

AI FOR AMERICANS FIRST

Protectionnisme IA, Énergie et Semi-conducteurs :
Trajectoires de Divergence US/Europe 2024-2030

Analyse Géostratégique et Économique Intégrée

Chapitre IV

Fabrice Pizzi

Université Sorbonne

Master Intelligence Économique — Intelligence Warfare

75% compute IA mondial = USA **\$675B** capex US
2026 **7-12×** ratio US/EU

Paris — Février 2026

7 chapitres • 4 scénarios prospectifs • 3 zones géographiques

Mots-clés : intelligence artificielle, protectionnisme technologique,
semi-conducteurs, export controls, compute souverain, géopolitique de
l'IA, France, États-Unis, Chine

CHAPITRE IV

Mécanismes de l'avantage concurrentiel américain par l'IA

Le Chapitre III a établi les faits : les États-Unis concentrent 75 % du compute IA mondial, 45 % de la consommation énergétique des data centers, et contrôlent 70 % du marché cloud européen via trois hyperscalers. Ce chapitre analyse les *mécanismes* par lesquels cette asymétrie de compute se transforme en avantage concurrentiel mesurable. Nous identifions quatre canaux de transmission : les coûts d'entraînement comme barrière à l'entrée, la concentration du cloud comme vecteur de dépendance, la productivité différenciée par région, et la captation des rentes d'innovation par les premiers entrants.

4.1 Les coûts d'entraînement comme barrière à l'entrée

4.1.1 L'explosion exponentielle des coûts de training

Le canal le plus documenté est celui des coûts d'entraînement des modèles de fondation. Martens (Bruegel, octobre 2024) établit que le coût d'entraînement d'un modèle frontier est passé de **1 000 dollars en 2017 à près de 200 millions de dollars en 2024**, et pourrait atteindre **plusieurs milliards d'ici 2030**.¹ Cette croissance exponentielle obéit aux *lois d'échelle* (scaling laws) identifiées par Kaplan et al. (2020) : l'amélioration de la performance cognitive des modèles est soumise à des rendements constants dans les inputs complémentaires (paramètres, données d'entraînement, capacité de calcul). Autrement dit, chaque gain marginal de performance nécessite une augmentation proportionnelle du compute, des données et de la taille du modèle.

Cottier et al. (2024), cités par Martens, décomposent les coûts de training en cinq composantes : personnel, puces IA, coûts serveurs, interconnexion réseau et énergie. Ils estiment le coût d'infrastructure (hardware + data center) à **dix fois le coût du training lui-même**, avec un taux d'amortissement de 140 %/an (dépréciation complète en 8,5 mois, au rythme de renouvellement des générations de puces). L'infrastructure de GPT-4 à fin 2023 aurait coûté environ 800 millions de dollars. Par extrapolation, les coûts d'infrastructure pourraient atteindre

500 milliards d'ici 2030, répliqués à travers une demi-douzaine d'hyperscalers.²

4.1.2 L'avantage US : compute abondant et subventionné

Dans ce contexte, l'accès au compute de pointe devient le facteur discriminant. Les Big Tech américaines (Microsoft, Meta, Google, Amazon, xAI) ont investi collectivement **320 milliards de dollars en 2025** dans l'infrastructure IA (serveurs, data centers, puces), contre 230 milliards en 2024.³ Meta développe un système de calcul IA avec 350 000 processeurs H100, estimé à plus de 10 milliards de dollars. Comme le note Martens, « *cette barrière à l'entrée est totalement hors de portée du financement public ; seules les plus grandes entreprises peuvent y prétendre* ». ⁴

Pour les acteurs européens, cette asymétrie se traduit concrètement. Le coût d'entraînement d'un modèle comparable à GPT-4 est estimé à environ 100 millions de dollars aux États-Unis (avec accès direct aux GPU, énergie bon marché, amortissement sur base installée massive). En Europe, le même entraînement coûterait 2 à 5 fois plus cher, en raison de : (i) l'accès indirect au compute via le cloud US (marges hyperscaler), (ii) des coûts énergétiques 2 à 3 fois supérieurs (Chapitre III), et (iii) l'absence d'économies d'échelle comparables. L'écart de coût FLOP (metric M2 du CACI) est estimé entre 2,4× et 3,6×.⁵

4.1.3 La co-opétition forcée : startups européennes et Big Tech US

Les coûts exponentiels créent un mécanisme de dépendance structurelle. Comme l'analyse Martens (Bruegel), les startups IA — y compris européennes — sont contraintes de coopérer avec les Big Tech américaines pour accéder à l'infrastructure de calcul et aux canaux de distribution. Cette *co-opétition* (coopération forcée avec un concurrent) place les startups dans une position subordonnée : elles dépendent des plateformes US pour entraîner leurs modèles, puis se retrouvent en concurrence directe avec les modèles propriétaires de ces mêmes plateformes. L'autorité britannique de la concurrence (CMA, 2024) et la Commission européenne ont identifié ces accords comme potentiellement anti-concurrentiels.⁶ Azoulay et al. (2024), cités par Martens, arguent que le partage de ressources est la seule voie pour éviter la concentration, sauf si un acteur « renard » utilise l'ouverture comme stratégie concurrentielle — ce que Meta a partiellement fait avec LLaMA.

4.2 La concentration du cloud comme vecteur de dépendance géopolitique

4.2.1 Le marché cloud européen : 70 % sous contrôle US

Le deuxième mécanisme est la concentration du marché cloud européen. Synergy Research Group (juillet 2025) établit que le marché européen des services cloud d'infrastructure a atteint **61 milliards d'euros en 2024** (multiplié par six depuis 2017) et devrait dépasser 75 milliards en 2025 (+24 %). **Amazon (AWS), Microsoft (Azure) et Google (GCP) représentent 70 % de ce marché.** La part des fournisseurs européens est passée de 29 % en 2017 à 15 % en 2022, où elle stagne depuis.⁷

Parmi les fournisseurs européens, SAP et Deutsche Telekom sont les premiers, avec chacun environ 2 % du marché, suivis d'OVHcloud, Telecom Italia et Orange. John Dinsdale, analyste en chef de Synergy, observe que « *les fournisseurs US investissent 10 milliards d'euros par trimestre en capex en Europe* », et que les trois hyperscalers opèrent désormais plus de 140 data centers hyperscale en Europe.⁸ La croissance la plus rapide (140-160 %) concerne les services spécifiquement liés à l'IA générative (GPU-as-a-Service, GenAI PaaS), un segment où les acteurs européens sont quasiment absents.

Indicateur	2017	2022	2024	Tendance
Marché cloud EU (€ Md)	~10	~35	61	×6 en 7 ans
Part AWS+Azure+GCP	~50 %	~67 %	70 %	Croissante
Part fournisseurs EU	29 %	15 %	15 %	Stable (plancher)
Capex US en Europe	n.d.	n.d.	~40 € Md/an	10 € Md/trimestre
DC hyperscale en EU	~40	~80	>140	+75 % en 2 ans

Tableau 8. *Marché cloud européen : domination des hyperscalers US. Source : Synergy Research Group (juillet 2025).*

4.2.2 Du lock-in commercial au lock-in géopolitique

Cette concentration dépasse le simple enjeu commercial. Le concept de *vendor lock-in géopolitique* (défini au Chapitre I) décrit une situation où la dépendance technique envers un fournisseur étranger se double d'un risque réglementaire lié aux décisions souveraines du pays d'origine de ce fournisseur. Le CLOUD Act (2018) oblige les fournisseurs cloud américains à fournir aux autorités US les données hébergées sur leurs serveurs, y compris ceux situés en Europe. En mai 2025, le directeur juridique de Microsoft France a reconnu sous serment l'incapacité de garantir que les données de citoyens français ne seraient jamais transmises aux autorités américaines.⁹

Dans le contexte de l'IA, le lock-in géopolitique opère à trois niveaux. Au niveau de l'**infrastructure** (IaaS), les workloads IA sont déployés sur des serveurs physiquement contrôlés par des entités soumises à la juridiction américaine. Au niveau des **plateformes** (PaaS), les frameworks d'entraînement et d'inférence sont intégrés aux écosystèmes propriétaires (Azure ML, SageMaker, Vertex AI). Au niveau des **modèles** (MaaS — Model-as-a-Service), l'accès aux modèles de pointe (GPT-4, Claude, Gemini) passe par des API contrôlées par des entreprises US, avec des conditions d'utilisation qui peuvent être modifiées unilatéralement. Le protectionnisme Section 232, en différenciant le coût d'accès au hardware, renforce ces trois niveaux simultanément.

4.2.3 Le rapport Draghi et le diagnostic européen

Le rapport Draghi (septembre 2024), commandité par la Commission européenne, constitue la reconnaissance institutionnelle la plus complète de cette dépendance. Draghi observe que les restrictions européennes sur le stockage et le traitement des données « *créent des coûts de conformité élevés et entravent la création de jeux de données intégrés de grande taille pour l'entraînement des modèles IA* ». Il recommande la création d'une infrastructure cloud hyperscale européenne propre, la consolidation des fournisseurs cloud européens, et un investissement massif dans l'IA, le calcul quantique et les supercalculateurs EuroHPC.¹⁰ Martens (Bruegel, 2025) modère cependant cet optimisme, observant que les supercalculateurs EuroHPC ne sont pas conçus pour l'IA générative et que la mise à niveau aux standards compétitifs « *dépasse la capacité financière des budgets européens* ».¹¹

4.3 Productivité différenciée : le dividende du compute abondant

4.3.1 Le potentiel macroéconomique théorique

Le troisième canal de transmission est l'impact de l'IA sur la productivité du travail, et son asymétrie régionale. Le McKinsey Global Institute (mai 2024, décembre 2025) estime que l'IA générative pourrait permettre à l'Europe d'atteindre un taux de croissance annuel de la productivité de **3 % par an d'ici 2030** dans un scénario d'adoption accélérée avec redéploiement proactif de la main-d'œuvre.¹² Ce chiffre est suffisant pour combler l'essentiel de l'écart de productivité avec les États-Unis. Toutefois, en scénario d'adoption lente, la croissance de la productivité ne serait que de **0,3 %** — proche des niveaux actuels et insuffisante pour financer le modèle social européen.

Le FMI (Working Paper, mars 2025) fournit une analyse plus granulaire.¹³ En croisant les indices d'exposition des tâches à l'IA (Felten et al., Eloundou et al.) avec les données sectorielles, le FMI estime des gains de productivité totale des facteurs (TFP) *nettement plus élevés dans les économies avancées que dans les pays à revenu intermédiaire*, en raison de leur plus grande part de valeur ajoutée dans les secteurs à forte exposition IA (services financiers, conseil, technologies) et de leurs niveaux de salaires plus élevés, qui justifient davantage l'automatisation. Les études microéconomiques sous-jacentes (essais contrôlés randomisés) montrent des gains allant de 14 % pour les tâches peu qualifiées à plus de 50 % pour les ingénieurs logiciel.

4.3.2 Le gap conditionné par l'accès au compute

Or, la réalisation de ces gains de productivité est conditionnée par l'accès au compute. C'est ici que l'asymétrie diagnosticée au Chapitre III produit ses effets. La Commission française de l'Intelligence Artificielle (2024) estime un impact potentiel de l'IA sur la croissance française atteignant 1,3 point par an, par analogie avec les effets de l'électricité — mais cette estimation suppose un accès non contraint au compute de pointe. De même, McKinsey (décembre 2025) conditionne le scénario optimiste européen à une « *adoption accélérée* » qui implique des investissements massifs dans l'infrastructure.¹⁴

L'écart d'investissement est considérable. McKinsey (janvier 2026) estime que les grandes entreprises européennes accusent un déficit annuel de **700 milliards de dollars en R&D et dépenses d'investissement** par rapport à leurs homologues américaines. Sur les seules technologies

numériques, les entreprises US ont investi **2 000 milliards d’euros de plus** que les entreprises européennes sur la période 2021–2025.¹⁵ L’écart annualisé entre janvier 2024 et septembre 2025 s’élève à 580 milliards d’euros pour l’investissement corporate et 300 milliards pour les startups et scale-ups.

Ce déficit d’investissement, combiné à l’asymétrie de compute et aux coûts énergétiques, produit un **gap de productivité IA conditionné**. Dans notre modélisation, nous distinguons le potentiel théorique (identique pour US et EU à structure sectorielle comparable) du *potentiel réalisable*, qui dépend de l’accès effectif au compute. Avec un CACI(US)/CACI(EU) de 7–12× (Chapitre III), le potentiel réalisable européen est structurellement inférieur.

Dimension	États-Unis	Europe (UE)	Écart
Productivité IA potentielle (McKinsey, scen. acc.)	+2,5–3,5 %/an	+2,5–3,0 %/an	Comparable
Productivité IA réalisable (sous contraintes compute)	+2,5–3,0 %/an	+0,8–1,5 %/an	-1,5 à -2 pts
Investissement IA annuel (corporate + VC)	~450 Md\$	~50–70 Md€	~7:1
Coût FLOP (\$/TFlop, training)	~0,5	~1,2–1,8	2,4–3,6×
CACI ratio (calibration 2024)	Référence	7–12× inférieur	7-12:1

Tableau 9. Productivité IA et avantage concurrentiel : US vs EU. Sources : McKinsey (2024, 2025, 2026), Bruegel/Martens (2024), calibration CACI (auteur).

4.4 La captation des rentes d’innovation : un avantage auto-renforçant

4.4.1 Le mécanisme de first-mover advantage en IA

Le quatrième canal est le plus structurant à long terme. La théorie des technologies génériques (GPT, Bresnahan & Trajtenberg, 1995) prédit que les premiers adoptants d’une technologie générale captent des *rentes d’innovation disproportionnées*, qui deviennent auto-renforçantes via trois mécanismes :

Premièrement, les **effets d'échelle dans les données**. Les entreprises qui déploient l'IA les premières accumulent des données d'usage (feedback loops, fine-tuning) qui améliorent la performance de leurs modèles. Ces données sont non-rivales mais exclusives : une fois capturées par un acteur, elles ne bénéficient pas aux concurrents. McKinsey estime que 75 % du potentiel économique de l'IA générative se concentre sur quatre fonctions clés (service client, développement logiciel, marketing, R&D), où les données d'usage sont particulièrement déterminantes.¹⁶

Deuxièmement, les **effets de réseau dans les plateformes**. Les hyperscalers US bénéficient d'un cercle vertueux : plus de clients → plus de revenus → plus d'investissement en infrastructure → meilleure offre → plus de clients. Ce mécanisme explique pourquoi, malgré les discussions sur la souveraineté numérique, la part des fournisseurs européens stagne à 15 % : l'écart de qualité de service (latence, éventail de services, support technique, écosystème d'intégration) est trop important pour être comblé par les seuls arguments de souveraineté.

Troisièmement, la **captation du talent**. Les entreprises US offrent des rémunérations incluant salaires et equity qui attirent les meilleurs chercheurs et ingénieurs IA mondiaux, y compris européens. Martens note que le personnel qualifié en IA est rare et que les entreprises se livrent une concurrence féroce pour les recruter, avec des packages de rémunération « *difficiles à égaler par les institutions européennes* ». Ce *brain drain* IA affaiblit le facteur L(r) du CACI européen et renforce l'avantage américain dans le capital humain.¹⁷

4.4.2 Le risque d'irréversibilité

Ces trois mécanismes produisent un effet cumulatif potentiellement irréversible. Comme l'analyse l'OCDE (Martens, 2021) par analogie avec les vagues technologiques antérieures (internet haut débit, cloud), les retardataires dans l'adoption d'une technologie générale deviennent structurellement dépendants des plateformes étrangères, subissent des marges comprimées, et perdent progressivement leur autonomie stratégique. Le cas de la pénurie de semi-conducteurs de 2022, qui a coûté 100 milliards d'euros au seul secteur automobile européen, illustre les conséquences concrètes d'une dépendance technologique non anticipée.¹⁸

La trajectoire est d'autant plus préoccupante que l'UE est absente de **quatre des huit segments de la chaîne de valeur de l'IA générative** tels que cartographiés par McKinsey (2024) : semi-conducteurs de conception IA (dominés par Nvidia), plateformes cloud IA (AWS, Azure,

GCP), modèles de fondation (OpenAI, Google, Anthropic, Meta), et outils de développement IA. Elle n'est compétitive que sur les segments d'application aval : applications sectorielles (SAP, Siemens), intégration industrielle, et semi-conducteurs spécialisés (power semiconductors : Infineon, STMicroelectronics, NXP, avec ~15 % de part mondiale).¹⁹

4.5 Synthèse : la mécanique de l'avantage concurrentiel US

Les quatre canaux identifiés forment un système renforçant. L'asymétrie de compute (Chapitre III) alimente des coûts de training différenciés (4.1), qui renforcent la dépendance au cloud US (4.2), qui contraint la productivité réalisable (4.3), ce qui freine les investissements européens et renforce la captation des rentes par les acteurs américains (4.4), qui à leur tour élargissent l'écart de compute. Ce cercle constitue ce que Farrell et Newman (2019) qualifieraient d'*interdependence weaponization* structurelle : la dépendance européenne aux infrastructures américaines, initialement fondée sur l'efficacité économique, devient un levier de pouvoir géopolitique lorsqu'elle est instrumentalisée par des mesures comme la Section 232.

Le protectionnisme Trump (Chapitre III, phase 4) n'a pas *créé* cette asymétrie — elle préexistait massivement. Il l'a **institutionnalisée** et **légalisée** via un mécanisme tarifaire qui différencie le coût d'accès au compute selon la nationalité. Ce faisant, il transforme un avantage de fait en un avantage de droit, dont le démantèlement est politiquement et juridiquement bien plus difficile. Le Chapitre V examine comment ces mécanismes pourraient évoluer dans quatre scénarios prospectifs.

Notes

¹ Martens, B. (2024), « The Tension Between Exploding AI Investment Costs and Slow Productivity Growth », Working Paper 18/2024, Bruegel, p. 5-8. La série temporelle est calibrée sur les données Epoch AI de coûts de training des modèles frontier.

² Cottier, B. et al. (2024), cités dans Martens (2024), op. cit. Le facteur 10× entre coût d'infrastructure et coût de training est une estimation centrale, sujette à variation selon l'utilisation post-training de l'infrastructure.

³ IEA (2025), Energy and AI, Paris, p. 42. L'IEA compile les annonces d'investissement de Meta, Amazon, Alphabet, Microsoft et xAI pour 2025.

⁴ Martens, B. (2024), « Why Artificial Intelligence Is Creating Fundamental Challenges for Competition Policy », Policy Brief 16/2024, Bruegel. Meta prévoit un système de 350 000 H100 (à ~30 000 \$/unité), soit >10 Md\$ de hardware seul.

⁵ Estimation auteur à partir des données Bruegel (coûts training), Eurostat (tarifs électriques industriels), EIA (prix US), et Epoch AI (performance/prix des GPU). La fourchette 2,4-3,6× dépend du mode d'accès (cloud vs on-premise) et des négociations PPA.

⁶ Competition & Markets Authority (2024), « AI Foundation Models — Update Report », Londres. La Commission européenne a lancé en 2024 un appel à contributions sur la concurrence dans l'IA générative.

⁷ Synergy Research Group (juillet 2025), « European Cloud Providers' Local Market Share Now Holds Steady at 15% ». Le marché est défini comme IaaS + PaaS + hosted private cloud.

⁸ Dinsdale, J. (2025), cité dans Synergy Research Group, op. cit. : « US cloud providers continue to invest €10 billion every quarter in European capex, most of which comes from the big three. »

⁹ Data Center Dynamics (juillet 2025), « European Cloud Providers Hold 15% of Local Market Share ». L'article rapporte le témoignage d'Anton Carniax, directeur juridique de Microsoft France, devant une commission parlementaire.

¹⁰ Draghi, M. (septembre 2024), The Future of European Competitiveness, rapport commandité par la Commission européenne. Cité dans Martens, B. (2025), « Catch-Up with the US or Prosper Below the Tech Frontier? An EU Artificial Intelligence Strategy », Bruegel Policy Brief.

¹¹ Martens, B. (2025), op. cit. L'auteur note que l'approche EuroHPC, centrée sur le hardware, caractérise les politiques numériques européennes depuis des décennies et est « particulièrement mal orientée » en matière d'IA.

¹² McKinsey Global Institute (mai 2024), « A New Future of Work: The Race to Deploy AI and Raise Skills in Europe and Beyond ». Le chiffre de 3 % suppose un scénario d'adoption accélérée avec redéploiement complet de la main-d'œuvre.

¹³ FMI (mars 2025), « Artificial Intelligence and Productivity in Europe », Working Paper WP/25/067. L'analyse utilise les indices d'exposition Felten et al. (2021) et Eloundou et al. (2024).

¹⁴ McKinsey (décembre 2025), « Accelerating Europe's AI Adoption: The Role of Sovereign AI Capabilities ». Le rapport estime la croissance de la productivité du travail européenne à 0,2 % en moyenne sur 2020-2025, contre un potentiel de 3,1 % avec adoption complète de l'IA.

¹⁵ McKinsey (janvier 2026), « Transforming Europe: Bold Moves to Lift a Continent ». L'écart de 700 Md\$/an est calculé sur les 150 plus grandes entreprises technologiques (US vs EU) en capex + R&D (données S&P Capital IQ).

¹⁶ McKinsey Global Institute (2024), « The Economic Potential of Generative AI: The Next Productivity Frontier ». Les quatre fonctions (service client, développement logiciel, marketing/ventes, R&D) représentent ~75 % du potentiel économique identifié sur 63 cas d'usage.

¹⁷ Martens, B. (2024), Working Paper 18/2024, op. cit. Le coût du personnel (salaires + equity) est souvent la composante la plus importante du coût de développement d'un modèle IA.

¹⁸ McKinsey (octobre 2024), « Time to Place Our Bets: Europe's AI Opportunity ». Le chiffre de 100 Md€ de perte PIB pour l'industrie automobile européenne provient d'Allianz (septembre 2022) et de l'OICA.

¹⁹ McKinsey (octobre 2024), op. cit. Les semi-conducteurs européens (Infineon, STMicro, NXP) représentent ~15 % du marché mondial de la conception-fabrication intégrée (données Omdia/Informa, 2024), mais principalement sur les segments power et automotive, pas sur les GPU/ASIC IA.

Licence et Avertissement Ce travail, "**America-First-IA**", est mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International (CC BY-NC-SA 4.0).

Vous êtes libre de partager et d'adapter le matériel à des fins non commerciales, à condition de créditer de manière appropriée Fabrice Pizzi (Université Paris Sorbonne) et de diffuser vos contributions avec la même licence. Ce document est fourni à des fins éducatives et de recherche uniquement.