

QUESTIONS-RÉPONSES

Volume 3 — Février 2026

AI FOR AMERICANS FIRST

Comprendre le ratio CACI 7-12:1

*Pourquoi l'Europe ne peut pas « simplement payer »
et pourquoi utiliser le cloud US n'est pas la solution*

Fabrice Pizzi

Université Paris Sorbonne — Intelligence Économique & Cybersécurité

À propos de ce document

Ce troisième volume répond à **la question la plus fréquente** posée sur l'étude : « Pourquoi l'Europe a-t-elle 7 à 12 fois moins de compute effectif que les US alors qu'elle pourrait simplement investir davantage — ou utiliser le cloud américain ? » Ce document décompose la réponse en **10 questions progressives** allant du plus intuitif au plus stratégique. Il est conçu pour être lu indépendamment des Volumes 1 et 2.

Repository : <https://mo0ogly.github.io/America-First-IA/>

PARTIE A — Comprendre le ratio : pourquoi 7-12:1 et pas 1,5:1 ?

Q1. Le ratio CACI US/EU est de 7-12:1. Mais les US sont plus riches que l'Europe. N'est-ce pas normal ? « Tu payes, tu as » ? [Grand public / Décideur]

C'est l'objection la plus naturelle — et la plus répandue. Mais elle repose sur une confusion entre le compute brut et le CACI.

Le CACI est normalisé par le PIB. C'est toute la subtilité de l'indice. La formule est : $\text{CACI}(r) = [\text{F}(r) \times \text{E}(r)^{-1}] / [\text{PIB}(r) \times \text{L}(r)]$. Le ratio 7-12:1 ne dit pas « les US ont 7-12× plus de compute en absolu ». Ça, c'est le ratio brut, qui est plutôt ~15:1 (Epoch AI). Le CACI dit quelque chose de beaucoup plus grave :

« À taille économique comparable et à capital humain comparable, les acteurs américains disposent de 7 à 12 fois plus de compute effectif que les européens. »

Faisons le calcul simple. Le PIB US est d'environ 28 000 Md\$, le PIB EU d'environ 18 000 Md\$. Le PIB américain est donc ~1,5× plus grand que l'europeen. Si c'était juste une question d'investissement proportionnel au PIB (« tu payes, tu as »), le ratio de compute devrait être de 1,5:1.

Or il est de 7-12:1. La question est : d'où viennent les 5 à 10× restants ? C'est la contribution centrale de l'étude : ces 5-10× viennent de facteurs structurels que l'investissement seul ne peut pas compenser.

Q2. D'accord. Alors d'où viennent ces 5-10× supplémentaires ? [Grand public / Décideur]

Trois facteurs structurels expliquent l'écart entre le ratio « attendu » (1,5:1) et le ratio « observé » (7-12:1). Aucun des trois ne se résout en « dépensant plus ».

Facteur 1 : le coût de l'énergie (×2-3 en défaveur EU)

Un FLOP de calcul IA en Europe coûte 2 à 3 fois plus cher qu'aux États-Unis. L'électricité industrielle coûte \$110-145/MWh en Europe contre \$50-65/MWh aux US. Concrètement : même si l'Europe investissait la même somme, elle obtiendrait 2 à 3 fois moins de compute utile. Un data center de 500 MW en Europe coûte ~60 % de plus en électricité qu'un data center identique au Texas ou en Virginie.

Ce n'est pas un choix d'investissement — c'est une contrainte structurelle liée au mix énergétique, à la fiscalité carbone, et à la régulation des marchés de l'électricité en Europe. C'est exactement pour ça que la recommandation n°1 est l'énergie dérogée dans les Special Compute Zones (\$50-60/MWh via PPA nucléaires).

Facteur 2 : les export controls + Section 232 (priorité de livraison)

Même avec le budget, vous ne pouvez pas acheter ce qui n'est pas livré. Nvidia produit un nombre fini de GPU H100/H200/B200 par trimestre. Les exemptions domestiques US signifient que les hyperscalers américains (Microsoft, Google, Amazon, Meta, xAI) sont livrés en priorité. En période de pénurie structurelle — ce qui est le cas depuis 2023 — les commandes européennes passent après, avec des délais de 3 à 6 mois.

De plus, la Section 232 (janvier 2026) impose 25 % de tarifs sur les semi-conducteurs IA avancés importés. **Même si les entreprises européennes achetant pour usage domestique ne paient pas directement ce tarif, l'effet d'annonce crée une incertitude réglementaire** : la Proclamation prévoit explicitement un élargissement possible d'ici juillet 2026. Quel investisseur va engager 2 Md€ sur un data center européen si les GPU risquent de coûter 25 % de plus dans 6 mois ?

Facteur 3 : l'effet de gravité capitalistique (fuite des investissements)

C'est le facteur le plus pernicieux. Les investisseurs rationnels — y compris européens et alliés — investissent aux US plutôt qu'en EU, parce que le retour sur investissement y est meilleur : énergie

moins chère, compute plus dense, marché plus grand, écosystème de talent plus concentré.

Le Japon a investi \$550 milliards sur le sol américain — c'est de l'argent qui aurait pu construire du compute japonais (ou européen via des partenariats). Les Émirats convergent vers les US. Même les entreprises européennes (**72-80 % des workloads IA EU sur infra US**) votent avec leurs pieds.

C'est le paradoxe du prisonnier : individuellement, chaque acteur optimise en investissant aux US. Collectivement, le résultat est une perte de souveraineté technologique pour tous les alliés.

Q3. En résumé, pourquoi « simplement investir plus » ne suffit pas ? [Grand public]

Parce que le problème n'est pas un problème de volume d'investissement. C'est un problème de structure. Voici la décomposition :

Composante	Ratio	Cause	Investissement seul suffit ?
PIB relatif	1,5:1	US plus gros que EU	Oui, proportionnel
Énergie	×2-3	Mix énerg., taxe carbone, marché élec.	Non — structurel
Priorité livraison	×1,5-2	Export controls, exemptions US	Non — géopolitique
Gravité capitaliste	×1,5-2	ROI supérieur aux US	Non — systémique
Total cumulé	7-12:1 Multiplicatif		NON

Tableau. Décomposition du ratio CACI US/EU. Source : calibration auteur (2024).

Les facteurs sont multiplicatifs, pas additifs. $1,5 \times 2-3 \times 1,5-2 \times 1,5-2 = 7 \text{ à } 12$. Même en résolvant un seul facteur, les autres maintiennent un écart considérable.

PARTIE B — « Mais on peut utiliser le cloud américain, non ? »

Q4. L'Europe utilise déjà massivement le cloud US. Le ratio est donc un faux problème ? [Décideur / Industriel]

C'est l'objection de 90 % des décideurs européens. Et à court terme, ils ont raison. Pourquoi investir dans du compute souverain si on peut louer celui des Américains ? AWS, Azure et GCP sont disponibles, performants, et immédiatement accessibles. 72 % des workloads IA européens tournent déjà dessus.

Mais c'est exactement la définition de la dépendance stratégique. Vous louez l'infrastructure critique de quelqu'un

d'autre, qui peut : (1) lire vos données, (2) vous couper l'accès, (3) vous faire payer plus cher, et (4) vous livrer après ses propres clients. Les quatre risques sont documentés et concrets.

Q5. Risque 1 : le CLOUD Act. Nos données sont-elles vraiment accessibles aux Américains ? [Juriste / DPO]

Oui. Le Clarifying Lawful Overseas Use of Data Act (CLOUD Act, 2018) autorise les autorités américaines à accéder aux données stockées sur les serveurs des entreprises américaines, **même si le serveur est physiquement hors du sol US.**

Concrètement : une banque française qui entraîne son modèle de scoring crédit sur Azure (serveur en Irlande) — le DoJ américain peut juridiquement exiger l'accès. Un modèle de défense entraîné sur AWS — la NSA peut y accéder. Un algorithme de trading propriétaire sur GCP — la SEC peut le demander.

Ce n'est pas une théorie conspirationniste : c'est le droit américain en vigueur. Le CLOUD Act ne fait pas de distinction entre les données américaines et européennes — il s'applique à l'entreprise (AWS, Microsoft, Google), pas au territoire.

Implication : tout modèle IA entraîné par une entreprise européenne sur une infrastructure cloud américaine est potentiellement accessible aux autorités américaines. C'est un risque de souveraineté, pas un risque théorique.

Q6. Risque 2 : l'accès cloud peut-il vraiment être coupé ou restreint ? [Géopoliticien / Stratégiste]

C'est déjà arrivé — et le mécanisme est en place pour l'élargir. C'est le « chokepoint effect » de Farrell & Newman (2019) appliqué au cloud. Les US ont déjà :

- **Coupé l'accès cloud aux entités chinoises** (Entity List, BIS). Les entreprises chinoises listées ne peuvent plus accéder à AWS, Azure ou GCP. Du jour au lendemain.
- **Restreint l'accès cloud par pays** (AI Diffusion Rule, quotas Tier 2). Les pays classés Tier 2 ont des plafonds sur le volume de compute cloud qu'ils peuvent consommer.
- **Envisagé l'élargissement des restrictions** (Section 232, Proclamation prévoit un élargissement possible juillet 2026).

Aujourd'hui, l'Europe est Tier 1 — accès libre. Mais le mécanisme juridique est en place. Le jour où un conflit commercial éclate (représailles AI Act, taxe GAFA 2.0, désaccord OTAN), il suffit d'un executive order pour restreindre l'accès cloud européen. Le Scénario B de l'étude (« vassalisation numérique ») modélise exactement ce cas.

Rappel historique : l'Europe achetait le gaz russe parce que c'était moins cher et plus simple que de construire des alternatives. Jusqu'en février 2022. Le compute US, c'est le gaz russe de l'IA — sauf que la dépendance est encore plus profonde, parce qu'on ne stocke pas du compute comme on stocke du gaz.

Q7. Risque 3 : même sans coupure, le cloud US ne coûte-t-il pas plus cher pour les Européens ? [Industriel / CFO]

Oui. Le FLOP coûte \$0,5/TFlop aux US et \$1,2-1,8/TFlop en EU (Tableau 9, Chapitre IV). Ce différentiel de 2,4-3,6× vient de trois facteurs :

La latence réseau : un modèle entraîné en Virginie depuis Paris subit des délais de transfert de données qui réduisent l'efficacité du training distribué. Pour les applications temps réel (manufacturing autonome, véhicules connectés, trading), la latence est rédhibitoire.

Les coûts de transfert (egress fees) : les hyperscalers facturent le transfert de données hors de leurs régions. Rapatrier un modèle entraîné aux US vers l'EU coûte. Stocker et transférer des datasets européens vers les US coûte. Ces frais s'empilent.

Les marges régionales : AWS, Azure et GCP pratiquent des tarifs 15-25 % plus élevés en région EU qu'en région US pour le même GPU-heure. C'est documenté par Bruegel.

Résultat : les startups françaises paient 2 à 3× plus cher le même GPU-heure que leurs concurrentes de la Silicon Valley. Même en utilisant le cloud US, le terrain de jeu n'est pas égal.

Q8. Risque 4 : quel impact concret sur le time-to-market des entreprises européennes ? [Industriel / Startup]

McKinsey documente un allongement de 25-40 % du time-to-market pour les entreprises européennes par rapport à leurs concurrentes américaines. Ce retard vient de trois facteurs cumulés :

1) Accès au compute plus lent : files d'attente pour les GPU, délais de livraison, quotas de capacité cloud en région EU.

2) Coût plus élevé : les startups EU doivent lever plus de fonds pour le même volume de compute, ce qui allonge les cycles de levée et retarde les projets.

3) Conformité AI Act : 3-5 % de budget supplémentaire en conformité réglementaire (Accenture), qui s'empile sur le surcoût compute.

En IA, 6 mois de retard = un cycle de modèle de retard = la mort compétitive pour une startup. Quand GPT-5 sort en mars et que votre concurrent américain l'intègre en avril, vous ne pouvez pas attendre septembre. Le marché sera pris.

PARTIE C — Alors on fait quoi ?

Q9. Si « payer plus » ne suffit pas et si le cloud US est un piège, quelle est la solution ? [Décideur / Politique]

L'objectif n'est pas l'autarcie technologique mais la capacité de choix. Pas 0 % de cloud US, mais 30-40 % des workloads sensibles sur du cloud souverain certifié d'ici 2029. Avoir un plan B le jour où le plan A se ferme.

Les cinq leviers sont détaillés au Chapitre VII :

1) Special Compute Zones avec énergie dérogée à \$50-60/MWh via PPA nucléaires. C'est le levier n°1 : sans énergie compétitive, rien d'autre ne fonctionne.

2) Intégration nucléaire-IA : la France a 63 GW nucléaires, EDF peut dédier 2 GW immédiatement. C'est le seul mix énergétique EU qui permet une trajectoire crédible. 250 MW d'ici fin 2026, 6 EPR 2 en construction, SMR à horizon 2030+.

3) Réserves stratégiques de GPU : contrats-cadres EU avec Nvidia, AMD, Intel Foundry sur 18-36 mois. Sécuriser les volumes avant que les restrictions ne s'élargissent. Découpler l'approvisionnement EU des aléas de la politique US.

4) AI Act comme levier offensif : conditionner l'accès au marché EU à la localisation du compute. Reconnaissance mutuelle avec Japon, Corée, Singapour. CLOUD Act Shield européen.

5) Modèle Mistral (analogie Airbus) : pas remplacer OpenAI, mais construire une alternative crédible. ASML a investi 1,3 Md€ (11 % du capital) — le signal industriel est là. La question n'est pas « Mistral peut-il battre OpenAI ? » mais « l'Europe peut-elle se passer de Mistral ? » La réponse est non.

Q10. En une phrase, pourquoi ce ratio de 7-12:1 est-il un problème existentiel et pas juste un retard rattrapable ? [Grand public]

Parce que le compute produit des effets d'agglomération auto-renforçants. Plus vous avez de compute, plus les talents viennent, plus les investisseurs viennent, plus le compute augmente. C'est un cercle vertueux pour les US et un cercle vicieux pour l'EU. Chaque année d'inaction élargit le ratio au lieu de le réduire.

C'est pourquoi la fenêtre 2026-2028 est critique. Après 2028, les positions seront consolidées : les AI Gigafactories US

seront opérationnelles, les écosystèmes cloud figés, les talents installés. Le Scénario D (« Réponse EU massive ») est encore possible aujourd’hui. Il ne le sera plus en 2030.

Le ratio 7-12:1 n'est pas un retard temporaire : c'est une divergence structurelle qui, sans intervention, devient irréversible. C'est la différence entre « on est en retard » et « on a perdu la capacité de rattraper ».

« La question n'est plus de savoir si la recomposition de l'ordre technologique mondial aura lieu — elle est en cours — mais si nous en serons les architectes ou les sujets. »

— Fabrice Pizzi, *AI for Americans First, 2026*

— FIN DU VOLUME 3 —

Ce document peut être lu indépendamment des Volumes 1 (thèse, CACI, méthodologie, scénarios, limites) et 2 (25 questions techniques, géopolitiques, opérationnelles).

Repository : <https://mo0ogly.github.io/America-First-IA/>