

IA générative dans l'enseignement supérieur

L'IA à l'université : pratiques de recherche et d'enseignement 17 juin 2026

Olivier WONG

Vice-président numérique

Président du bureau

Coordinateur



IA : Intelligence Artificielle

Ce contenu *sous licence ouverte* [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du Programme d'Investissements d'Avenir portant la référence ANR-21-DMES-0001

Quel impact de l'IA générative sur l'apprentissage ?

Une pré-publication du MIT qui a fait du bruit (10 juin 2025)

“While LLMs offer immediate convenience, our findings highlight potential cognitive costs. Over four months, LLM users consistently underperformed at neural, linguistic, and behavioral levels”

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., ... Maes, P. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. *arXiv [Cs.AI]*. Retrieved from <http://arxiv.org/abs/2506.08872>

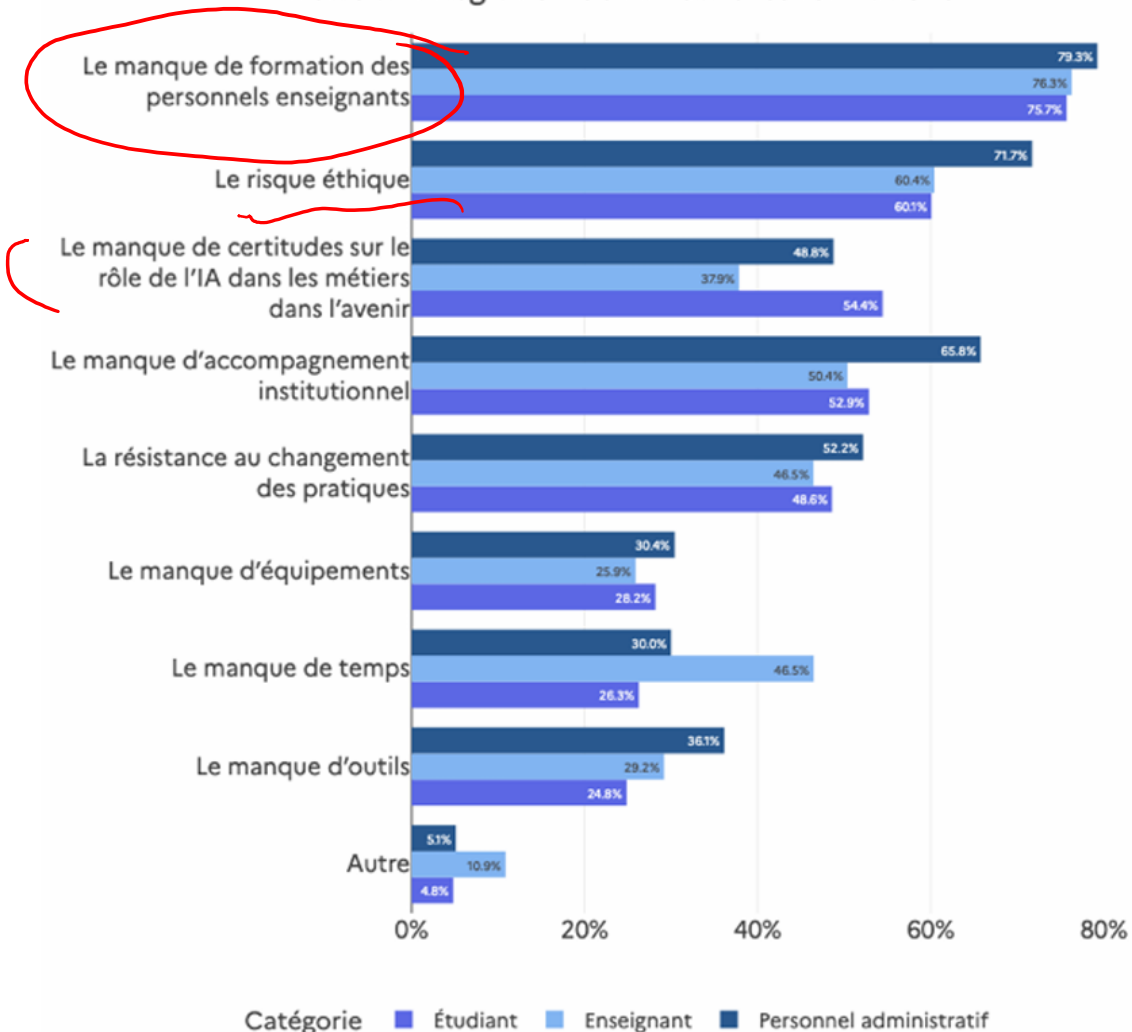
Une méta-analyse publiée dans Nature (26 mars 2026) sur l'impact de ChatGPT sur les résultats d'apprentissage des élèves :

- significantly improves student learning outcomes, boosting cognitive skills and non-cognitive skills
- combined with traditional methods : greatly enhances learning
- impact is greater on declarative than procedural knowledge

Wu, X., Zhu, P., Zhang, J. *et al.* ChatGPT's impact on student learning outcomes: a meta-analysis of 35 experimental studies. *Humanit Soc Sci Commun* **13**, 684 (2026). <https://doi.org/10.1057/s41599-026-07019-z> licence CC-BY-NC-ND 4.0

Rapport IA et ESR, juillet 2025

Selon vous, quels sont aujourd'hui les principaux freins à l'intégration de l'IA dans les formations ?



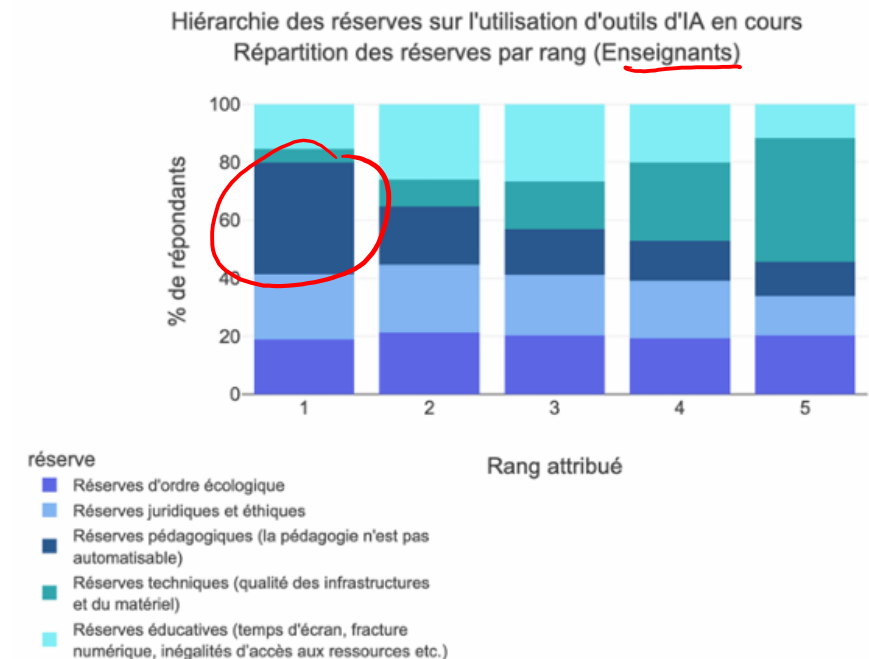
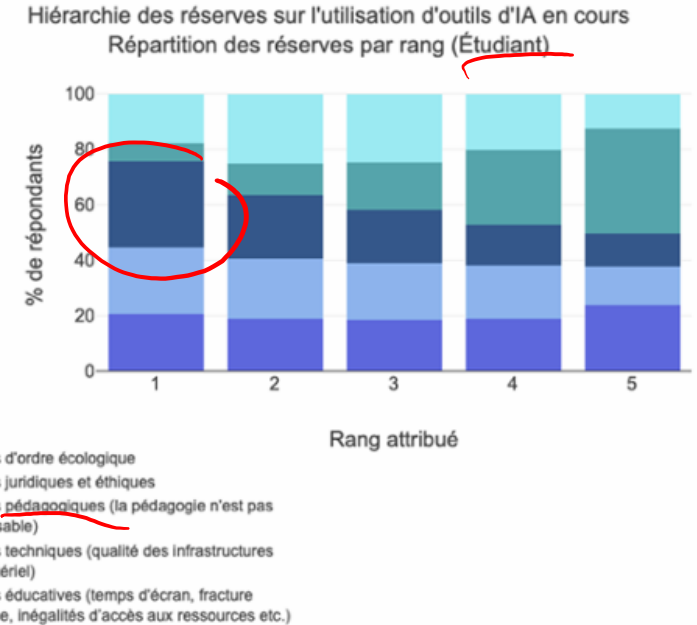
Rapport IA et ESR, juillet 2025

Perception par les acteurs des risques d'impact de l'IA sur les capacités cognitives humaines :

- possible **diminution de la pensée** critique, de la créativité et de la capacité à résoudre des problèmes de manière autonome
- appréhension face à une **dépendance** accrue aux outils d'IA, pouvant mener à une **baisse de l'autonomie** intellectuelle, une **baisse de l'effort** intellectuel et à une dégradation de la qualité de l'apprentissage
- risque significatif de **désapprentissage et d'appauvrissement cognitif**
- **perte de motivation** à apprendre : obstacle à l'apprentissage réel, à la compréhension approfondie des sujets et à la capacité de développer une pensée personnelle et originale

Conséquences potentielles :

- une **baisse de la qualité** du travail
- une **désadaptation** au monde professionnel
- une **dévalorisation de l'effort** intellectuel



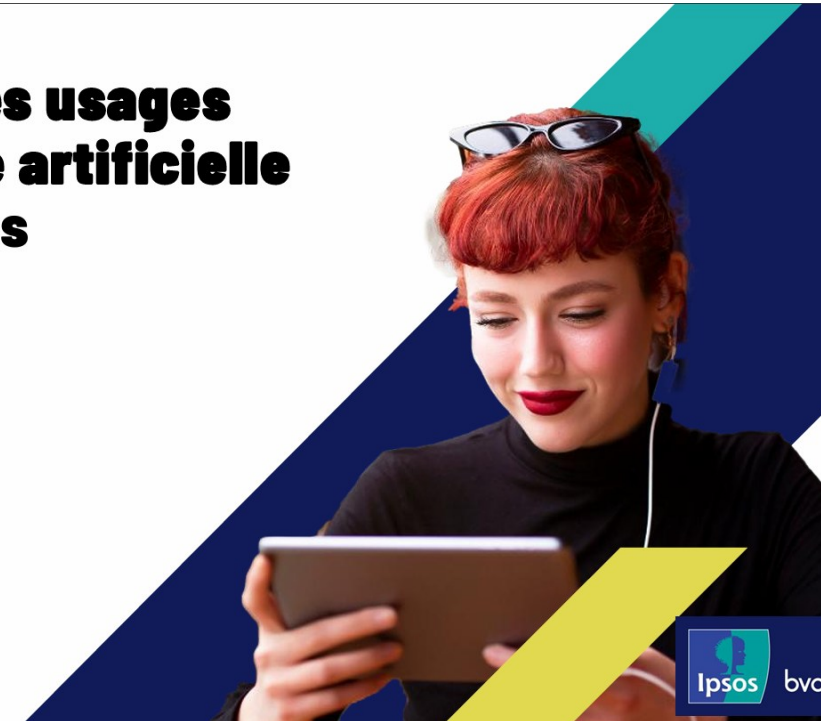
Lecture rapide de l'étude IPSOS BVA pour EPITA (fév. 2026)

Principales conclusions

- L'IA, déjà très présente dans le quotidien étudiant
- Utilisation dans le cadre des études : des dérives préoccupantes
- Une maîtrise technique, mais des lacunes sur les enjeux critiques
- Face aux inquiétudes, une attente d'accompagnement

Observatoire des usages de l'intelligence artificielle par les étudiants

Février 2026



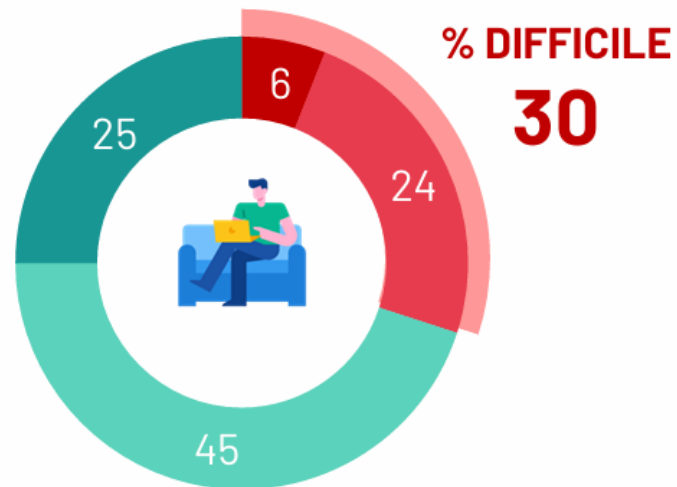
L'IA, déjà très présente dans le quotidien étudiant

Autre signe témoignant de la place importante qu'a prise l'IA : **de nombreux étudiants reconnaissent qu'ils auraient du mal à s'en passer, surtout pour leurs études**

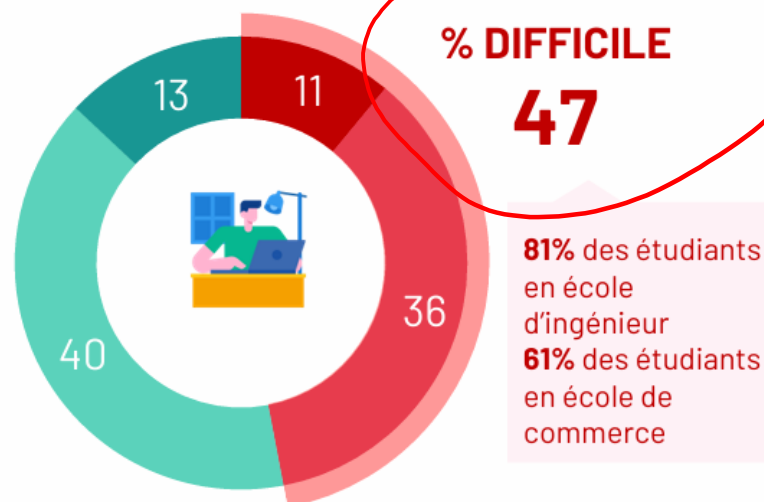
Question : Dans quelle mesure serait-il difficile pour vous de vous passer de l'IA générative pour ... ?

(Base : A ceux qui utilisent l'IA)

Pour vous aider dans
votre vie personnelle



Pour vous aider dans vos études
(exercices, accompagnement...)



81% des étudiants
en école
d'ingénieur
61% des étudiants
en école de
commerce

Parmi les étudiants
dans leur ensemble,

1 sur 2

aurait du mal à se
passer de l'IA dans
au moins un de ces
cadres

■ Très difficile ■ Assez difficile ■ Assez facile ■ Très facile

L'IA, déjà très présente dans le quotidien étudiant

Dans le cadre de leurs études, ils déclarent l'utiliser avant tout pour expliquer, développer ou résumer des informations mais aussi pour générer des devoirs en partie ou totalité



Question : Parmi les suivants, quels sont tous les usages que vous avez fait de l'IA générative dans le cadre de vos études ? Pour ...

Plusieurs réponses possibles par répondant, total supérieur à 100% (Base : A ceux qui utilisent l'IA dans leurs études, au moins de temps en temps)

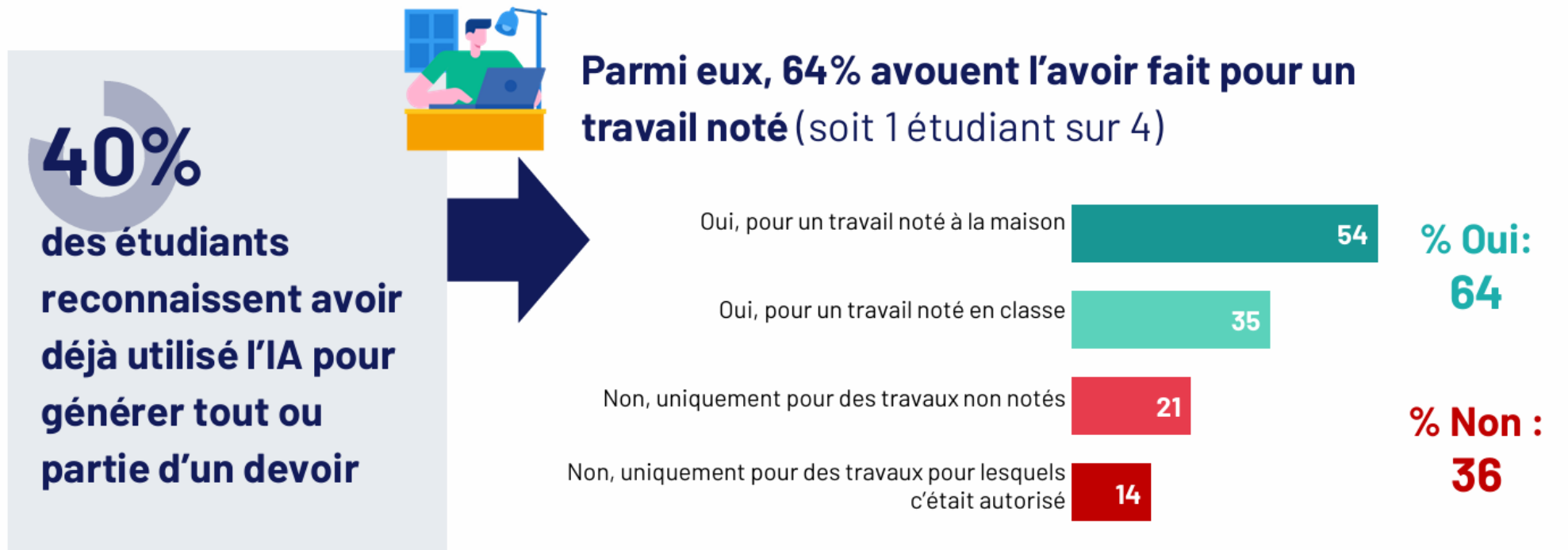


Utilisation dans le cadre des études : des dérives préoccupantes

La majorité de ceux qui utilisent l'IA pour générer des exercices et devoirs le font y compris pour des travaux notés

Question : Et plus spécifiquement, avez-vous déjà utilisé l'IA pour des devoirs notés alors que ce n'était pas autorisé ?

Plusieurs réponses possibles par répondant, total supérieur à 100% (Base : A ceux qui disent l'avoir utilisé pour s'aider pour des devoirs)



Utilisation dans le cadre des études : des dérives préoccupantes

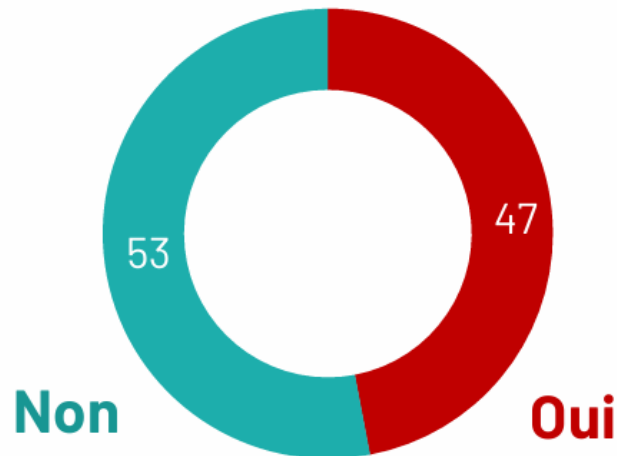
Plus généralement, près de la moitié des étudiants reconnaît s'aider de l'IA pour des exercices même quand cela n'est pas autorisé



Question : Concernant votre formation, chacune des situations suivantes s'applique-t-elle à la vôtre ?

... vous vous aidez de l'IA pour faire vos exercices, même quand cela n'est pas autorisé

Base : A tous



47% se sont déjà aidés de l'IA pour un devoir alors que ce n'était pas autorisé

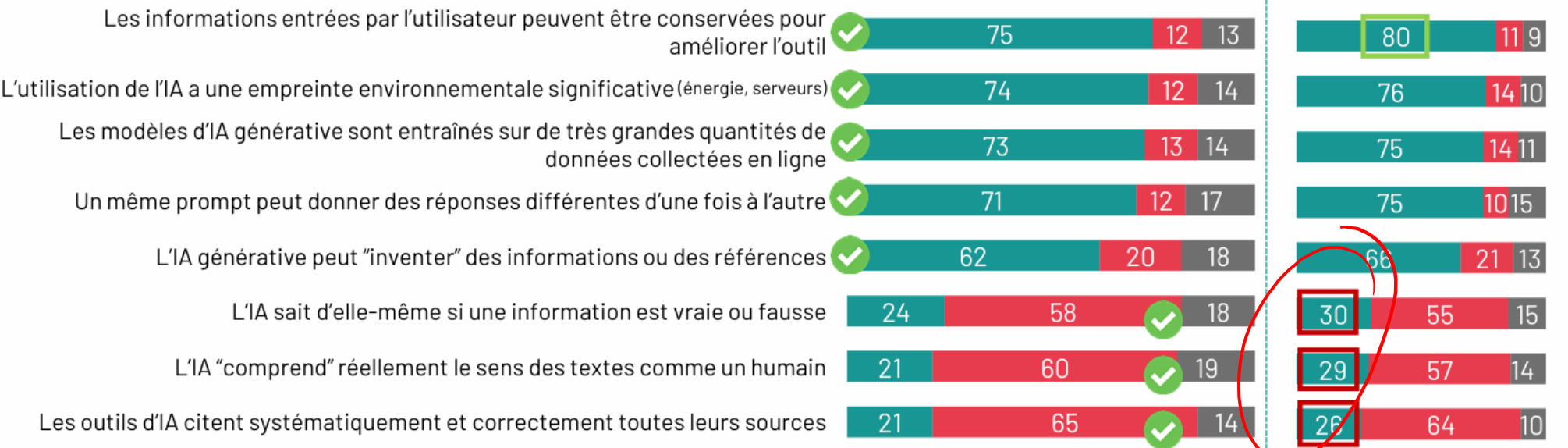
66% des étudiants d'IUT, 65% des étudiants d'école de commerce, gestion et comptabilité, contre 43% des étudiants d'écoles d'ingénieur

Une maîtrise technique, mais des lacunes

Testés plus précisément sur leurs connaissances, ils maîtrisent effectivement les principaux rudiments de l'IA, mais 1 sur 5 pense qu'elle ne peut pas inventer des références ou des informations



Question : Pour chacun des énoncés suivants sur le fonctionnement de l'IA, diriez-vous qu'il est vrai ou faux ?
(Base : A tous)



Vrai Faux Je ne sais pas

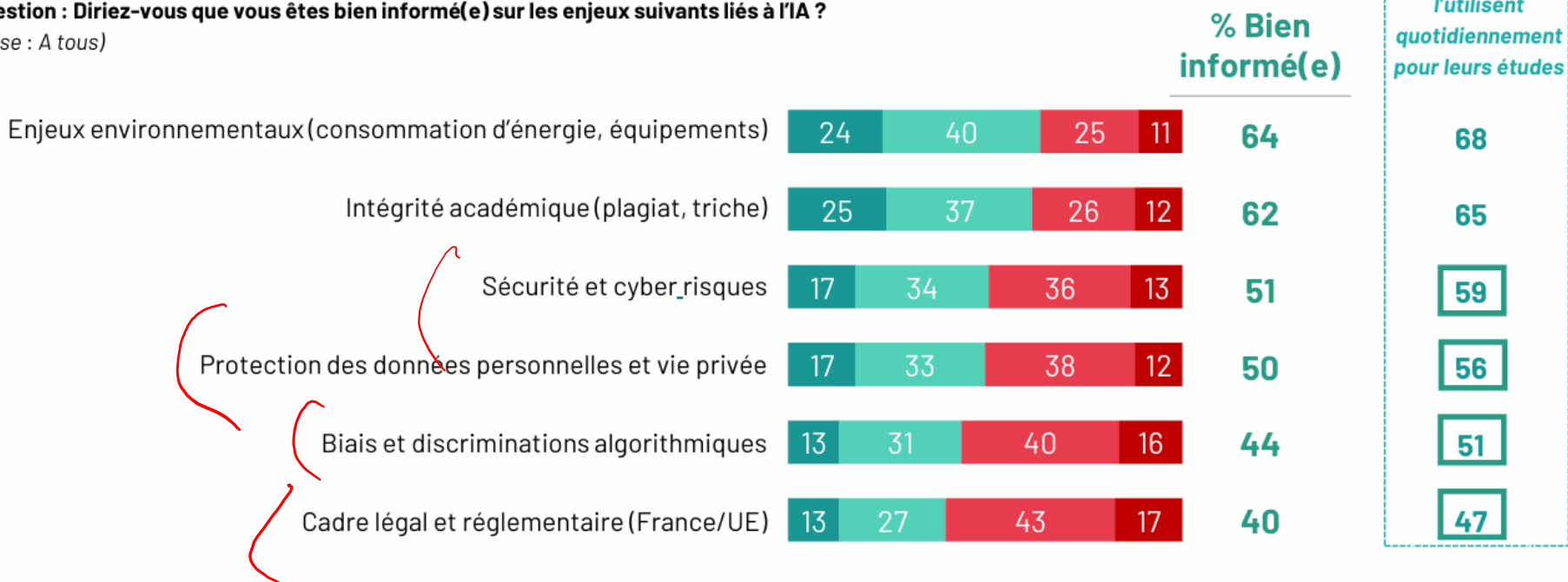


Une maîtrise technique, mais des lacunes sur les enjeux critiques

Plus que de compétences pratiques, les étudiants manquent d'informations sur tous les enjeux autour de l'IA



Question : Diriez-vous que vous êtes bien informé(e) sur les enjeux suivants liés à l'IA ?
(Base : A tous)



81% se disent mal informés sur au moins un enjeu cité, et 56% sur au moins la moitié des enjeux cités

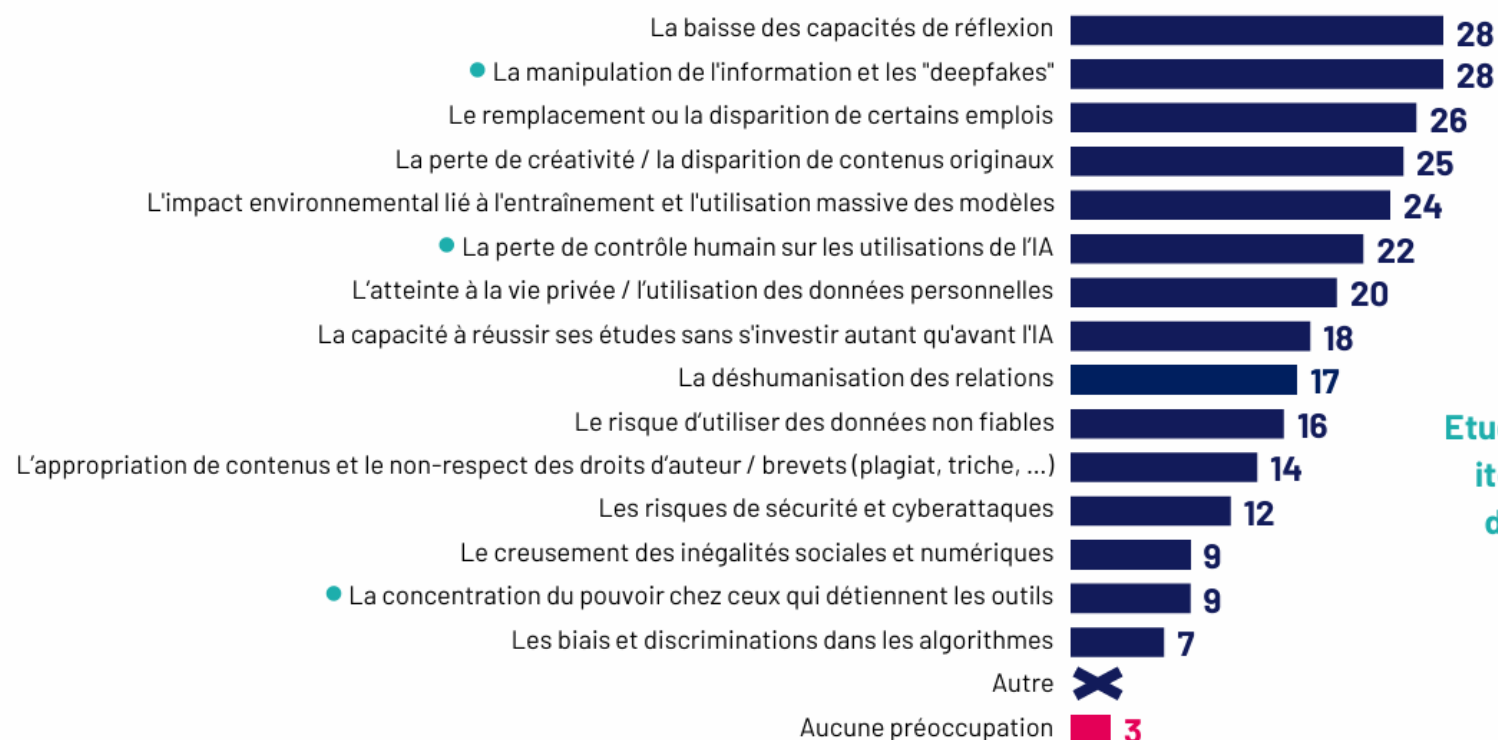
■ Très bien informé(e) ■ Plutôt bien informé(e) ■ Plutôt mal informé(e) ■ Pas du tout informé(e)

Face aux inquiétudes

Les préoccupations autour de l'IA sont très diverses et relèvent aussi bien des baisses de capacité de réflexion et de créativité que des risques de manipulation ou d'impact sur l'emploi

Question : Quelles sont vos trois principales préoccupations concernant l'émergence de l'intelligence artificielle ?

Jusqu'à trois réponses possibles par répondant, total supérieur à 100% (Base : A tous)



Etudiants ayant cité au moins un item portant sur les risques de désinformation et de perte de contrôle : 51%

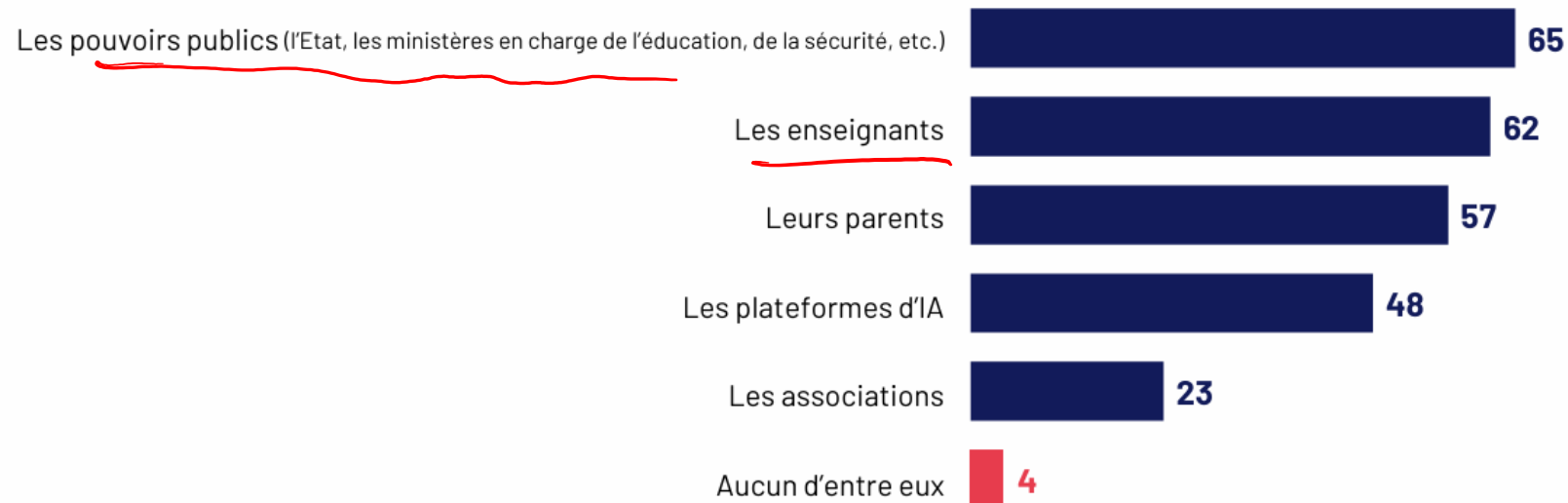


Face aux inquiétudes, une attente d'accompagnement

Dans ce contexte, les étudiants estiment que les pouvoirs publics et les enseignants ont un rôle clé à jouer pour protéger les jeunes



Question : Parmi les acteurs suivants, quels sont ceux qui doivent agir pour protéger les jeunes des risques liés au développement de l'IA ?
Plusieurs réponses possibles par répondant, total supérieur à 100% (Base : A tous)



Éléments de réflexion

Alignement pédagogique (Biggs, Lebrun) à l'heure de l'IAg

- Objectifs → maîtrise des outils ? insertion pro + processus apprentissage
- Activités → individualisation
- Evaluation → devoir maison à oublier ? repenser
- Moyens technologiques → renforcement des oraux

Sujets spécifiques

- Chaine de confiance
- Détecteur d'IAg ?

Fuite de données

Apple a bloqué l'utilisation d'outils d'IA pour certains de ses employés, devenant ainsi la dernière grande entreprise à restreindre l'utilisation de plateformes d'IA générative sur le lieu de travail, en raison de la crainte que les employés ne divulguent des données internes sensibles.

Forbes, 21 mai 2023

<https://www.forbes.fr/technologie/apple-rejoint-une-liste-croissante-dentreprises-qui-repriment-lutilisation-de-chatgpt-par-leurs-employes/>

... Production

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.

Le Monde, 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

... Protection ?

OpenAI is now fighting a court order to preserve all ChatGPT user logs—including deleted chats and sensitive chats logged through its API business offering—after news organizations suing over copyright claims accused the AI company of destroying evidence.

Ars Technica, 06 juin 2025

<https://arstechnica.com/tech-policy/2025/06/openai-says-court-forcing-it-to-save-all-chatgpt-logs-is-a-privacy-nightmare/>

Exploitation publicitaire

OpenAI announced on Friday that it will be testing ads in its free version for logged-in, adult US users. (...) the move comes as OpenAI is urgently trying to figure out how to bring in more revenue

(...) Inserting ads into chatbot conversations could be controversial, given users' sometimes personal and intimate conversations. And the move will amp up the pressure on OpenAI to ensure it doesn't recommend products that could be potentially dangerous or harmful, especially after the company has faced lawsuits alleging that ChatGPT encouraged users' suicides.

CNN, 16 janvier 2026

<https://edition.cnn.com/2026/01/16/tech/chatgpt-ads-openai>

Mise en danger

La Floride attaque OpenAI et son patron Sam Altman pour mise en danger de jeunes utilisateurs de ChatGPT.

Le procureur reproche à OpenAI de ne pas avoir mis en place de règles plus contraignantes pour vérifier l'âge des utilisateurs.

(...) conversations avec ChatGPT en se faisant passer pour un adolescent. Le robot conversationnel a notamment produit des conseils sur la manière de cacher ses habitudes alimentaires à son entourage, préparer un suicide ou s'automutiler de façon «sûre».

Le Figaro, 02 juin 2026

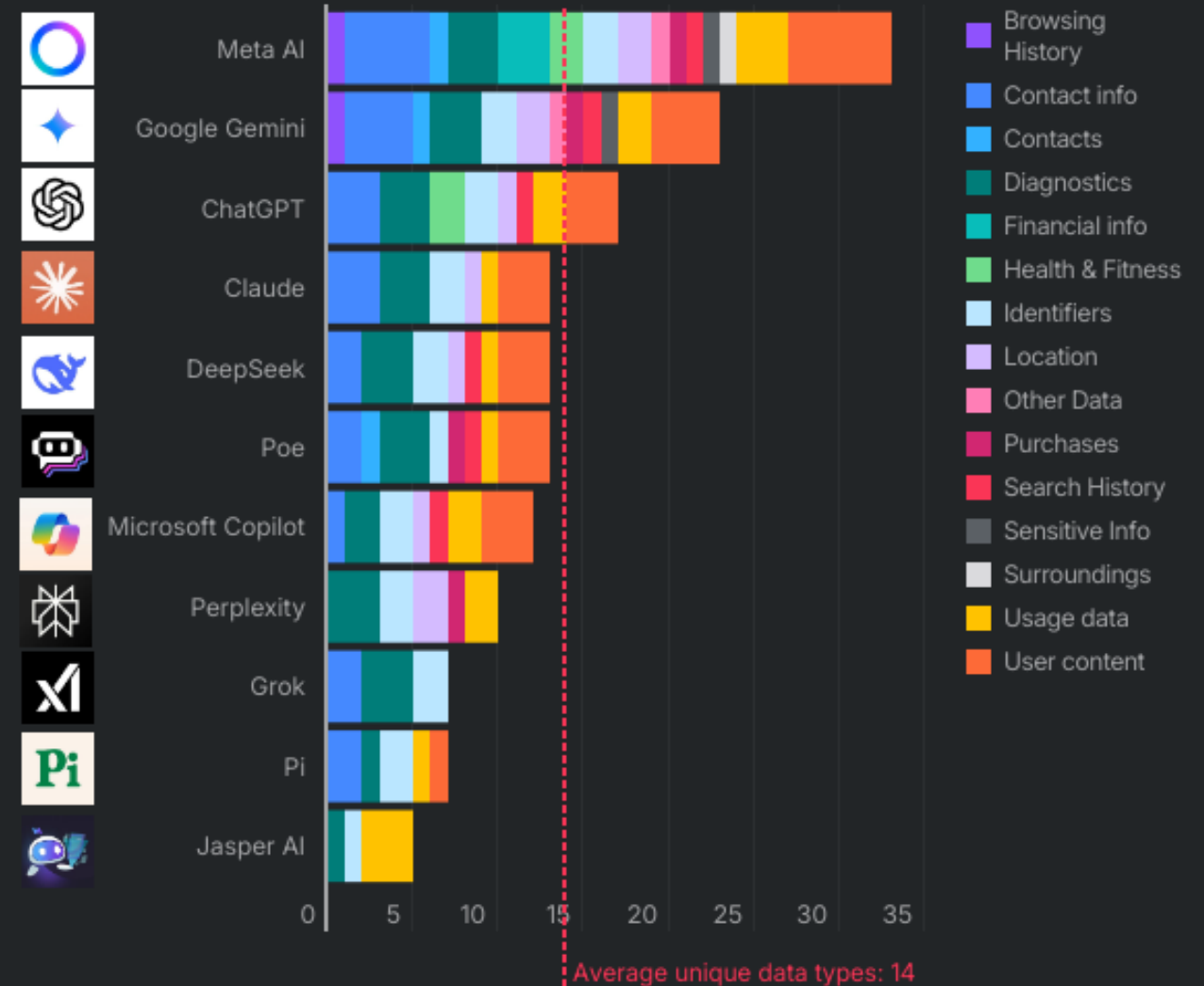
<https://www.lefigaro.fr/secteur/high-tech/la-floride-attaque-openai-et-son-patron-sam-altman-pour-mise-en-danger-de-jeunes-utilisateurs-de-chatgpt-20260601>

Services d'IA générative & collecte de données

DATA UPDATE: MARCH, 2026

Meta AI collects over 90% of user data types

ChatGPT may now collect 17 out of 35 data types, a 70% increase from the 10 data types identified in last year's AI chatbots review

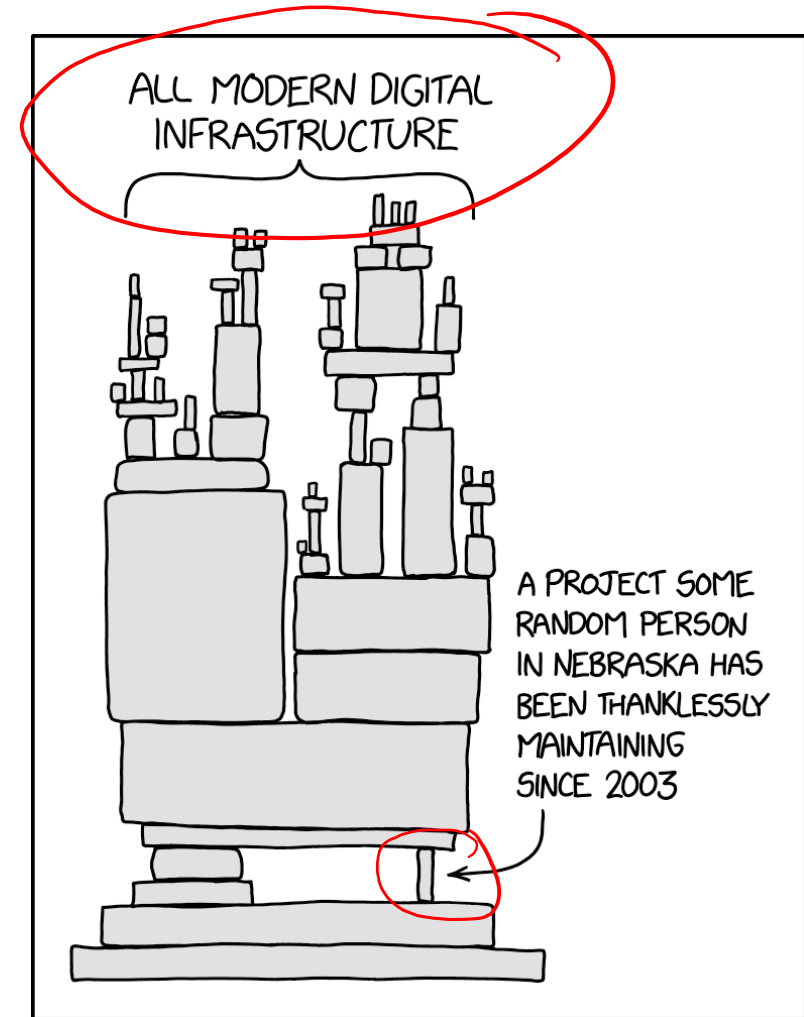


This image is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License - <https://creativecommons.org/licenses/by-nc-sa/3.0/>



Dépendances dans le numérique

- **Choisir et diversifier nos dépendances**
 - Infrastructures et matériel : Etats-Unis, Asie (Taiwan, Chine)
 - Logiciels et services : marché mondial
 - Usages : captation et ancrage par les plateformes
-
- **Quelles dépendances ?**
 - Chaine d'approvisionnement
 - Répartition de la valeur entre acteurs
 - Déplacement des enjeux sur les données



Souveraineté numérique ? Autonomie stratégique ?

Souveraineté numérique

« Maîtrise complète allant du matériel au logiciel »

Avis de la Cour des Comptes : une souveraineté absolue serait « coûteuse et contre-productive »

ou

Autonomie stratégique

« Maîtriser ses dépendances afin d'accroître les choix possibles »

RAPPORT

Souveraineté numérique

De l'urgence d'organiser
la coopération entre le public,
le privé et les communs
numériques



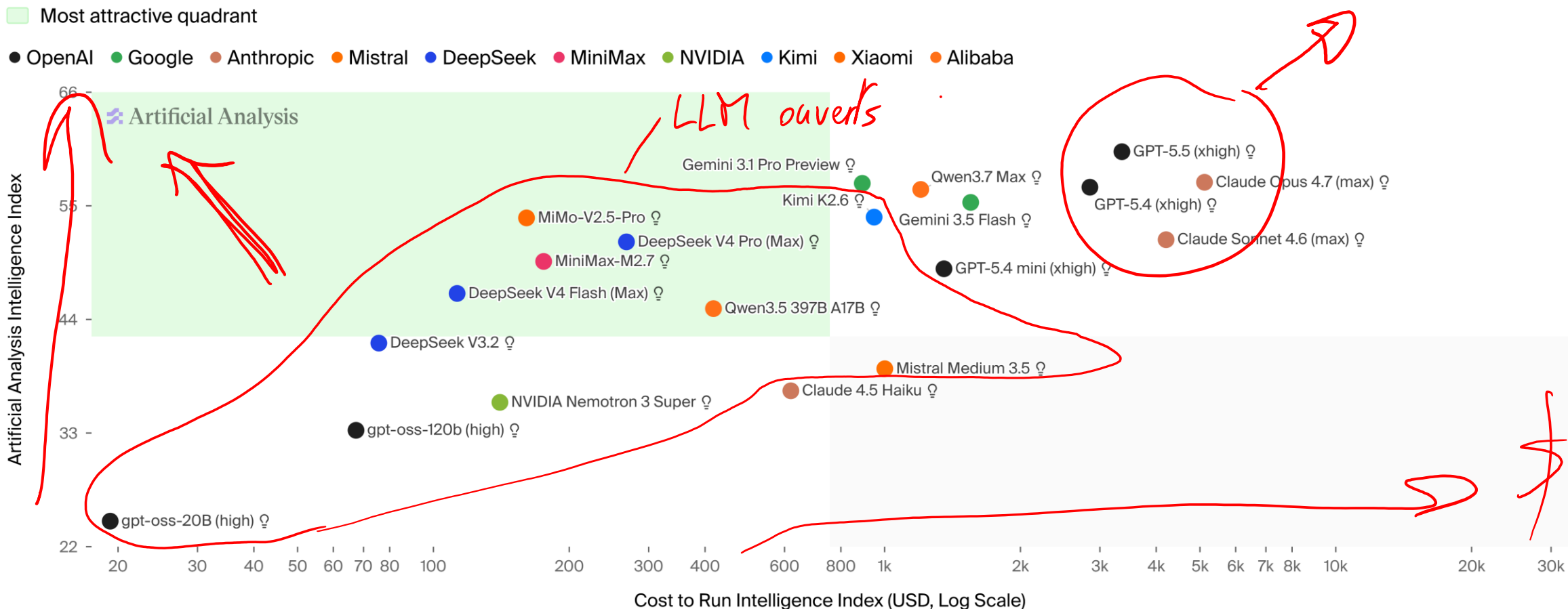
8 piliers : souveraineté numérique (autonomie stratégique)

- **Contrôle et liberté de choix** : capacité d'un utilisateur (Etat, entreprise, citoyen) à prendre ses propres décisions concernant les technologies sans y être contraint par un tiers
- **Réversibilité** : capacité à changer de fournisseur / infra sans coût prohibitif, ni perte de données
- **Diversification** : éviter une dépendance monopolistique, via une stratégie multifournisseurs
- **Sécurité** : protection au juste niveau des données selon leur sensibilité & la réduction des vulnérabilités face aux cyberattaques et à l'espionnage
- **Régulation** : capacité à imposer des règles strictes aux acteurs opérant sur le territoire national ou traitant des données de ressortissants nationaux
- **Sobriété** : éliminer les fonctionnalités inutiles, promouvoir des solutions répondant précisément aux besoins des utilisateurs, afin de concilier usages numériques et engagements climatiques & budgétaires
- **Formation** : soutenir la recherche, l'attraction des talents, et le renforcement des transferts entre la recherche et l'industrie

Modèles : performance & coûts

Intelligence vs. Cost to Run Artificial Analysis Intelligence Index

Artificial Analysis Intelligence Index · Cost to run Intelligence Index



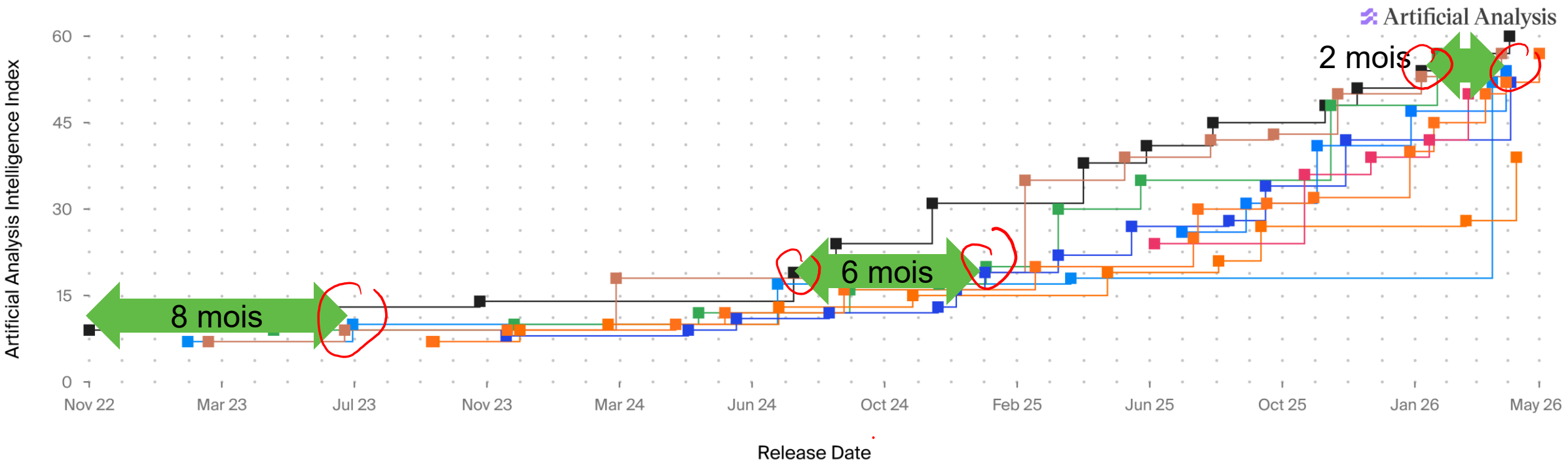
Evolution des performances

Frontier Language Model Intelligence, Over Time

Artificial Analysis Intelligence Index v4.0 incorporates 10 evaluations: GDPval-AA, τ^2 -Bench Telecom, Terminal-Bench Hard, SciCode, AA-LCR, AA-Omniscience, IFBench, Humanity's Last Exam, GPQA Diamond, CritPt

← Délai entre LLM ouvert et propriétaire à « niveau égal » →

● OpenAI ● Meta ● Google ● Anthropic ● Mistral ● DeepSeek ● MiniMax ● Kimi ● Alibaba



Pendant la ruée vers l'or vendez des pelles et des pioches



NVIDIA Corp
NASDAQ : NVDA

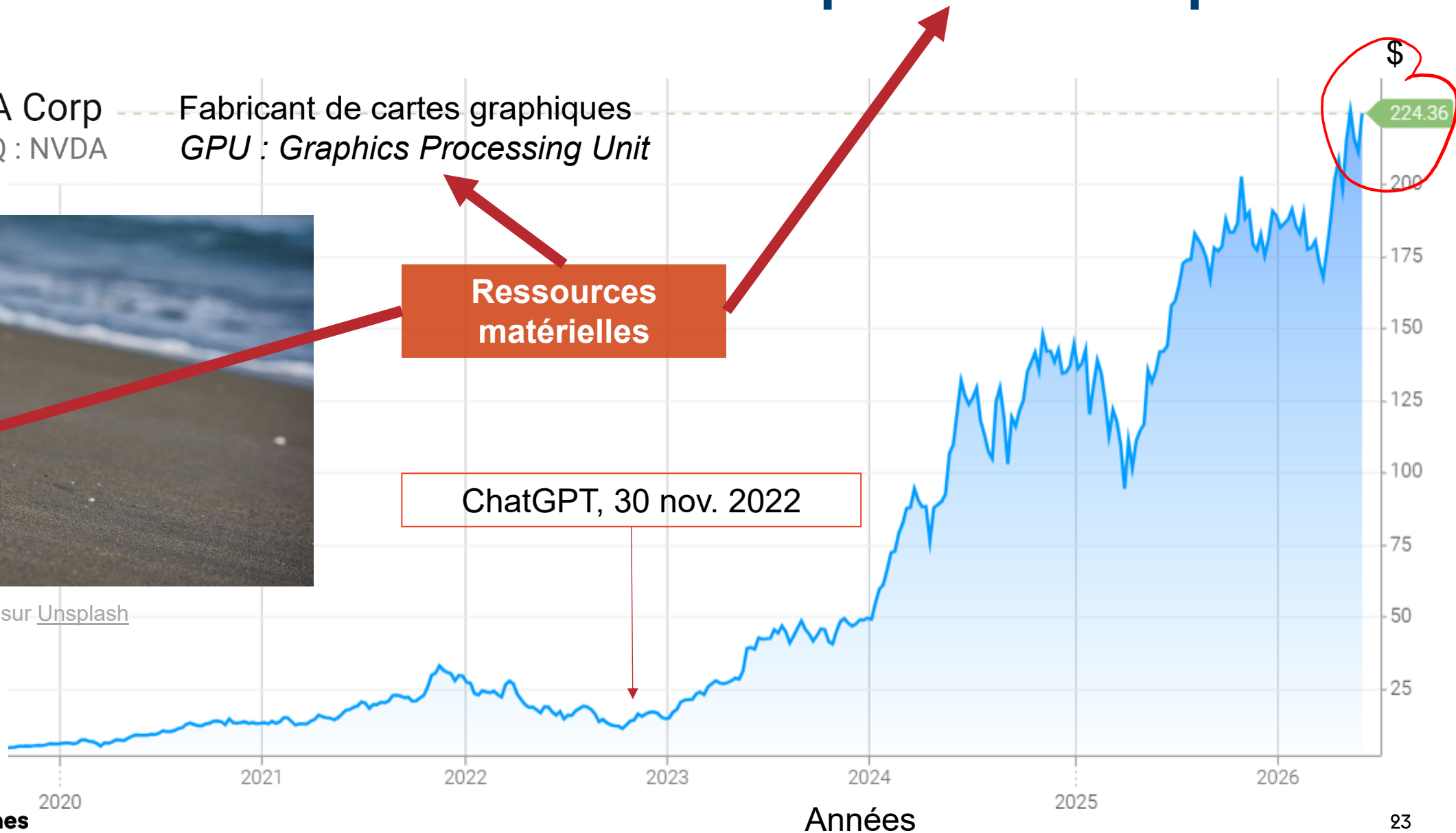
Fabricant de cartes graphiques
GPU : Graphics Processing Unit



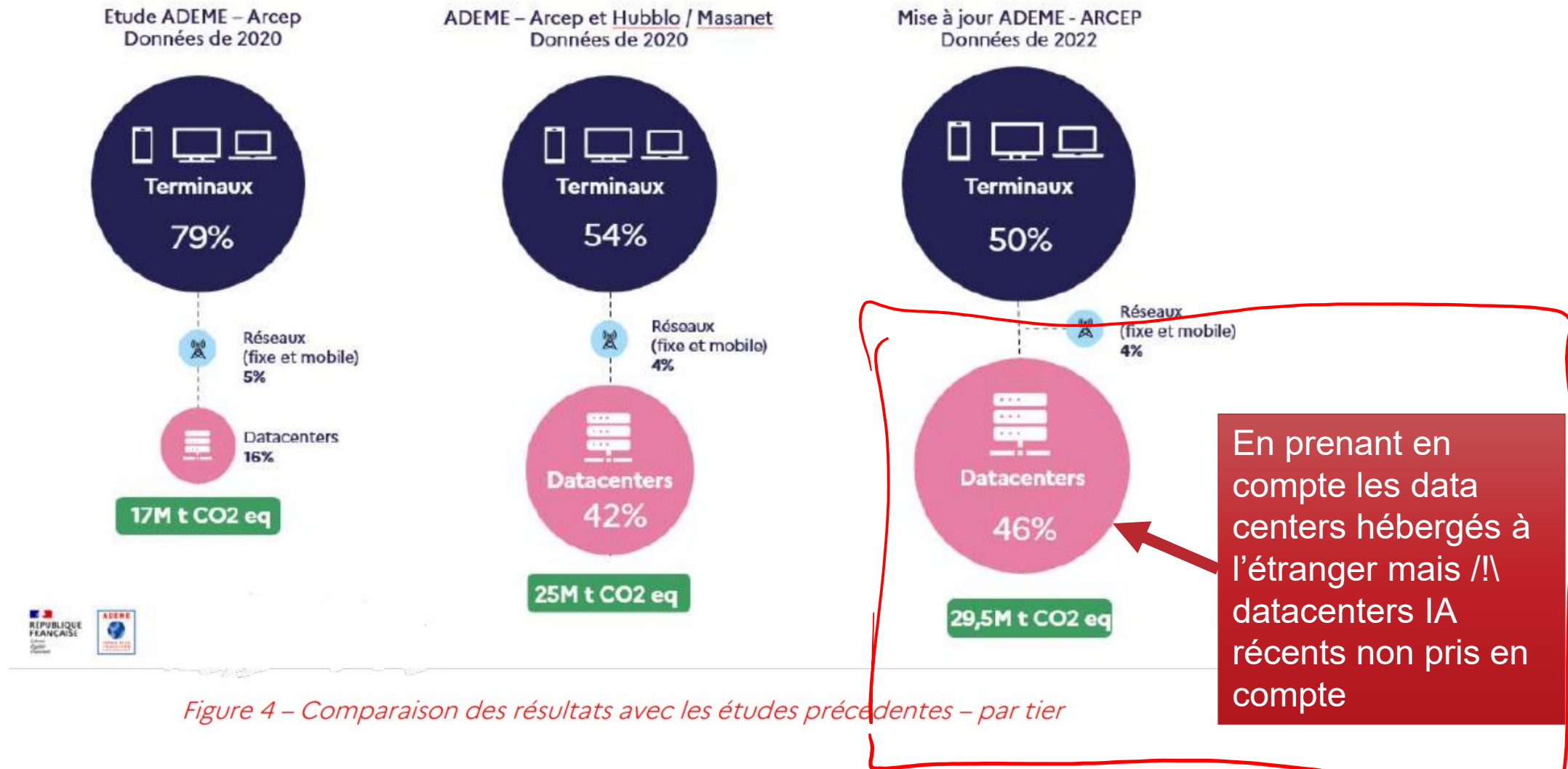
Photo de [Alejandro Alas](#) sur [Unsplash](#)

Ressources
matérielles

ChatGPT, 30 nov. 2022



Impact environnemental du numérique en France (actualisé)



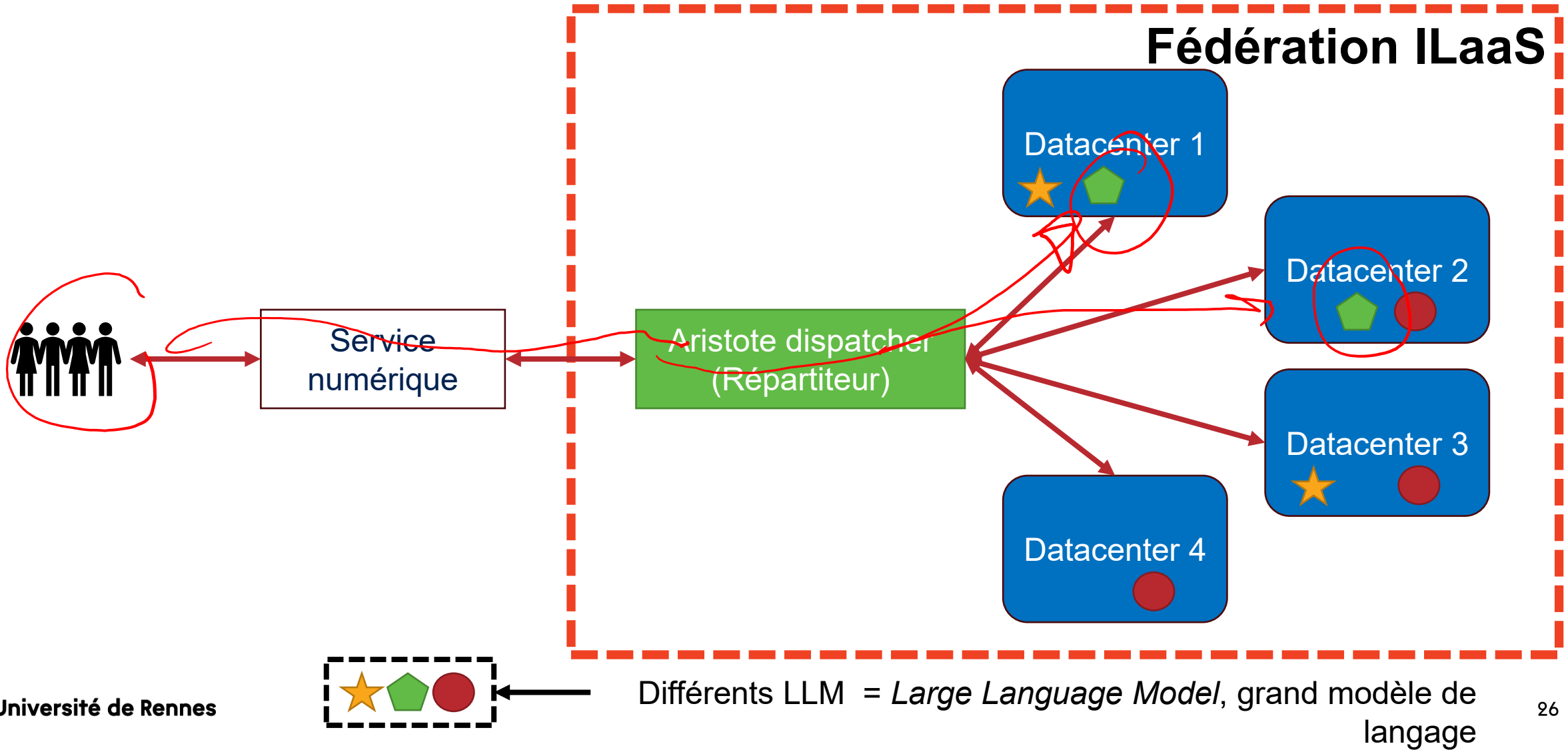
Robustesse

**Capacité d'un système à se maintenir
sa stabilité (à court terme)
et sa viabilité (à long terme)
malgré les fluctuations**

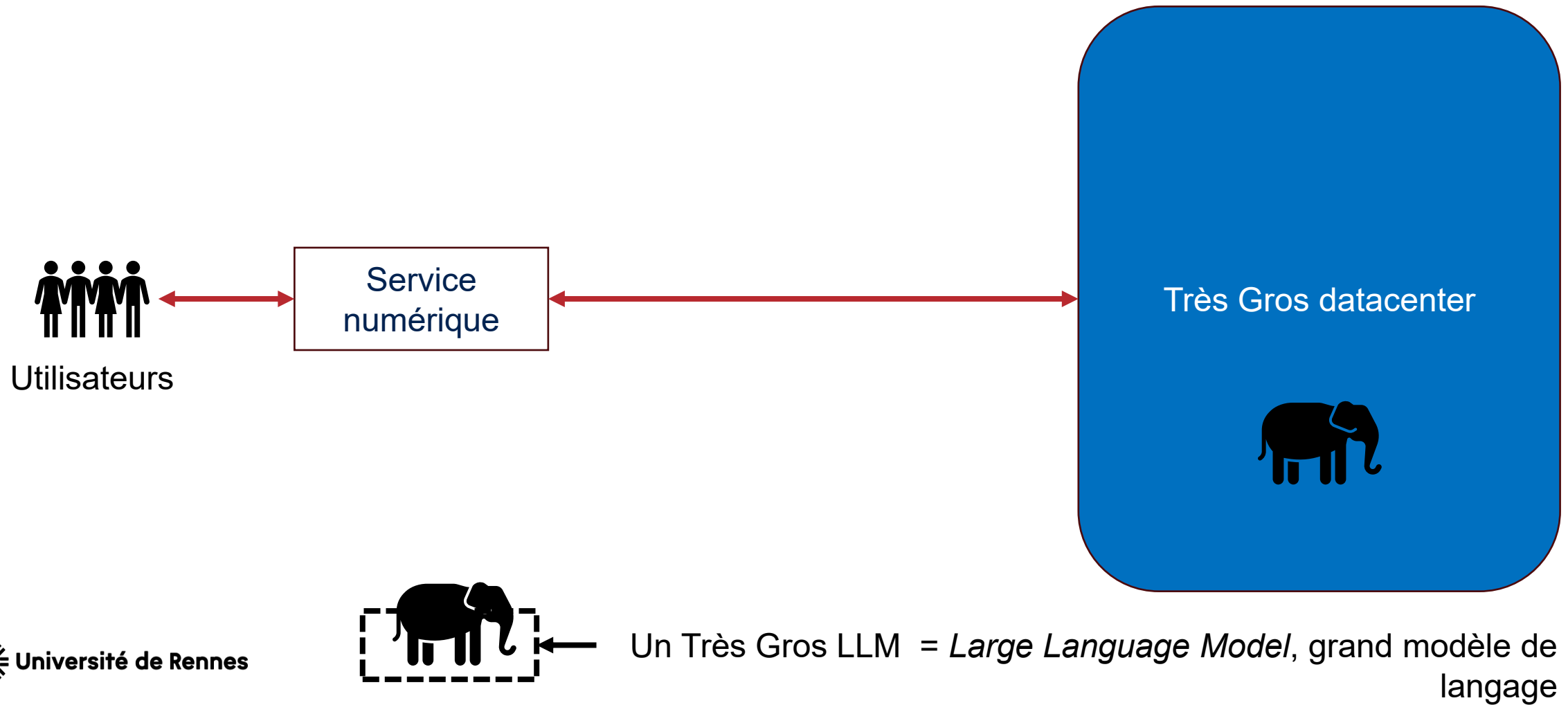
Robustesse – Institut Michel Serres



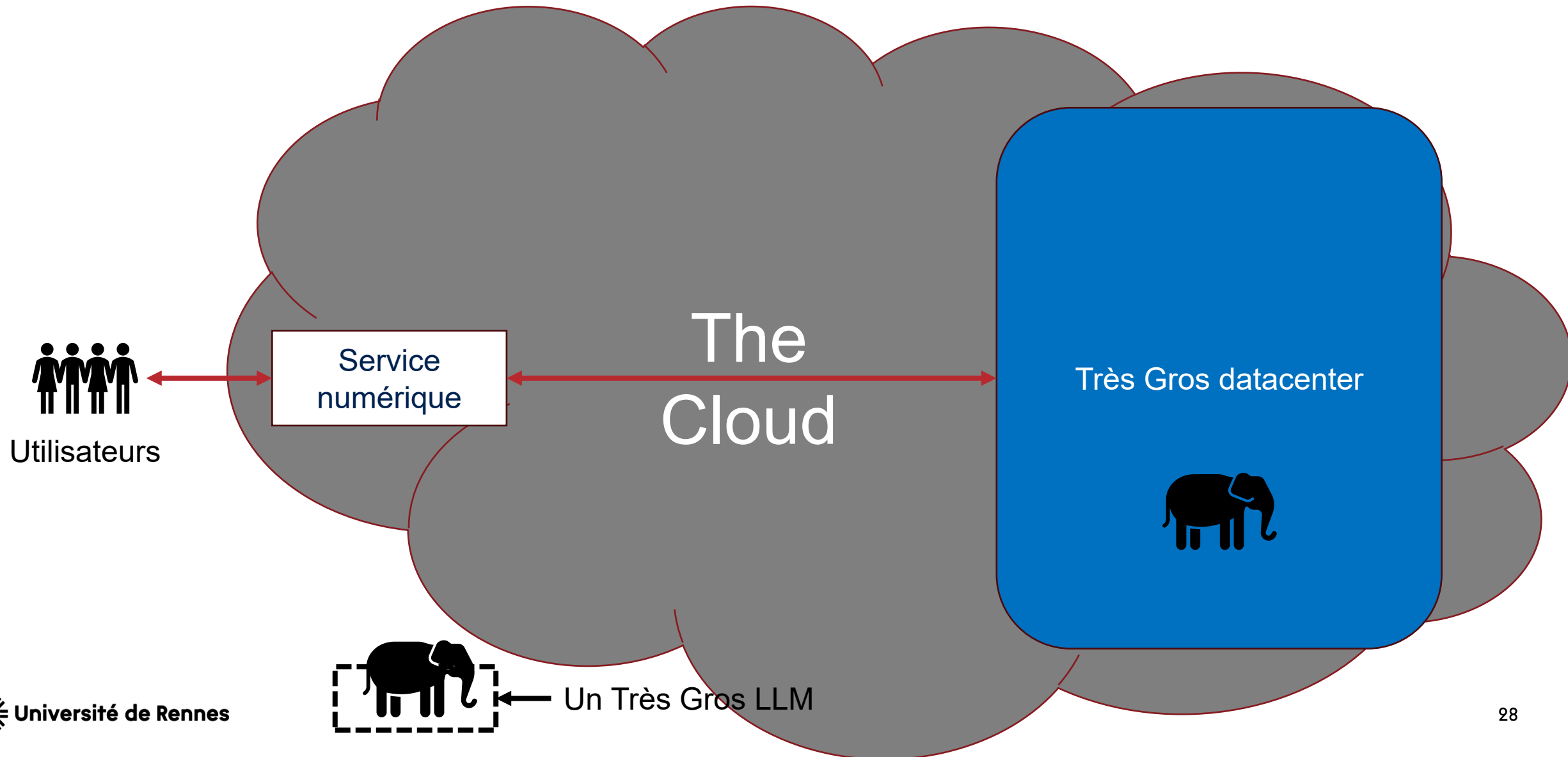
ILaaS : Inférence de LLM « as a Service »



Ce serait beaucoup plus PERFORMANT comme ça, non ?



Ce serait beaucoup plus PERFORMANT et PRATIQUE, non ?



« Qui peut le plus peut le moins »

Oui, ce serait PERFORMANT

mais est-ce pertinent ?



Photo de [Adrien Olichons](#) sur [Unsplash](#)



Photo de [personalgraphic.com](#) sur [Unsplash](#)



Photo de [Matthew Alexanders](#) sur [Unsplash](#)

Photo de [Nathan Dumlaos](#) sur [Unsplash](#)

Photo de [Lucas de Moura](#) sur [Unsplash](#)



One size fits all, really ?



17/06/2026

31

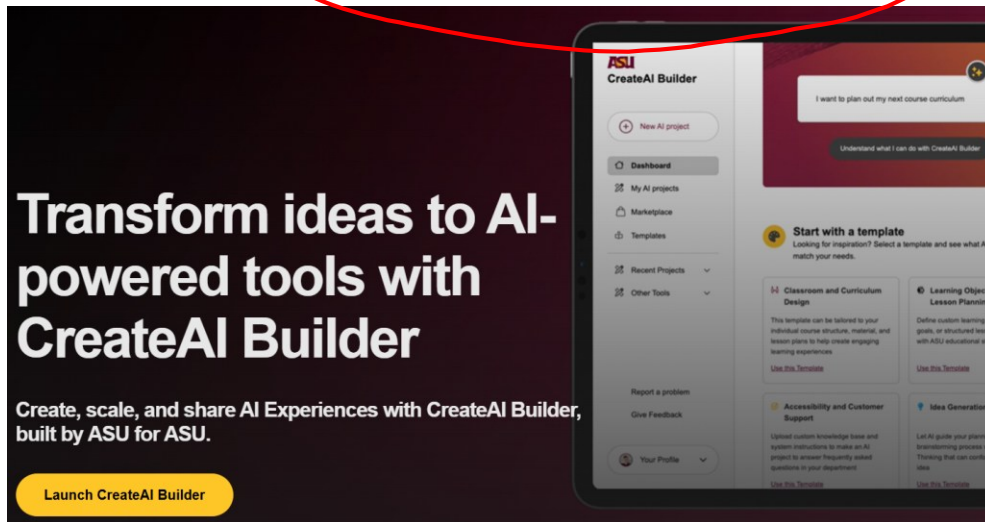
Exemples à l'international

Solutions locales pour une ou plusieurs universités

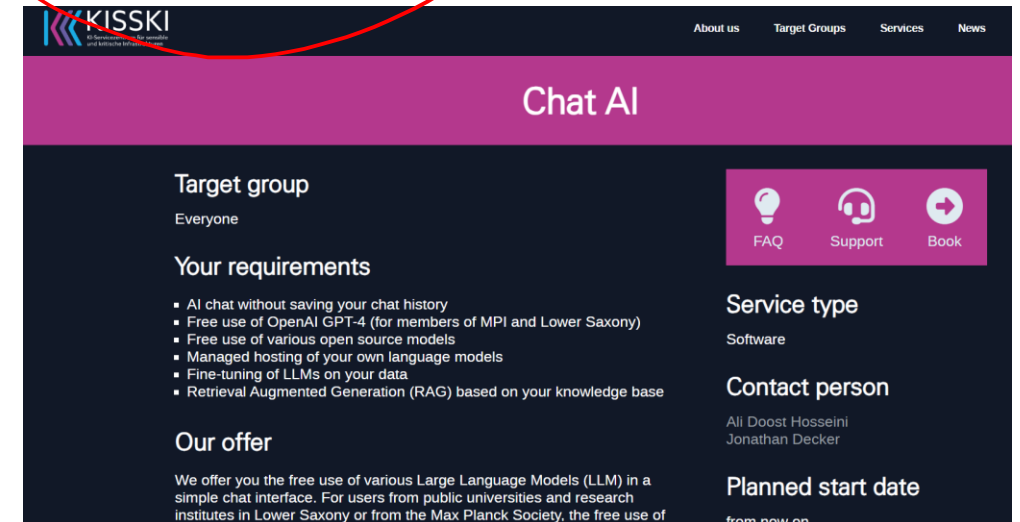
Interface web, accès API

LLM open weights & propriétaires

Etats-Unis, Arizona State University
CreateAI Builder ai.asu.edu/ai-tools



Allemagne, KISSKI/GWDG Chat AI
SLURM for HPC Services





Objectif du projet

**Mutualiser les infrastructures
numériques des partenaires pour
les mettre à disposition
d'utilisateurs bénéficiaires**

Offre portail apps.ilaas.fr

Vue publique & authentifiée (fédération Renater)

Applications

- open source ou commerciales
- Utilisant les inférences ILaaS

Plateformes

- mutualisées (ex: Dialogue) : raccordement par permissions
- instances dédiées (ex: n8n) : déploiement par organisation

Traitement des demandes

- Validation politique par établissement (VP-Num)
- Validation du GT-Projet

Applications disponibles sur ILaaS

Connexion universitaire



Dialogue

Publiez des bots (chat, mail, voix) combinant arbres guidés et agents d'IA en utilisant un éditeur visuel.



LibreChat

Tchatez avec un bot personnalisable et interopérable



Aristote

Générez titres, mots-clés et quiz en plusieurs langues à partir de vos vidéos Pédagogiques



n8n

Automatisez vos workflows en connectant services et API via une interface visuelle low code



OI

Rag@Renn

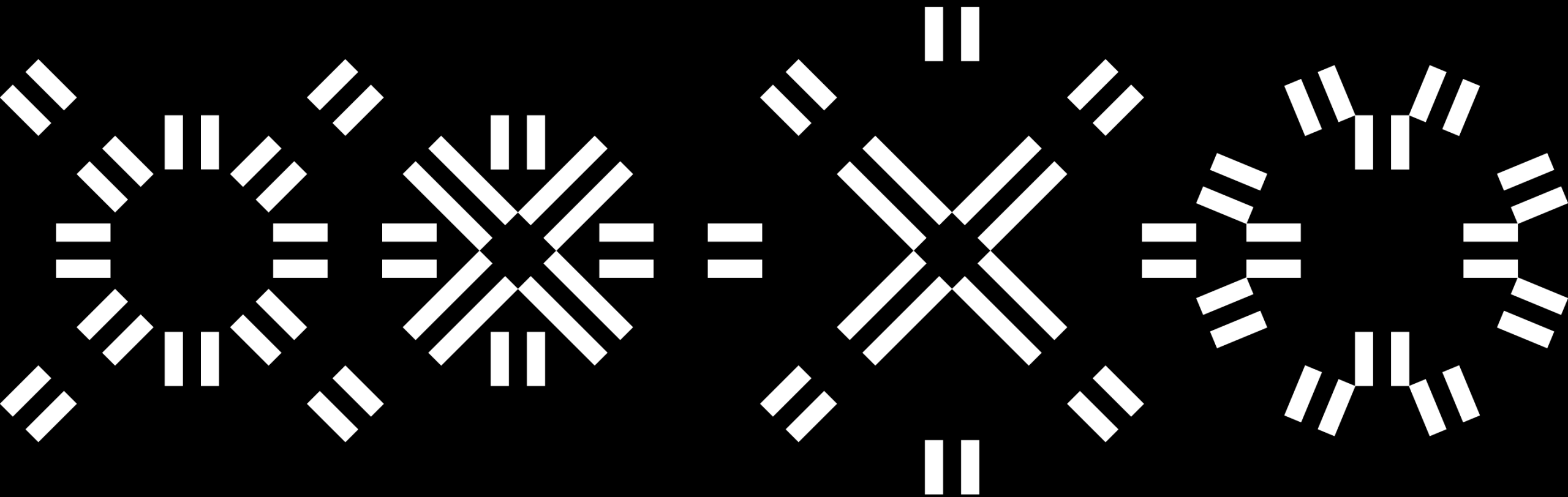
Travaillez sur vos documents avec un bot intégrant assistants et RAG (Retrieval Augmented Generation)



TIPS (Technologie Interactive pour Pratiques de Scénarios)

Dynamisez la méthode des cas et engagez vos étudiants dans une expérience d'apprentissage interactive et moderne.

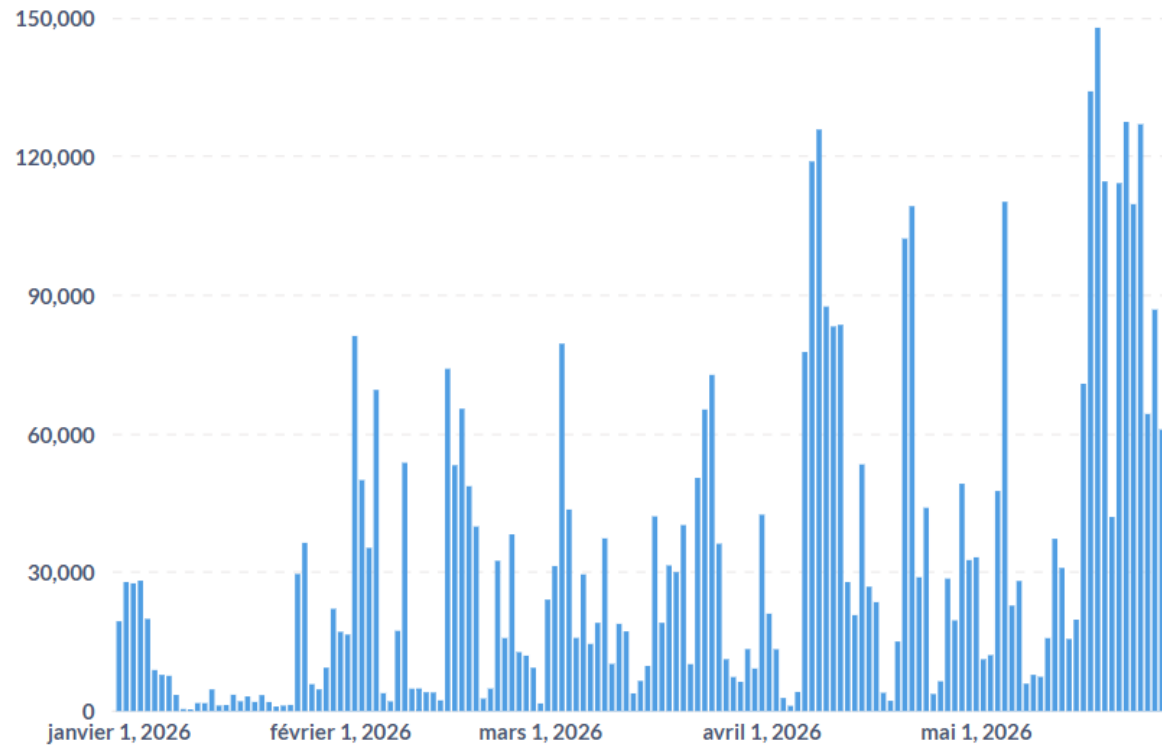




Usages constatés (mai 2026)

Nombre de requêtes par jour (150k max par jour)

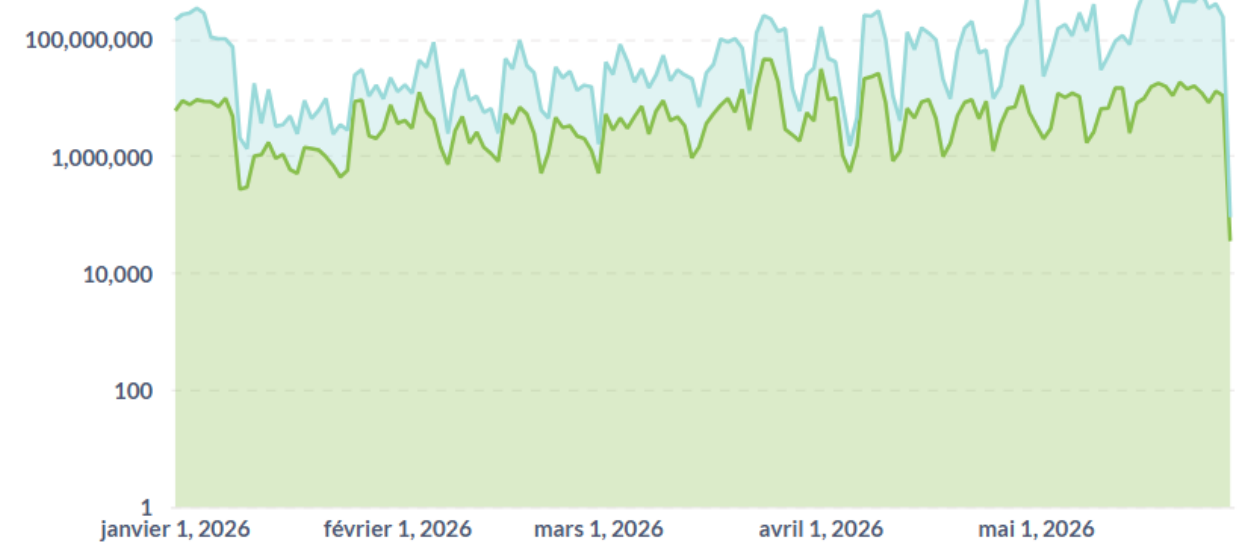
Du 1^{er} janvier au 28 mai 2026



Nombre de tokens en entrée et en sortie (max par jour : ~ 834M entrée / 45M sortie)

Du 1^{er} janvier au 28 mai 2026

Echelle logarithmique



Feuille de route 2026

Structuration

- Accord de consortium entre 24 établissements : Université de Rennes, Université Reims Champagne-Ardennes, Université Paris 1 Panthéon-Sorbonne, Université Paris Saclay, Université de Lille, Université Côte D'Azur, Université Paris Cité, Université de Tours, Université d'Orléans, Université de Technologie de Compiègne, Université de Strasbourg, Université de Lorraine, La Rochelle Université, Université de Picardie Jules Verne, Université Clermont Auvergne, Institut Mines-Telecom, Nantes Université, Université de Haute Alsace, CRIANN, Université d'Artois, Inria, Hcéres, Renater, CY Cergy Paris Université, Université Gustave Eiffel
- Consortium ouvert au fil de l'eau
- Modèle économique, facturation, dimensionnement de l'infrastructure cible
- Consolidation infrastructure socle

Portail applications ILaaS : accès à un large panel d'applications IA

- Logiciels EdTech sur des serveurs maîtrisés (ex : Nolej ; Pyxiscience, ...)
- Solutions Open Source développées au sein de l'écosystème ESR
- Environnement de développement agentiques (N8N, Dialogue, ...)

Recommandations institutionnelles

« Une **articulation avec la stratégie adoptée dans l'enseignement supérieur** et visant à la mise en place de partenariats privilégiés avec une IA souveraine est une piste à explorer (...) les **enjeux de souveraineté nationale sont indissociables de la mise en œuvre de solutions** d'IA, et plus particulièrement d'IA générative. »

Rapport « L'IA dans les établissements scolaires », mai 2025 (p.2 et 47)

« Recommandation 17 (MENESR, Établissements, Région, SGPI) : Développer des **data centers dédiés à l'inférence souveraine** de modèles d'IA, à l'instar de ceux dédiés à la recherche, et porter cette réalisation au niveau européen.

Ce développement **doit se faire en complément et en coordination avec le regroupement Ilaas** (acronyme pour inference LLM as a service) qui vise à faciliter, mutualiser les capacités de calcul des ESS afin de distribuer les requêtes d'inférence sur les ressources disponibles. Pensé dans une **logique d'équité d'accès**, un tel dispositif doit permettre de **réduire les inégalités de déploiement de l'IA** entre établissements. »

Rapport « L'IA dans l'ESR », juillet 2025 (p.76)

Enjeux stratégiques : infrastructure mutualisée

Soutenabilité : numérique responsable

- Optimisation de l'infrastructure et des services : utiles, utilisés, utilisables
- Accompagnement des usages vertueux
- Modèle économique soutenable

Sécurité

- Niveau de confiance partagé dans l'ESR
- Offre mutualisée autour des datacenters régionaux labellisés

Souveraineté

- Maîtrise du déploiement et du traitement LLM
- Maintien et développement des compétences techniques en IA

Résilience et robustesse

- Qualité de service, lissage des pics, gestion des indisponibilités
- Distribution régionale & lien avec les centres nationaux (CINES)

Offre API ilaas.fr (documentation sur le site)

Inférence LLM par API llm.ilaas.fr gateway compatible API OpenAI

- LLM
 - ❖ 6 open weights : Gemma 4 31B, Mistral Small 4 119B, Qwen 3.6 35B A3B
 - ❖ Arrêtés fin septembre 2026 : GPT oss 120B, Llama 3.x (8B et 70B), Mistral Small 3.2 24B,
 - ❖ 1 commercial : Mistral Medium 3
- Haute-disponibilité : répartition multi datacenters
- Répartition (charge, file d'attente)
- Accès prioritaire pour les contributeurs de ressources (coupe-file)
- Confidentialité : pas d'accès au contenu des requêtes (simple relais)

Transcription Speech-to-text (STT) [asynchrone stt.ilaas.fr](https://stt.ilaas.fr)

- Basée sur Whisper
- Workers distribués dans les DC ILaaS
- Fonctionnalités : génération de métadonnées (titres, résumés, mots-clés), traduction, sous-titrage paramétrable (nb de lignes, nb de caractères), génération de Quiz et comptes-rendus
- Confidentialité : suppression après traitement (vidéos, transcripts), anonymisation de l'utilisateur final

Embeddings : www.ilaas.fr/rag-ilaas swagger rag-api.ilaas.fr/docs

- Modèles bge-m3 (emdeddings) et bge-reranker-v2-m3 (reranker)

Projet : genèse & avancement

Lancement du projet fin 2024 à 5 établissements

Point d'étape mars 2026 :

- 211 clés distribuées pour des services opérés par 88 établissements
- Fédération opérée par 6 établissements dans leurs Data Centres Labellisés en Région de l'ESR



Capacité	Matériel	Répartition
Socle permanent	8 H200 répartis sur 4 nœuds	4 nœuds
Variable selon besoins et disponibilité	8 RTX 6000 Pro + 4 L40(s) 16 H100 + 11 H200	16 nœuds
Prochainement (sous réserves vu le contexte d'inflation)	8 B300 14 H200 ou équivalent	A déterminer

Partenariats & soutiens

AMUE : projet IA souveraine

- Contribution technique
- Offre API Mistral Medium (ILaaS / Rennes)
- Plateforme « Le Chat » (CINES & ILaaS)

Esup

- Participation aux journées techniques
- Collaborations atelier IA

MESRE

- Concertation stratégique
- Soutien financier (fonds d'amorçage CoSIN, DGRI)
- Webinaires en commun IA Sup (DGESIP)



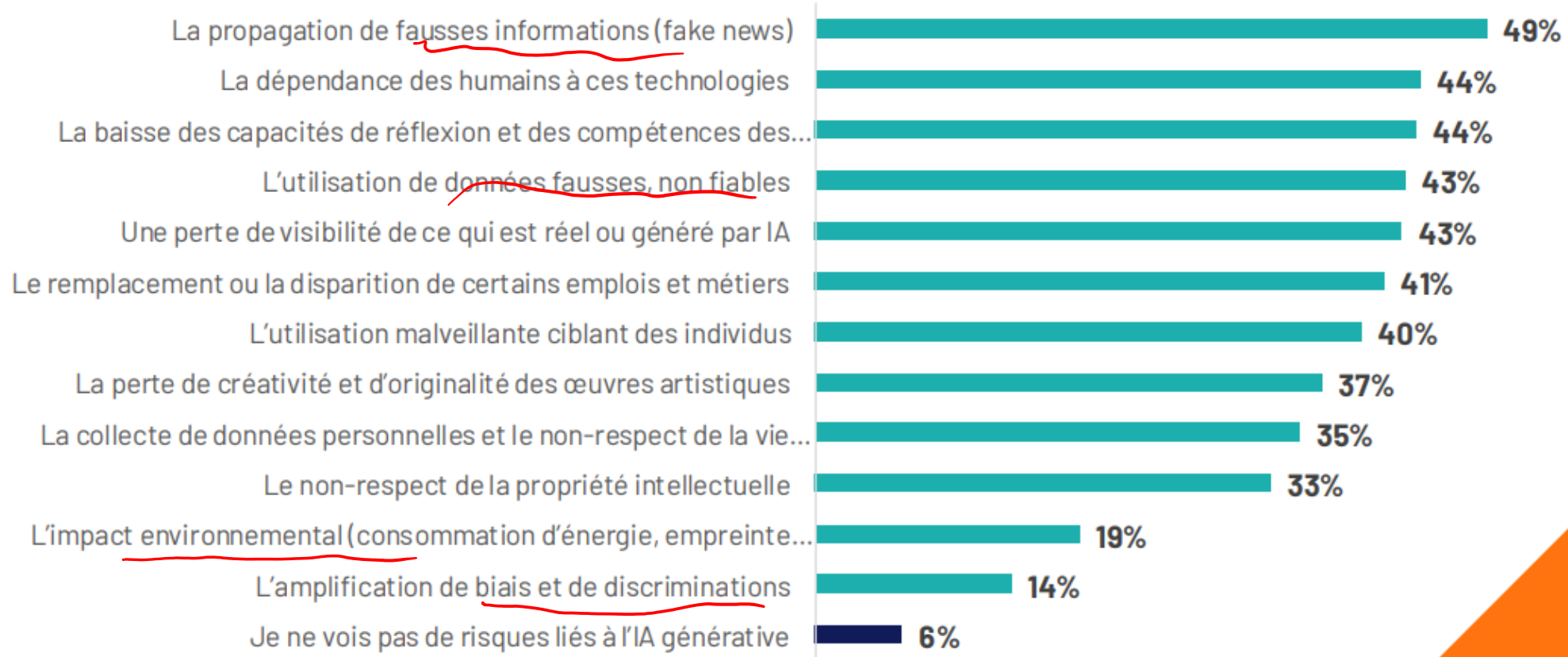
Enjeux de l'IA générative

- Données
- Impact environnemental
- Dépendances & souveraineté



Etude Ipsos « L'usage de l'IA par les français » fév. 2025

Selon vous, quels sont les principaux risques liés à l'utilisation des IA génératives ?



Base = 155 utilisateurs professionnels / 328 utilisateurs privés

[L'usage de l'intelligence artificielle par les Français](#), février 2025

Impact environnemental du numérique en France

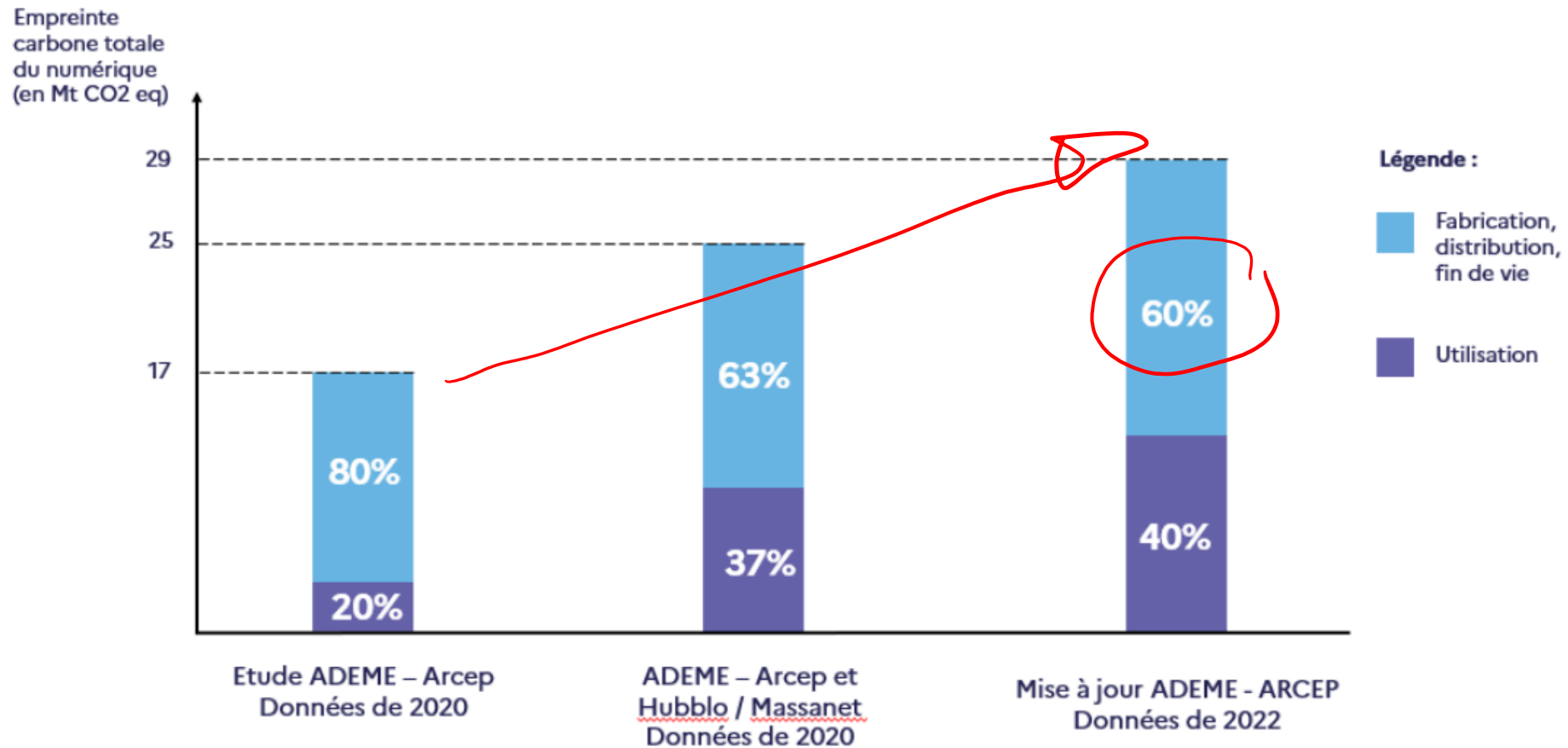


Figure 5 – Comparaison des résultats avec les études précédentes – par phase du cycle de vie

Trajectoire énergétique : prospective

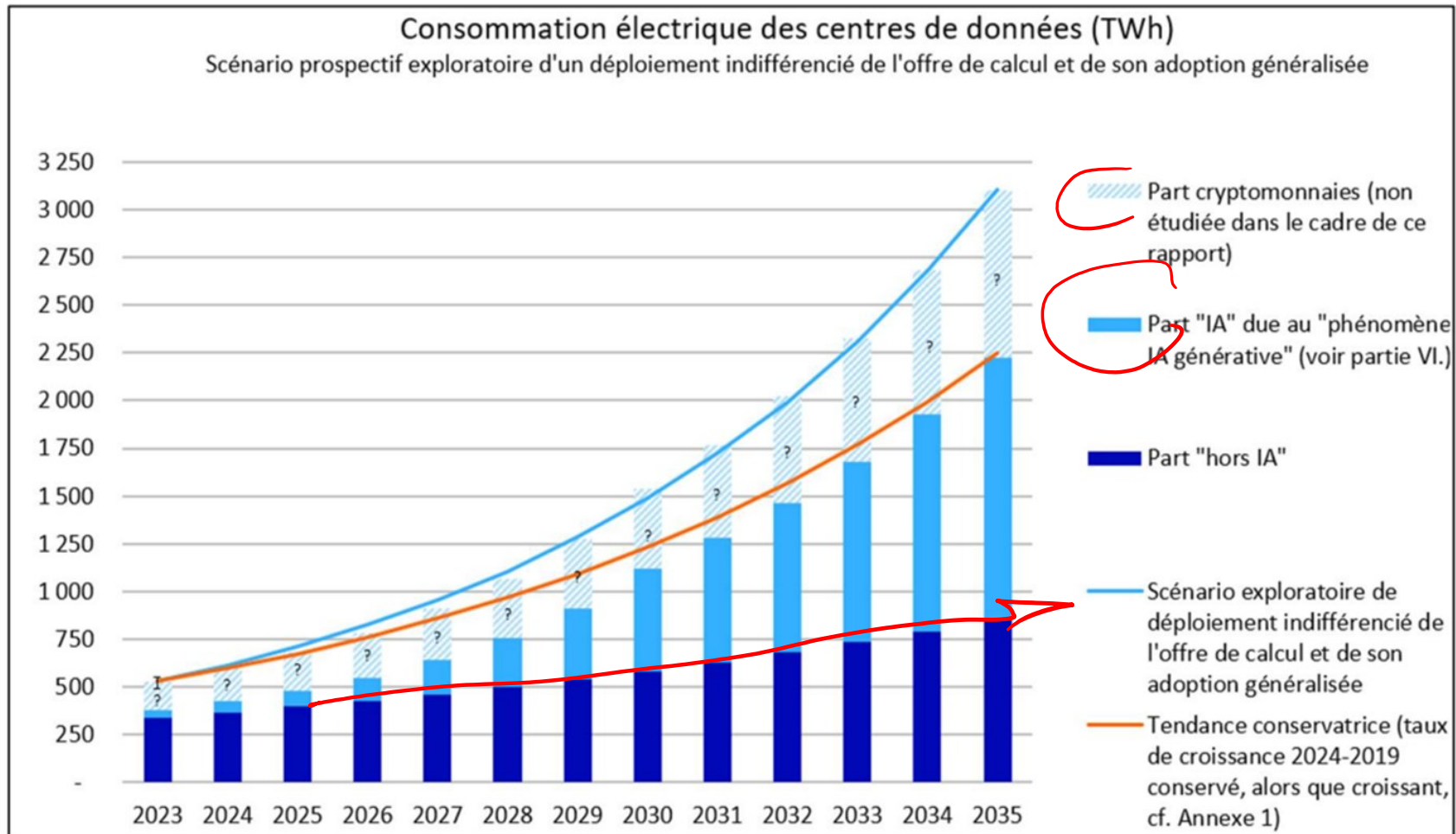


Figure 10 - Consommation d'électricité des centres de données en phase d'usage (TWh) de notre scénario prospectif et exploratoire d'un déploiement indifférencié de l'offre de calcul et de son adoption généralisée en comparaison à une tendance conservatrice. Source : (The Shift Project, 2025a)

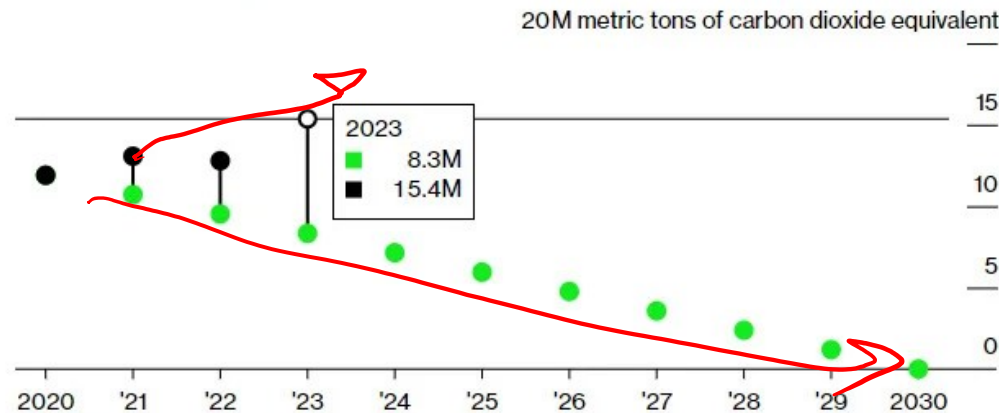
Microsoft's AI Push Imperils Climate Goal as Carbon Emissions Jump 30%

“The company’s goal to be carbon negative by 2030 is harder to reach, but President Brad Smith says the good AI can do for the world will outweigh its environmental impact.”

Microsoft's Emissions

Artificial intelligence is putting the tech giant's climate goals in peril

● Climate plan (simulated) ● Actual



Source: Microsoft (Scope 1, 2 and 3 "management criteria" data)

Note: Green dots represent linear decline to carbon negative goal.

<https://www.itforbusiness.fr/europe-chine-carbone-des-nuages-gris-au-dessus-de-microsoft-76735>

<https://financialpost.com/pmnbusiness-pmn/microsofts-ai-push-imperils-climate-goal-as-carbon-emissions-jump-30>

Limites

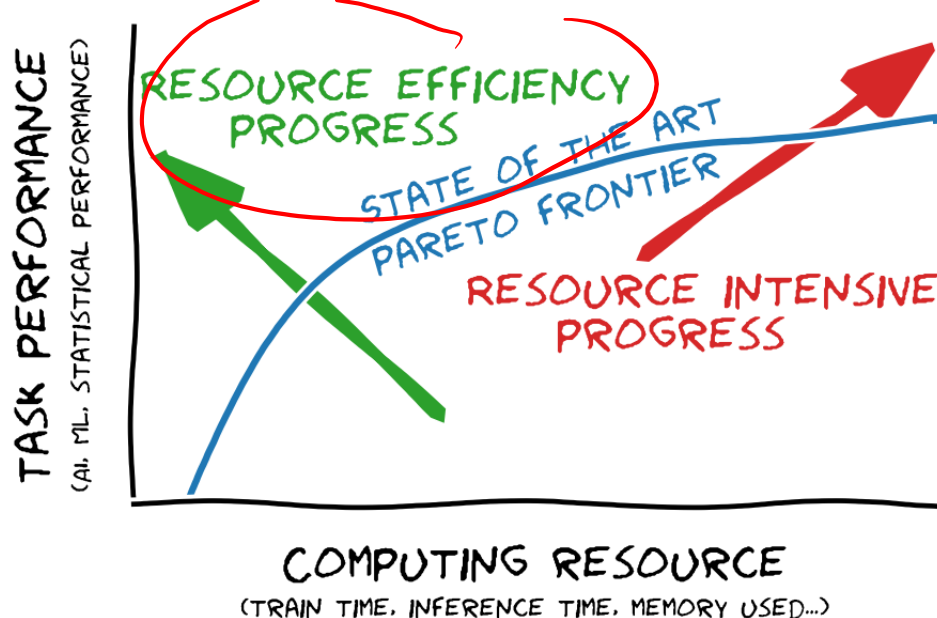


Figure 1. An explosion in model size – Left: The increase in model size means it is more and more expensive to run them in terms of RAM. Right: resources we need are increasing faster than available compute. Data from Epoch (2023), specific details in Appendix A.

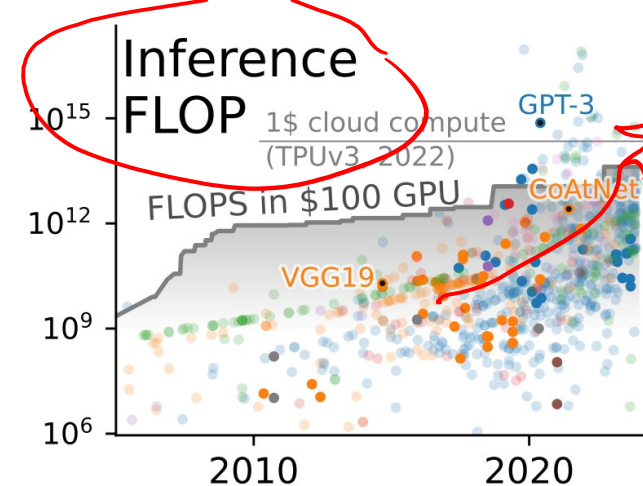
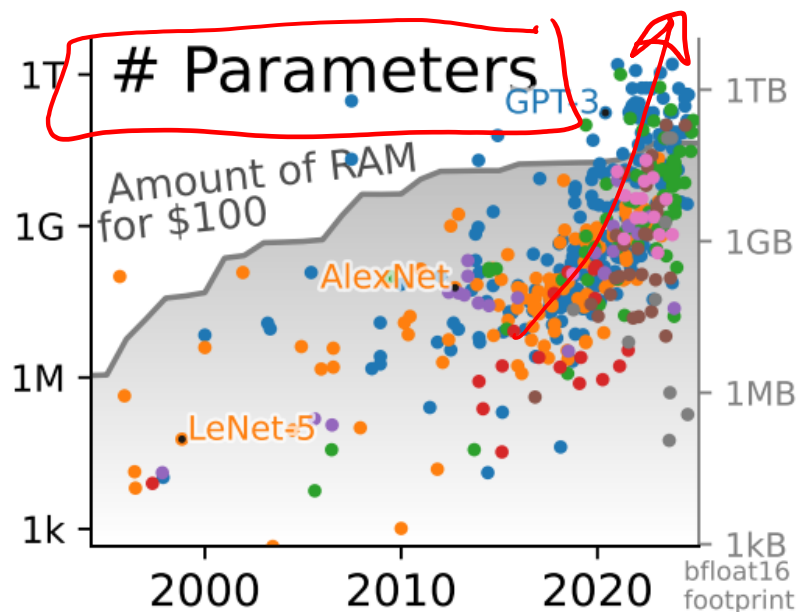
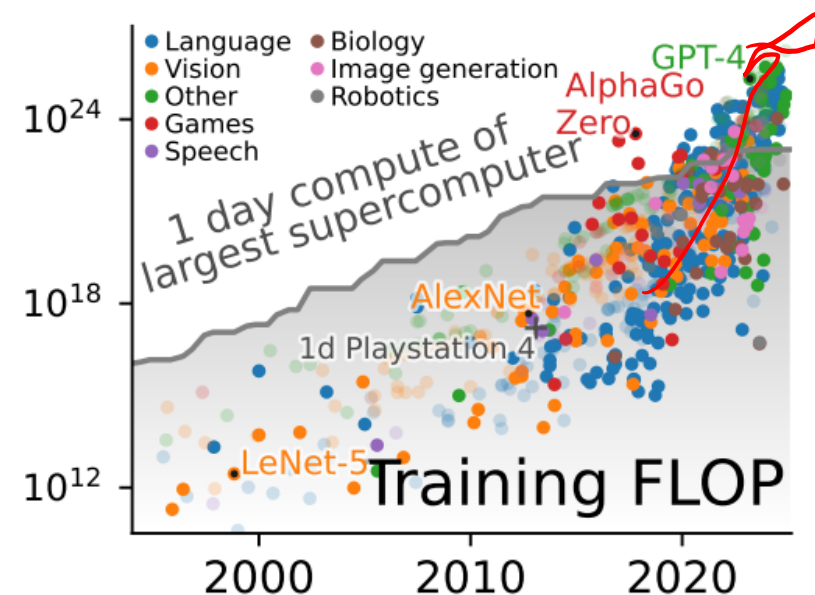
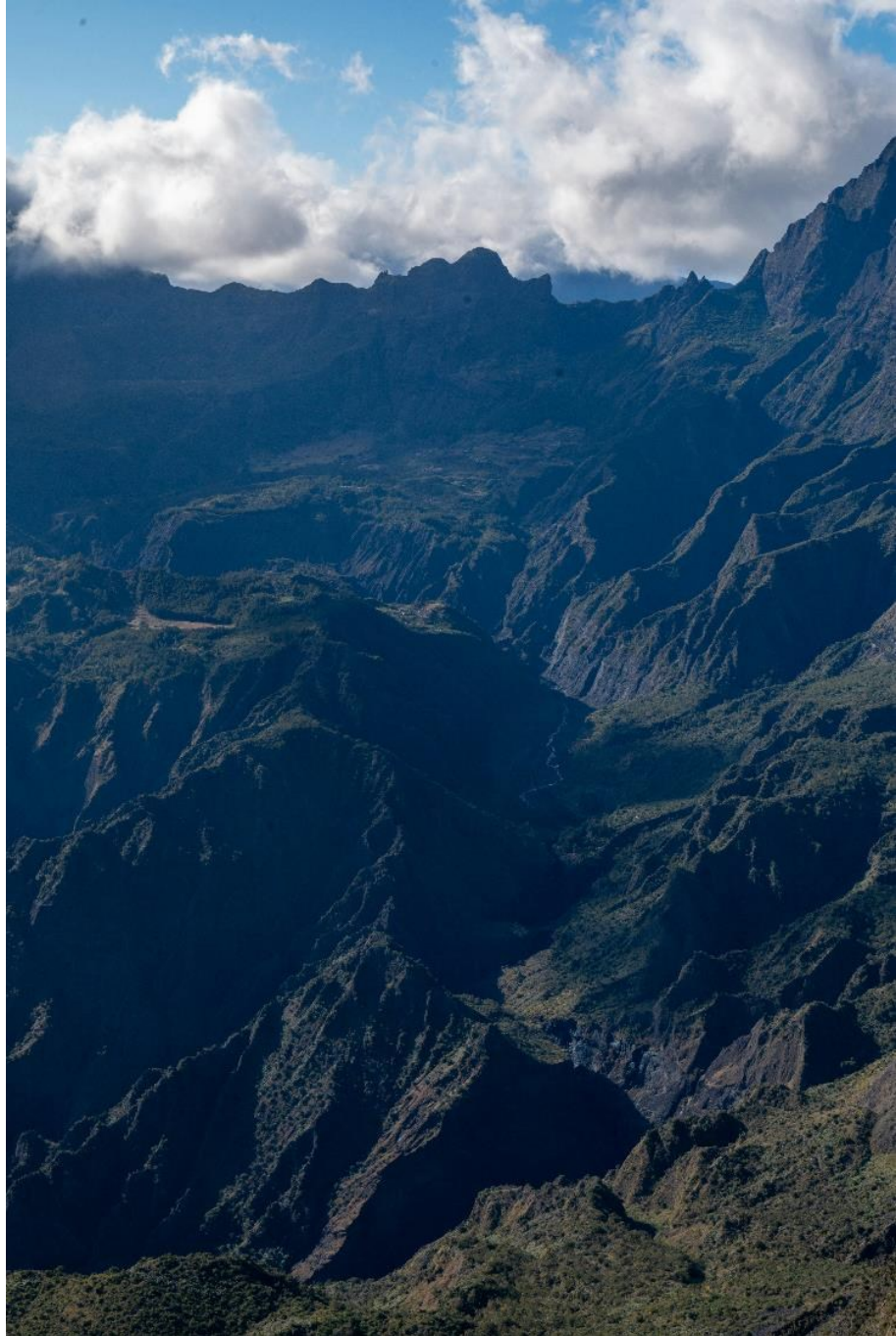


Figure 3. The cost of a single inference is growing faster than compute is improving



Stratégie numérique

- Tenir compte des limites de notre environnement
- Agir en expérimentant
 - Apprendre
 - Coconstruire
 - Définir un(des) chemin(s)



Souveraineté numérique : un choix possible

Conséquences d'une stratégie de « Laisser faire »

- Dépendance forte à un monopole délétère
- Erosion de la capacité à construire un numérique éthique et sobre
- Flux financier de 4 à 7 milliards d'€/an vers les Etats-Unis
(avec une hypothèse de généralisation de ChatGPT aux secteurs Education et ESR sur les tarifs éducation & pro allant de de 10 à 200€/mois/utilisateur)

Comment faire ?

- IAg ouvertes : alternative crédible aux IAg « grand public »
- Robustesse : utiliser plusieurs LLM sobres et spécialisés
- Utiliser en local sur son poste, ou déployer de façon mutualisée pour un ou plusieurs établissements
- Equilibrer
 - ❖ En interne : données à sensibilité élevée (stratégiques, personnelles, confidentielles, confiées par un tiers, propriété intellectuelle...)
 - ❖ En externe : données non sensibles, pour un ratio favorable capacité / coût (environnemental & budgétaire)

IA : tendance de fond

IA générative utile, utilisable, utilisée... Sobriété numérique

Stratégie « juste nécessaire » à établir

- Collaborer : communautés internes, autres acteurs publics
- Mutualisation infrastructures ESR : fédération inférence projet ILaaS
- Déterminer les cas d'usages pertinents : bilan multicritères
- Cadrage : [charte & guide de recommandations](#) (travaux ESR) en lien avec les recommandations de l'atelier E IAgén & pédagogie : Discernement, sobriété, vérification, équité inclusion, transparence, agentivité

Agir & construire ensemble

- Réfléchir en approche systémique
 - ❖ Expérimenter solutions et pratiques
 - ❖ Partage retours, cas d'usages, réflexions
- Travailler l'adéquation entre usages et choix de la technologie, jusqu'à l'absence volontaire de technologie
- Face aux incertitudes, viser une stratégie de robustesse, non de performance à tout prix



<https://www.demos-pro3.fr/proposition-charte-ia-esr/>