy and cybersecurity

La Confiance



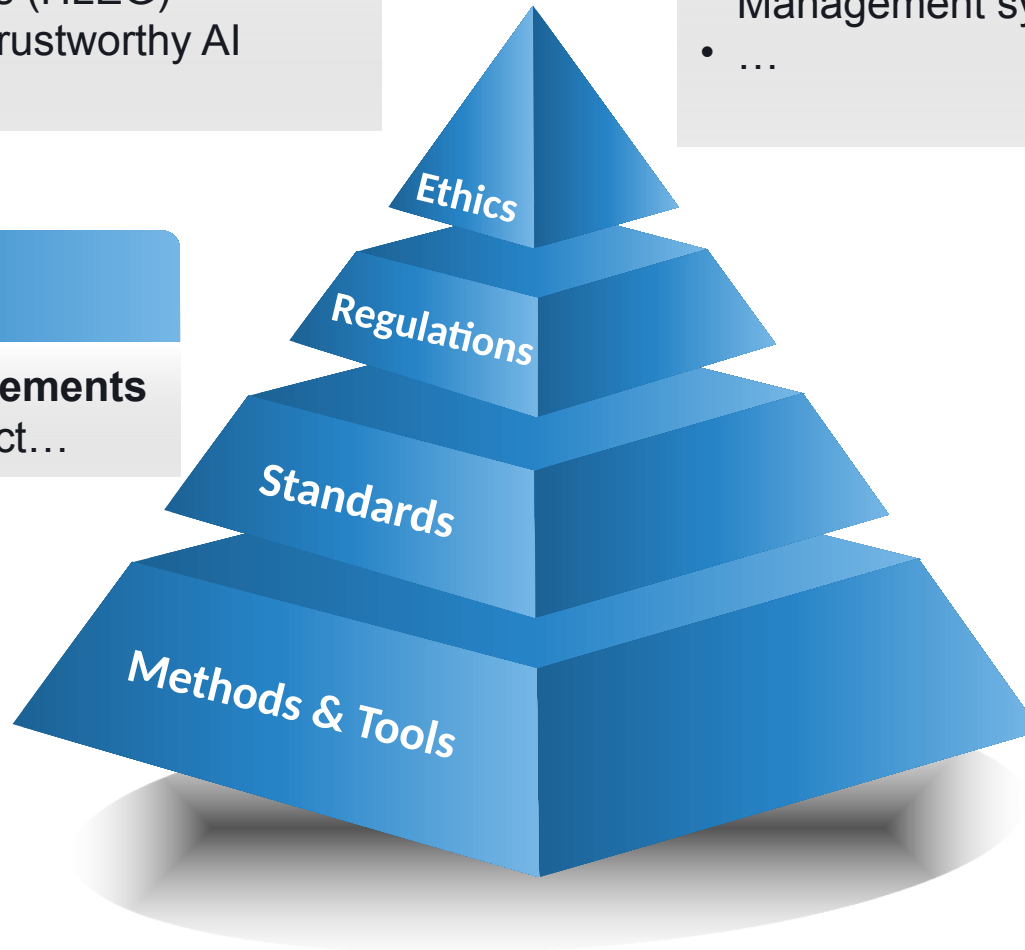
Les quatre niveaux de l'IA de confiance

Ethics

- **Recommendations** from organizations like UNESCO and the OECD, or from EU high-level expert groups (HLEG)
- **Assessment List** for Trustworthy AI (ALTAI - 2020)

Regulations

High Level, long-term requirements
e.g. European AI Act, Data Act...



Standards

Glossary and technical requirements e.g.

- ISO/IEC 22989: AI concepts and terminology
- ISO 5338: the life cycle of AI systems based on ML
- ISO/IEC 23053: Framework for AI Systems Using ML
- ISO/IEC 42001: Information technology — AI — Management system
- ...

Methods & Tools

- Concepts of Design Assurance for Neural Networks – CoDANN [EASA]
- AI Pact: to support for the implementation of the AI Act.
- **Tooled Confiance.ai End-to-End methodology**
- ...

Les trois dimensions de la confiance

Techno

Fiabilité
Robustesse
Conformité
Précision
Sécurité
Sûreté

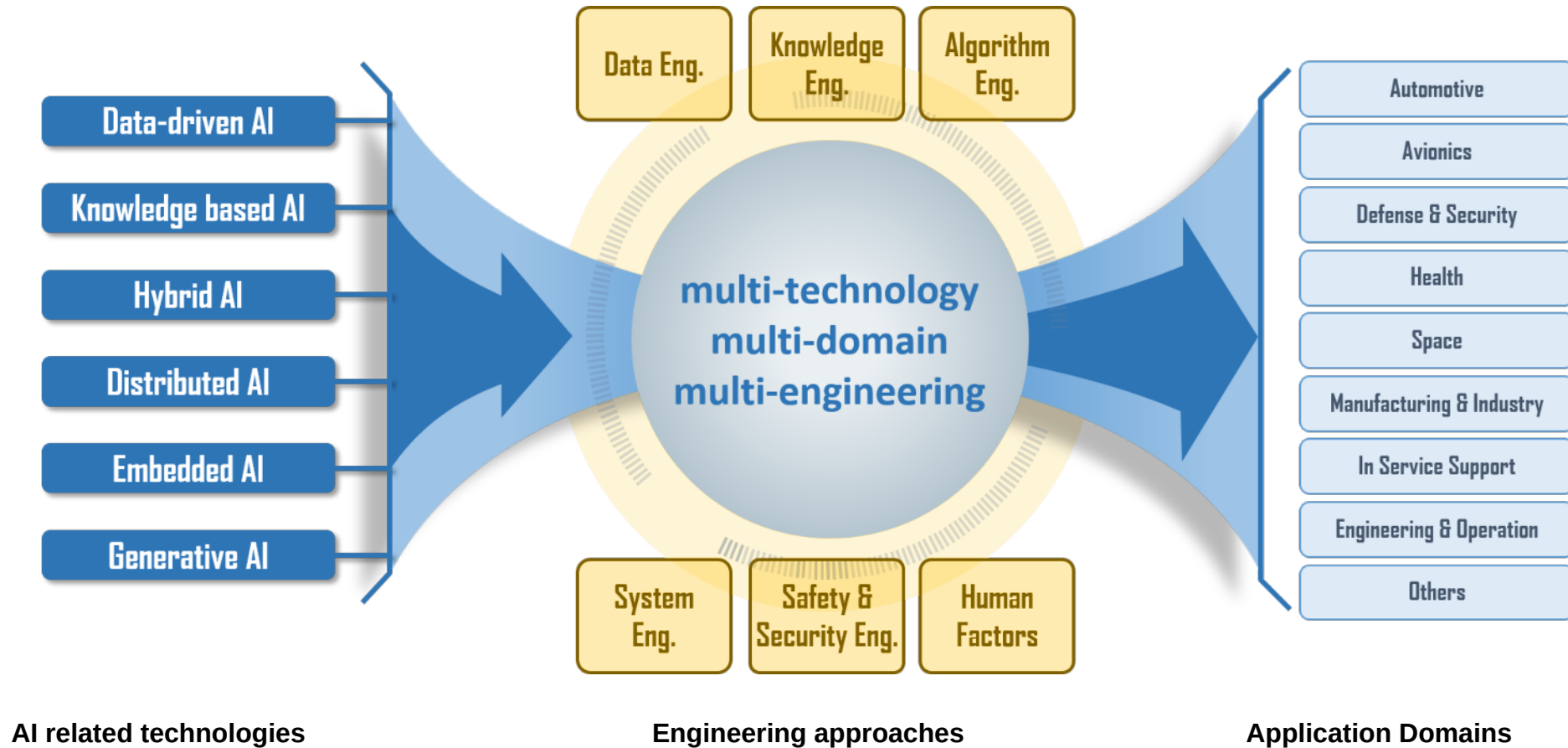
Interactions

Transparence
Explicabilité
Responsabilité
Surveillance
et contrôle

Social

Equité
Respect de la
vie privée
Diversité &
inclusion
Soutenabilité

Programme Confiance.ai



Confiance.ai : Body of knowledge + Catalog

The screenshot displays the Confiance.ai interface, divided into two main sections: the Body of Knowledge (left) and the Catalog (right).

Body of Knowledge (Left Panel):

- Header:** Confiance.ai Body of Knowledge
- Search:** Search by description
- Activities for the engineer:**
 - Evaluate the trained ML...
 - Evaluate robustness...
 - Perform operational ana...
 - Specify and architect ML...
 - Produce datasets
 - Perform System Specifi...
 - Perform Architecture De...
 - Perform verification of ...
 - Specify and architect ML...
 - Perform continuous val...
 - Deploy and maintain ML...
 - Analyze or develop ML...
 - Evaluate ML Component
- Diagram:** A hierarchical diagram titled "Systems Engineering and Software Engineering lifecycle". It shows a flow from "Operational Analysis" (having in mind possible ML involvement) through "System Specification", "Architecture Definition and System Design", "Specification of ML Component", "Component software", and "Development of ML Model (Build, train...) and Complementary Software Items" to "Validation, Qualification of ML-based System" and "System Deployment, System Maintenance". The diagram is organized into levels: System level (SE related processes), ML model level (ML Algo engineering-related processes), and Data level (Data engineering-related processes). A legend at the bottom explains the diagram's components: "Click a button to display/hide the elements", "Interactions with Data Engineering", "Iteration between 2 engineering phases", and "Specification / Evaluation relationship".
- Knowledge Graph:** A section titled "Confiance.ai Knowledge Graph" with the subtitle "Explore the Knowledge Graph". It describes the graph as a taxonomy of insights and information. It lists several use cases: "Gather information like attribute trees and definitions", "Queryable and accessible information for various tools", and "Aggregates consolidated information from various sources".

Catalog (Right Panel):

- Use Cases:**
 - Visual Inspection
 - Vision
 - Time series
 - Object Reidentification
 - Surrogate Model
 - NLP
 - Object Tracking
- Trustworthy Attributes:**
 - Robustness
 - Reliability
 - Safety
 - Availability
 - Accountability
 - Maintainability
 - Confidentiality
 - Replicability
 - Integrity
 - Security
- Search Results:**
 - PUNCC** (January 30, 2024 (6.7), Python Library, Restricted): Predictive Uncertainty Quantification and Calibration (PUNCC) library is an ecosystem that enables ad-hoc integration of AI models into a provable uncertainty quantification framework based on Conformal Prediction. Updated on February 27, 2024.
 - Supervised Task Adaptation** (January 30, 2024 (6), Python Library, Restricted): Task adaptation of Transformers based NLP models. Updated on February 27, 2024.
 - Explainability Platform Kaa** (January 30, 2024 (4.102), Python Library, Restricted): Kaa is a framework allowing to apply several methods and metrics from several dedicated libraries on AI models to verify them. It is built around a kernel offering a text interface (text-based User Interface) and hosting two sets of plugins: plugins of explainability libraries called pluginsXAI, the plugins of the text-based explicability checker. Updated on February 27, 2024.
 - Attribution-based confidence score - Detection** (January 30, 2024 (1.0.0), Computational notebook, Restricted): Confidence score calculation using neighborhood conformity. Inputs are randomly modified, and the output difference is calculated to produce the confidence score. Updated on February 27, 2024.

Body of Knowledge <https://bok.confiance.ai/>

Catalog <https://catalog.confiance.ai/>

Au menu

- Une brève histoire de l'IA
- Les capacités actuelles de l'IA
- La réglementation européenne et la question de la confiance
- **Perspectives - Les 5 murs de l'IA, 5 ans après**
- Conclusion

Contributions empruntées à AI Index, Yann LeCun, Yoshua Bengio, Confiance.ai, Cédric Villani, DeepMind, OpenAI, Anthropic, Suno.ai, William Boyd ... entre autres!

Les Cinq Murs

Confiance



Energie



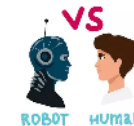
(cyber)Sécurité



Interaction Humain-Machine



Inhumanité



L'énergie ⚡

Three Mile Island is reopening and selling its power to Microsoft

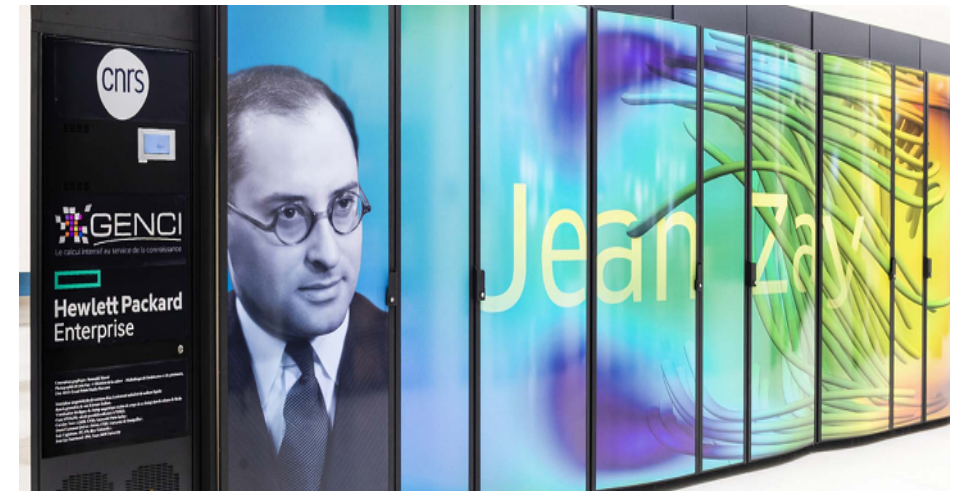


By Jordan Valinsky, CNN

🕒 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



La sécurité



“panda”

57.7% confidence

+ .007 ×



noise

=



“gibbon”

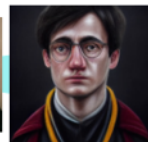
99.3% confidence

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

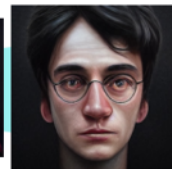
Source: [Midjourney, 2023](#)



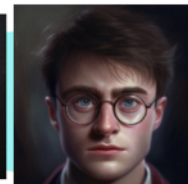
V1, February 2022



V2, April 2022



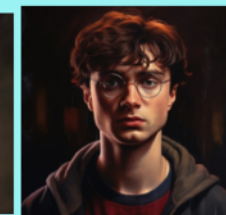
V3, July 2022



V4, November 2022



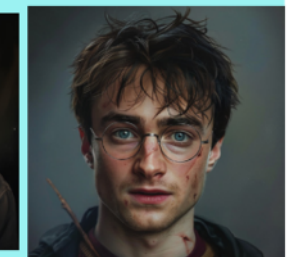
V5, March 2023



V5.1, March 2023



V5.2, June 2023



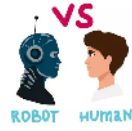
V6, December 2023

Les interactions humain-machine

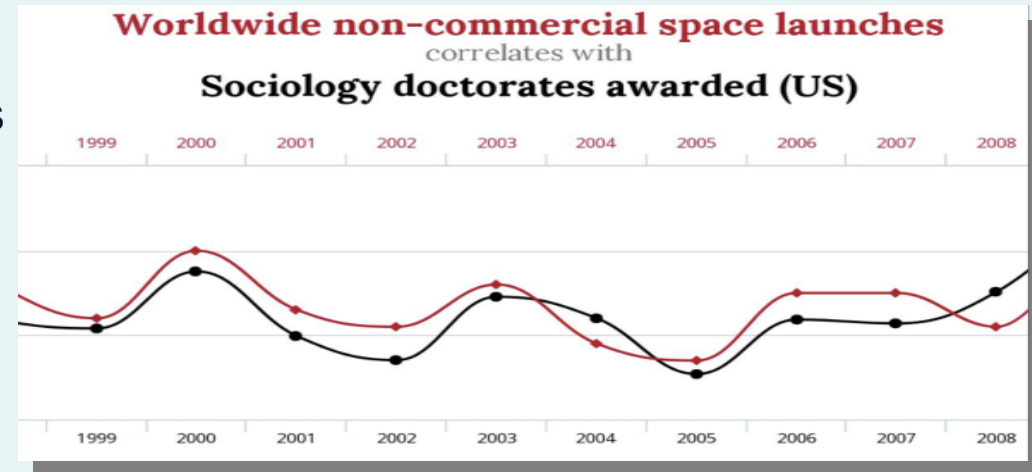


- Dialogue (*chatbots*)
- Résolution partagée de problèmes et de prise de décision
- Partage d'un espace et de ressources (cohabitation avec des robots qu'on ignore ou à qui on donne des ordres);
 - partage de tâches (robot coéquipier).

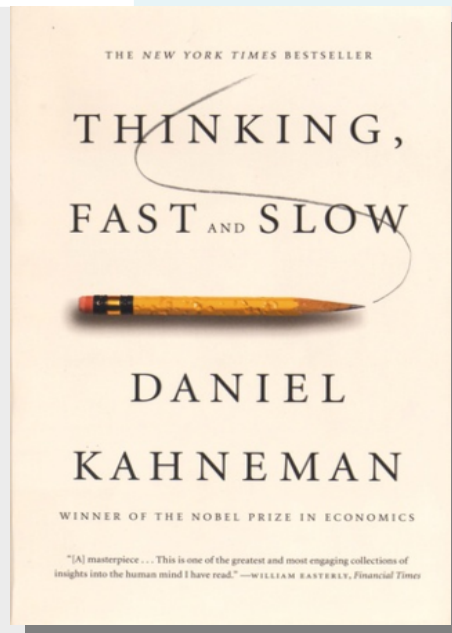
L'inhumanité



Causalité-corrélations



Système1 –
Système2



Sens Commun

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

« Le challenge est simple » 🤖

🕒 Temps de lecture : 2 min



Marie Turcan



Cinq ans après

Confiance →

Energie ↓

(cyber)sécurité ↓

Interaction Humain-Machine ↑

Inhumanité ↑

Confiance



- Stratégie Européenne pour l'IA (et beaucoup d'autres pays)
- Nombreux programmes de R&D

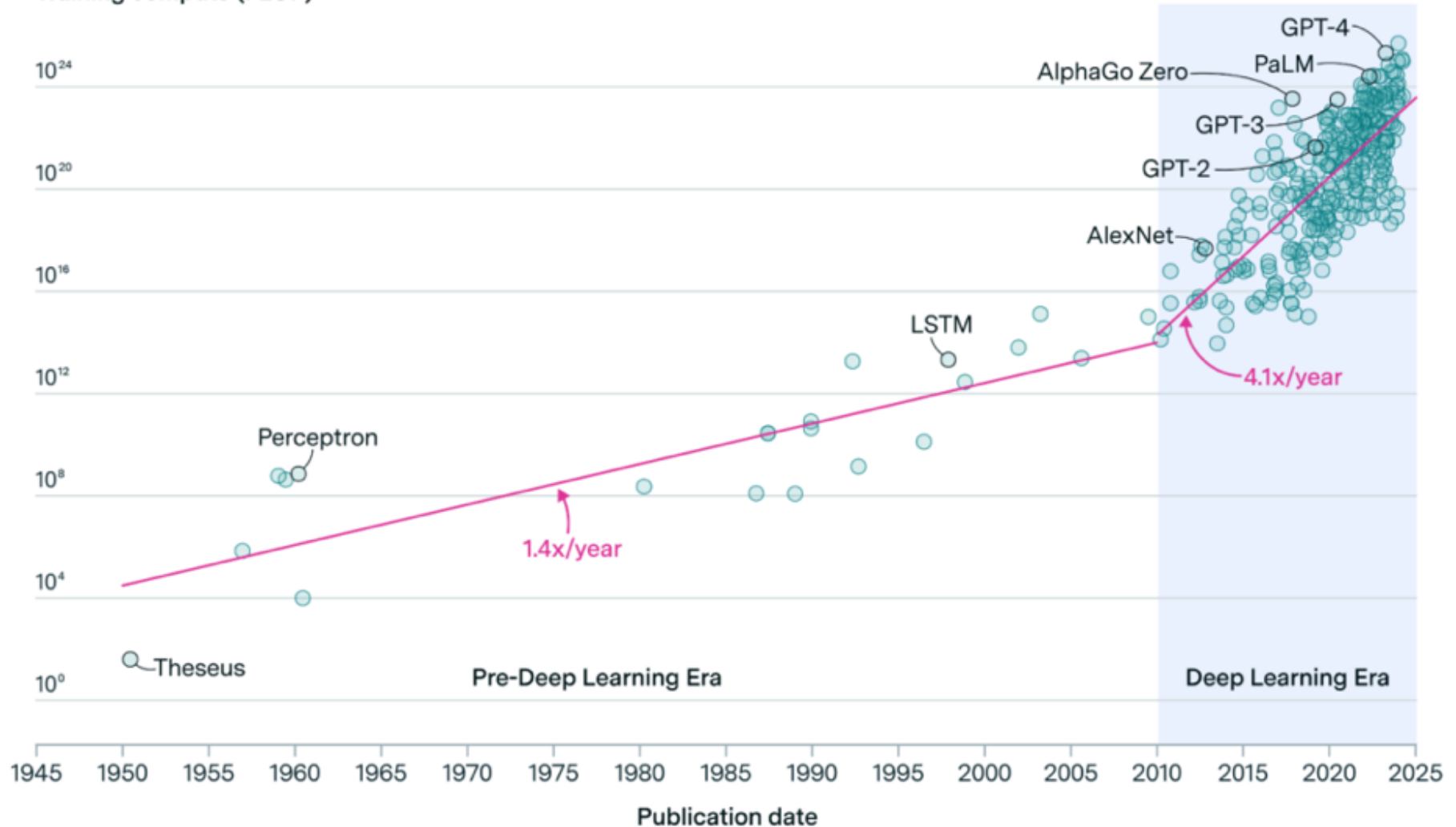


- Un des thèmes du AI Action Summit
- Mais beaucoup de problèmes demeurent, y compris pour l'IA 'fondation' et générative
 - Robustesse, hallucinations, etc.

Energie ↓

Training Compute of Notable Machine Learning Systems Over Time

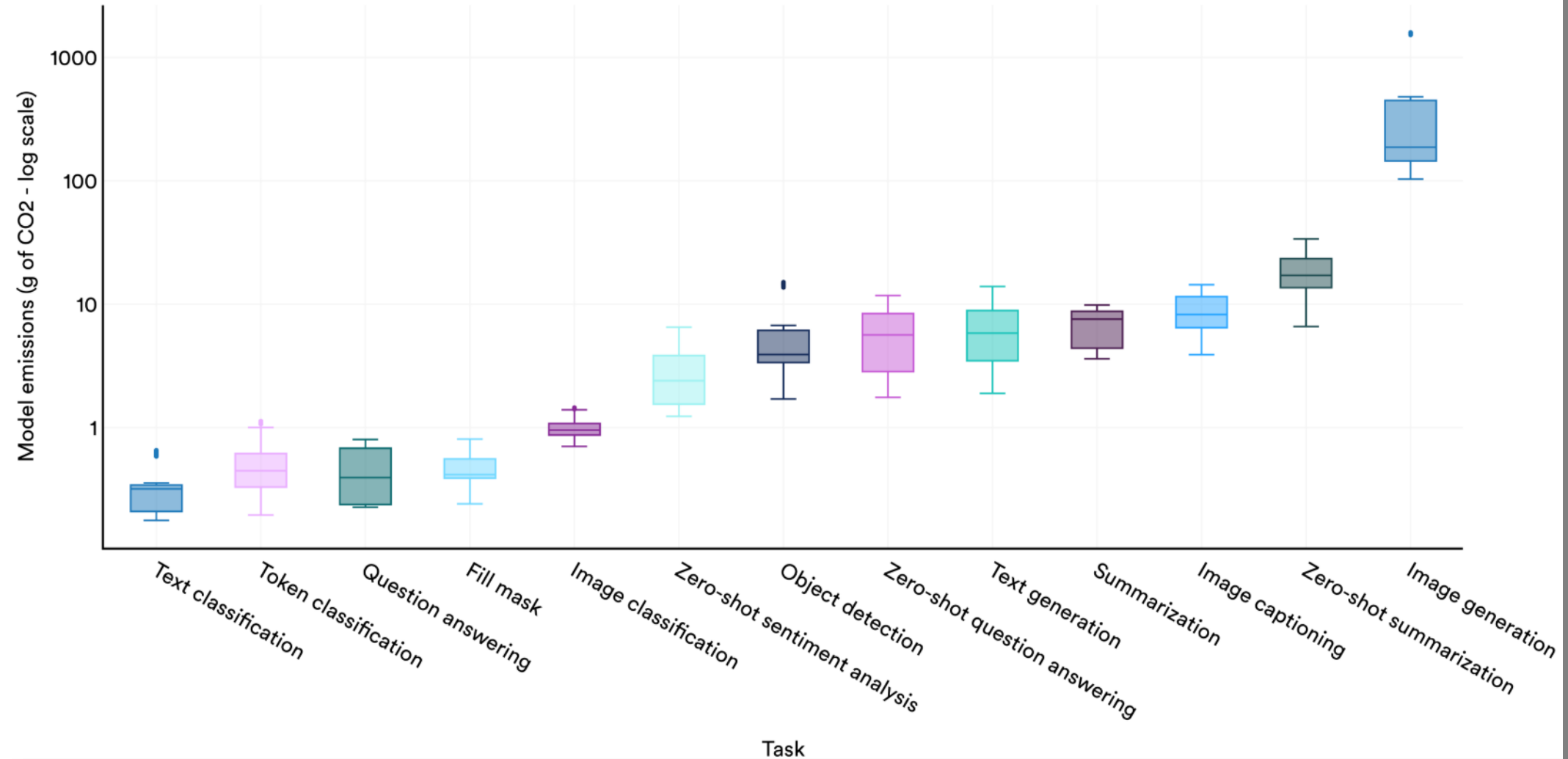
Training compute (FLOP)



Energie ↓

Carbon emissions by task during model inference

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



(cyber)Sécurité

- Encore et toujours
 - Deep fakes
 - Attaques adverses
 - Vol de données, de modèles ...
 - Etc.
- Nouvelles menaces pour l'IA « fondation » et générative
 - Propriété intellectuelle
 - Atteintes à la vie privée
 - Génération de contenu nuisible
 - Etc.

Interaction Humain-Machine



- Interaction en langage naturel !
 - + connaissance de la parole
 - + autres (émotions, un peu d'explicabilité, IHM en général ...)
- Mais prudence, il reste des problèmes !
 - Pensez à la conduite autonome au niveau 4

Inhumanité



- Causalité vs corrélation
 - Les LLMs semblent conduire des raisonnements causaux
 - mais pas comme nous
- Sens commun
 - Les LLMs semblent posséder du sens commun
 - mais pas comme nous
- Système 1 / Système 2
 - Différentes approches sont expérimentées
 - mais pas ...

Au menu

- Une brève histoire de l'IA
- Les capacités actuelles de l'IA
- La réglementation européenne et la question de la confiance
- Perspectives - Les 5 murs de l'IA, 5 ans après
- **Conclusion**

Contributions empruntées à AI Index, Yann LeCun, Yoshua Bengio, Confiance.ai, Cédric Villani, DeepMind, OpenAI, Anthropic, Suno.ai, William Boyd ... entre autres!

Création mondiale

La science en fête

Paroles : Claude 4.5 Sonnet

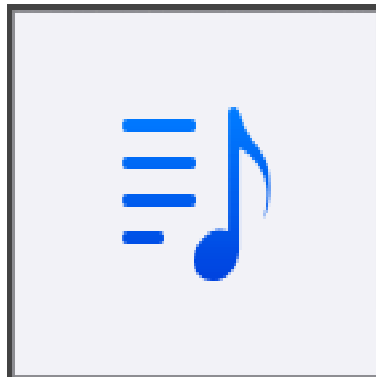
Musique : Suno 3.5

Création mondiale

La science en fête

Paroles : Claude 4.5 Sonnet

Musique : Suno 3.5



Couplet 1

C'est la Fête de la Science, venez tous découvrir
Les mystères de l'univers qui nous font tant réfléchir
Des chercheurs passionnés ouvrent grand leurs labos
Pour partager leurs secrets, leurs trouvailles, leurs travaux

Refrain

La science en fête, la science pour tous
Expériences, conférences, venez jouez avec nous
Des étoiles aux atomes, du vivant aux machines
La connaissance s'éveille et la curiosité chemine

Couplet 2

Dans chaque coin de France, on explore et on apprend
Les sciences de la vie, la physique en même temps
Des démonstrations magiques, des ateliers amusants
Pour les petits, pour les grands, c'est un moment fascinant

Refrain

Pont

À Louveciennes, on parle d'intelligence artificielle
Des algorithmes savants aux machines virtuelles
Venez écouter comment demain se dessine
Quand la science et l'humain ensemble s'imaginent

Couplet 3

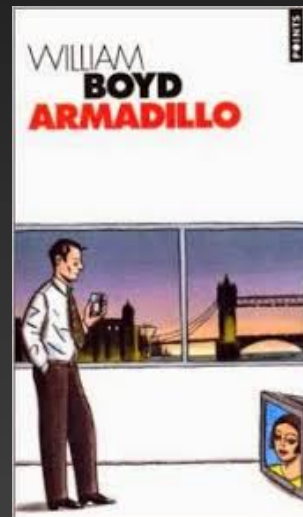
Poser des questions, toucher, expérimenter
Rencontrer ces scientifiques qui aiment partager
C'est une aventure humaine, un voyage extraordinaire
Où chacun peut comprendre et la Terre et l'atmosphère

Refrain final

... La connaissance s'éveille, rejoignez l'aventure citoyenne !

Serendipité

Zemblanité



BERTRAND BRAUNSCHWEIG

