



L'IA ALTERNATIVE : UNE OPPORTUNITÉ POUR L'EUROPE

Septembre
2026

Pourquoi l'Europe devrait moins miser sur la course à la taille et à la puissance brute des modèles de pointe, et davantage sur des solutions d'IA déployables, des données souveraines et une IA digne de confiance pour les institutions

Engin IYIDOĞAN

SKEMA PUBLIKA

SKEMA Publika est le think tank international de SKEMA Business School, il analyse les grandes mutations sociales, économiques, technologiques et géopolitiques pour éclairer la décision publique et privée. Sur la base des travaux de l'école et de contributions extérieures validés scientifiquement, le think tank alimente le débat public et émet des recommandations pour les décideurs nationaux et internationaux.

SKEMA Publika mobilise une approche interdisciplinaire et internationale, nourrie par le réseau mondial de campus de SKEMA et une communauté d'experts issus du monde académique et professionnel. Cette dimension internationale ne constitue pas seulement un réseau, mais un mode de pensée. Ainsi, Publika articule les dynamiques locales avec les transformations globales pour proposer un regard décentré et multipolaire sur les grands enjeux contemporains.

Les opinions exprimées dans ce texte n'engagent que leurs auteurs.

© Tous droits réservés, SKEMA Business School, 2026

Couverture : Drapeau bleu et jaune François Genon. 2019. Unsplash.com

SKEMA Publika

SKEMA Business School, Campus Grand Paris
5 Quai Marcel Dassault – CS 90067
92156 Suresnes Cedex, France

Tél. : +33.1.71.13.39.32
E-mail : publika@skema.edu
Site Internet : www.publika.skema.edu

COLLECTION

« Explorer les zones grises de la géopolitique, du numérique et des risques globaux ».

Dans la « collection incertitude » les travaux analysent les tensions, risques émergents et les dynamiques d'instabilité à l'échelle mondiale.

COMITE DE LECTURE

Frédérique Vidal est Directrice du Développement de SKEMA Publika et Directrice de la Stratégie et de l'Impact Scientifique de SKEMA Business School. Professeur des universités en biologie, elle a été Présidente de l'université Nice-Sophia-Antipolis entre 2012 et 2017, puis ministre de l'Enseignement Supérieur, de la Recherche et de l'Innovation, dans les Gouvernements Philippe et Castex de 2017 à 2022. Elle a été Conseillère spéciale auprès du Président de l'EFMD et est aussi actuellement Représentante permanente de la Principauté de Monaco auprès du Programme des Nations unies pour l'environnement et de la Commission baleinière internationale.

Sean Scull est Chargé de projets think tank, est Doctorant en science de l'information et de la communication à l'université Paul Valéry-Montpellier III. Il est diplômé en sciences politiques avec une spécialisation en relations internationales de l'université de Göteborg et d'un Master en politiques internationales avec une spécialisation en politique anglophone de l'université de Toulon. Sean a vécu et travaillé en Suède et aux États-Unis d'Amérique.

AUTEURS

Engin Iyidogan est professeur de finance à SKEMA Business School. Ses travaux portent notamment sur les applications de l'IA à la finance, la cryptoéconomie et l'avenir de la monnaie. Il est titulaire d'un doctorat en finance de l'Imperial College London.

RESUME

Pour faire émerger ses propres champions de l'intelligence artificielle (IA), l'Europe ne devrait pas évaluer ses perspectives dans ce domaine à la seule aune de sa capacité à rivaliser avec les plus grands laboratoires américains ou chinois dans la course aux modèles de pointe. La question économique la plus pertinente est plutôt de savoir dans quels domaines elle pourrait se forger un avantage dans le déploiement de l'IA. L'agenda 2025-2026 de la Commission européenne esquisse une réponse : usines d'IA et futurs *gigafactories* d'IA, plan d'action pour le continent de l'IA, stratégie pour l'union des données, programmes sectoriels de déploiement et cadre de gouvernance du règlement sur l'IA. Ces initiatives pourraient jeter les bases d'une stratégie européenne centrée sur l'adoption et le déploiement, alors même que le débat public reste largement focalisé sur les modèles de pointe. L'IA alternative, telle qu'elle est entendue ici, ne consiste ni à tourner le dos à la recherche de pointe ni à prôner un isolement réglementaire. Elle désigne une combinaison pragmatique de systèmes multimodaux, de petits et moyens modèles de langage, de modèles à poids ouverts, de génération augmentée par récupération, de déploiements en périphérie ou à proximité de la périphérie, de portefeuilles de modèles et d'espaces de données partagés. Ces approches sont particulièrement adaptées à nombre de secteurs industriels et réglementés en Europe. Elles traduisent également l'importance croissante accordée par la Commission à l'accès aux capacités de calcul, aux données sectorielles et à une adoption à grande échelle, plutôt qu'aux seules performances mesurées par les benchmarks.

Le paysage technologique de 2026 renforce la pertinence de cette approche. Les principaux fournisseurs proposent désormais des familles de modèles répondant à des exigences différentes en matière de coût, de latence et de capacités. La multimodalité, les variantes allégées, les modèles à poids ouverts et le déploiement directement sur les appareils constituent désormais des catégories de produits bien établies. L'adoption de l'IA par les entreprises progresse rapidement, mais les gains effectivement constatés restent limités et inégaux. Dans plusieurs enquêtes menées auprès de dirigeants aux États-Unis, au Royaume-Uni, en Allemagne et en Australie, la majorité des répondants ne font encore état d'aucun effet sur l'emploi ou la productivité au sein de leur entreprise, même s'ils anticipent des effets plus importants dans les trois prochaines années. Une étude danoise ne relève elle non plus aucun effet moyen détectable sur le nombre d'heures

travaillées ou les revenus après deux ans, tout en mettant en évidence des évolutions dans la nature des tâches et les transitions professionnelles. Ces résultats ne signifient pas que l'IA est dépourvue de valeur : ils montrent que cette valeur dépend de la refonte de l'organisation, de l'adéquation de l'outil aux processus de travail et de la gouvernance.

L'écart d'adoption constitue lui aussi un enjeu majeur. Aux États-Unis, les salariés utilisent l'IA générative (ou GenAI) dans leur travail plus fréquemment et de manière plus intensive que dans plusieurs grandes économies européennes. Selon l'enquête de la BCE sur l'accès des entreprises au financement, 38 % des entreprises de la zone euro considèrent leur niveau d'adoption comme avancé, tandis que 33 % n'en font encore qu'un usage occasionnel ou expérimental¹. Le défi européen ne tient donc pas uniquement aux capacités de pointe, mais aussi à une diffusion plus lente de l'IA au sein des entreprises et des administrations publiques.

Les systèmes de petite taille, multimodaux, décentralisés et adossés à des mécanismes de récupération de l'information peuvent réduire la latence et les coûts d'infrastructure, limiter les transferts inutiles vers le cloud et offrir une plus grande maîtrise locale. Ces caractéristiques présentent un intérêt particulier sur des marchés réglementés et fragmentés. La question de la gouvernance ne se limite donc pas aux grands modèles centralisés : elle englobe également les systèmes distribués, les couches d'orchestration et la traçabilité transfrontalière. Les exigences qui leur sont imposées doivent être proportionnées aux capacités et aux risques effectivement démontrés. La voie qui s'ouvre à l'Europe est ainsi plus étroite, mais aussi plus concrète qu'une simple imitation des laboratoires à la pointe de l'IA : elle réside dans des systèmes auditables, réactifs, fondés sur des connaissances sectorielles et acceptables aux yeux des institutions qui les utilisent. L'Europe pourrait y construire un avantage, à condition d'accompagner les investissements dans les infrastructures d'une véritable capacité de déploiement.

Dans ce *policy paper*, nous formulons **quatre priorités d'action**. L'Europe devrait : (1) faire des usines d'IA de véritables institutions au service du

¹ Ferrando, A., Lamboglia, S., Rariga, J., & Schmidt, M. (2026). Adopting and investing in AI: evidence from euro area firms in the SAFE. Banque centrale européenne. https://www.ecb.europa.eu/press/economic-bulletin/focus/2026/html/ecb.ebbox202602_06~c9f485c478.en.html

déploiement, reliées aux données sectorielles et aux capacités d'évaluation ; (2) mobiliser la commande publique afin de créer une demande pour des solutions d'IA auditable, portable et interopérable ; (3) aider les entreprises à passer de l'achat d'un modèle unique à la gestion de portefeuilles de modèles encadrés par une gouvernance adaptée ; et (4) mettre en place des infrastructures de validation mutualisées dans les secteurs à fortes exigences de confiance.

TABLE DES MATIÈRES

COMITE DE LECTURE	3
RESUME	i
Introduction.....	1
I. L'Europe à l'heure de l'IA alternative	5
A. À la frontière technologique : pourquoi les petits modèles comptent 5	
B. Économie de l'IA dans les entreprises et le secteur public.....	6
C. Gouvernance, souveraineté et IA avancée	10
II. Cinq thèses stratégiques	13
A. Une approche par portefeuille peut réduire le coût moyen de l'intelligence	13
B. Dans l'IA d'entreprise, les données propriétaires ne sont pas le nouveau pétrole : elles sont le nouveau fonds de roulement	13
C. La meilleure carte de l'Europe pourrait être le déploiement spécialisé, plutôt que l'imitation des acteurs de pointe	13
D. La gouvernance peut devenir une technologie de production.....	14
E. L'orchestration deviendra un enjeu organisationnel central de l'IA décentralisée	14
III. Recommandations	15
Conclusion	17
Annexe.....	18
A. Le marché de l'IA en 2026.....	18

INTRODUCTION

Selon l'organisation à but non lucratif *Rest of World*, les États-Unis et la Chine dominent le marché de l'IA, en attirant à eux seuls entre 70 et 80 % des investissements mondiaux dans ce domaine². Le Boston Consulting Group dresse le même constat. D'un côté, Washington conserve une longueur d'avance dans les modèles d'IA de pointe, les talents et la mobilisation des capitaux. De l'autre, Pékin se distingue par des modèles optimisés en termes de coûts et par l'adoption rapide de cette nouvelle technologie à l'échelle de la société³. Le grand absent semble être le Vieux Continent : aucune entreprise européenne ne paraît aujourd'hui en mesure de rivaliser sérieusement avec les géants technologiques établis, comme Google ou Microsoft, ni avec les nouveaux acteurs majeurs de l'IA, tels qu'OpenAI ou NVIDIA.

L'Europe souffre en effet de handicaps structurels dans la compétition qui l'oppose aux États-Unis et à la Chine dans le domaine de l'IA, en particulier lorsqu'il s'agit de suivre une stratégie fondée sur la course à la taille et à la puissance brute. Premièrement, le capital-risque européen ne dispose pas d'une capacité d'investissement comparable à celle des fonds américains ou de l'État chinois. Deuxièmement, l'Europe dispose de ressources énergétiques plus contraintes et s'impose des objectifs plus stricts en matière d'émissions, ce qui complique le déploiement de centres de données très énergivores⁴. Enfin, comme nous le soulignons dans le rapport de SKEMA Publika *Intelligence artificielle et souveraineté européenne*, le marché européen reste fragmenté, ce qui complique la mise en œuvre d'une stratégie cohérente et nuit à sa capacité à rivaliser à l'échelle mondiale.

² Chandran, R., & Chandran, R. (2026). The Great AI Divide : Navigating U.S. and Chinese dominance. Rest Of World. <https://restofworld.org/2026/ai-divide-america-china-world/>

³ Lang, N. et al. (2026). The Great Divide: How the US and China Are Splitting the AI World. Boston Consulting Group. <https://www.bcg.com/publications/2026/us-and-china-ai-strategy-causing-global-ai-divide>

⁴ AI Office publishes frontier AI expert findings on EU competitiveness, sovereignty and security. (2026). Shaping Europe's Digital Future. <https://digital-strategy.ec.europa.eu/en/library/ai-office-publishes-frontier-ai-expert-findings-eu-competitiveness-sovereignty-and-security>

Dans ce rapport, le terme « IA alternative » désigne une inflexion à la fois politique et technologique déjà perceptible en Europe. Le plan d'action pour le continent de l'IA articule infrastructures de calcul, accès aux données, adoption sectorielle et compétences ; la stratégie pour l'union des données fait de la disponibilité des données une condition de la compétitivité ; et EuroHPC présente les usines d'IA comme des écosystèmes réunissant puissance de calcul, données, accompagnement et compétences.⁵ Ces initiatives ne constituent pas encore une stratégie industrielle pleinement établie, mais elles en fournissent déjà plusieurs briques essentielles.

Premièrement, l'IA multimodale peut traiter plusieurs types d'entrées, notamment du texte, des images et des contenus audios. Cette capacité est particulièrement pertinente pour les entreprises, dont une grande partie des données se trouve non seulement sous forme de texte, mais aussi dans des documents numérisés, des images de produits, des fichiers audios ou des processus alimentés par des capteurs. Il n'existe pas de seuil unique en nombre de paramètres permettant de définir un petit modèle de langage (SLM) : une étude récente englobe ainsi des modèles comptant d'environ 100 millions à 5 milliards de paramètres.⁶ Dans ce policy paper, nous entendons par SLM un modèle conçu pour nécessiter sensiblement moins de puissance de calcul, offrir une latence plus faible ou requérir une capacité d'hébergement moindre que les systèmes de pointe à très grande échelle, notamment en vue d'un déploiement spécialisé ou local.

Deuxièmement, l'architecture de déploiement compte tout autant que la taille du modèle. L'IA en périphérie (Edge AI) traite les données directement sur des appareils locaux ou sur des nœuds de proximité, sans dépendre d'une connexion permanente à une infrastructure cloud. Les approches décentralisées étendent cette logique aux systèmes d'entraînement ou d'inférence, même si une grande partie de la littérature sur le sujet reste encore exploratoire.⁷ Pour les sites industriels, les hôpitaux et les administrations publiques en Europe, la question concrète est donc de

⁵ European Commission, [Le plan d'action pour le continent de l'IA](#) (2025); European Commission, [Stratégie européenne pour l'union des données](#) (2026); EuroHPC Joint Undertaking, [Usines d'IA](#) (2026).

⁶ Lu et al., [Small Language Models: Survey, Measurements, and Insights](#), arXiv:2409.15790, version 3 (2025).

⁷ Dong et al., [Beyond a Single AI Cluster: A Survey of Decentralized LLM Training](#), arXiv:2503.11023 (2025), prépublication d'une étude.

déterminer quelles données et quelles charges de travail ont réellement besoin de quitter l'environnement local.

Troisièmement, les fournisseurs proposent de plus en plus des familles de modèles adaptées à différents besoins en matière de capacités, de latence et de coûts.⁸ Dans ce rapport, un portefeuille de modèles désigne une architecture organisée de manière à orienter chaque tâche vers le modèle le plus approprié : un modèle plus puissant pour les tâches complexes mais occasionnelles ; un modèle plus rapide pour les tâches courantes ; un modèle multimodal pour les traitements documentaires et visuels ; un modèle local pour les données sensibles ; et des couches de récupération permettant de maintenir à jour les connaissances sectorielles. L'analogie avec la finance est éclairante : il s'agit d'arbitrer entre performance, coût, latence et contraintes de gouvernance, plutôt que d'utiliser le même modèle pour toutes les tâches.

Quatrièmement, la génération augmentée par récupération (RAG) permet de connecter un modèle, au moment de l'inférence, à des sources de connaissances externes, à la documentation interne ou à des bases de données faisant autorité. Cette approche est particulièrement pertinente dans les secteurs où des informations actualisées, propriétaires ou réglementées importent davantage qu'un préentraînement généraliste. L'opportunité pour l'Europe réside donc moins dans la reproduction des plus vastes campagnes d'entraînement que dans sa capacité à exploiter efficacement des connaissances institutionnelles fiables.

Le droit de l'Union européenne encadre les modèles d'IA à usage général ainsi que les obligations liées aux risques systémiques ; il ne définit pas l'intelligence artificielle générale (AGI) comme une catégorie réglementaire.⁹ L'AGI demeure un concept stratégique controversé. Ce rapport la considère comme un enjeu de long terme et se concentre sur les caractéristiques effectivement gouvernables des systèmes déployés, plutôt que sur des seuils ou des échéances hypothétiques. L'objectif n'est pas de créer un laboratoire unique, de dominer un benchmark ou de développer un *chatbot* souverain, mais de bâtir un écosystème interopérable permettant un déploiement fiable de l'IA. Autrement dit, nous estimons que l'Europe ne devrait pas

⁸ Documentation officielle des modèles : OpenAI, [Models](#); Anthropic, [Models overview](#); Google, [Gemini models](#); Mistral AI, [Models overview](#) (consultés le 26 août 2026).

⁹ Commission européenne, [Regulatory framework on AI](#) (mis à jour 2026).

chercher exclusivement à rivaliser dans la course aux modèles de pointe toujours plus massifs, mais privilégier des systèmes d'IA alternative adaptés à ses besoins propres. Les entreprises européennes n'ont pas nécessairement besoin d'une IA reposant avant tout sur la puissance brute et le passage à l'échelle : elles ont surtout besoin de solutions moins coûteuses et mieux adaptées aux réalités de l'écosystème européen – qu'il s'agisse de l'agriculture, des services, de la finance, de l'industrie ou du secteur public.

I. L'EUROPE À L'HEURE DE L'IA ALTERNATIVE

A. À LA FRONTIÈRE TECHNOLOGIQUE : POURQUOI LES PETITS MODÈLES COMPTENT

Le choix d'un modèle repose désormais sur de multiples critères : multimodalité, contrôle du raisonnement, longueur du contexte, utilisation d'outils, souplesse de déploiement, latence et coût. Pour répondre à ces différents besoins, les fournisseurs déclinent leurs offres entre modèles phares et versions plus légères. La taille reste un facteur important, mais l'architecture de déploiement fait désormais pleinement partie de la frontière technologique et commerciale.

Les modèles à poids ouverts (ou open-weight) élargissent encore le champ des possibles. Llama 4 Scout, de Meta, associe une multimodalité native à une architecture de type mélange d'experts ; la famille Gemma de Google se décline en plusieurs tailles et configurations adaptées au déploiement local ; Mistral propose depuis l'Europe des modèles ouverts ; IBM publie ses modèles Granite avec des fonctionnalités de traçabilité et SmolLM3 de Hugging Face est un modèle multilingue de 3 milliards de paramètres.¹⁰

Trois évolutions méritent une attention particulière. La multimodalité tend à devenir la norme, l'efficacité de l'inférence constitue désormais un enjeu stratégique et les capacités des petits modèles continuent de progresser. Le contrôle du raisonnement, les différents niveaux d'utilisation d'outils et les variantes à poids ouverts ont tous une incidence sur le coût et la gouvernance d'un système. Il convient donc de les considérer comme de véritables choix de déploiement, et non comme de simples considérations techniques de second plan.

Les modèles de pointe conservent toute leur importance, notamment dans la recherche scientifique, la programmation et les tâches de raisonnement ouvertes. Mais pour de nombreux autres usages en entreprise et dans le secteur public, la question concrète est de déterminer quelle combinaison de modèles permet d'atteindre le niveau de qualité requis, avec un coût, une

¹⁰ Documentation officielle des modèles : Meta, [Llama 4](#) ; Google, [Gemma model card](#) ; Mistral AI, [Mistral Small 4](#) ; IBM, [Granite 4.1](#) ; Hugging Face, [SmolLM3](#) (consultés le 26 août 2026).

latence et un risque de gouvernance aussi faibles que possible. Une entreprise disposant de solides capacités de routage, d'évaluation et de déploiement local peut ainsi créer davantage de valeur qu'une autre qui achète systématiquement des capacités haut de gamme hébergées à distance pour chacun de ses processus. Les atouts de l'Europe dans l'industrie manufacturière, la santé et les services publics offrent un terrain particulièrement favorable à cette approche par portefeuille.

B. ÉCONOMIE DE L'IA DANS LES ENTREPRISES ET LE SECTEUR PUBLIC

Les données disponibles sur les effets organisationnels de l'IA invitent à davantage de prudence que ne le laissent entendre de nombreux discours. Dans un document de travail du NBER publié en 2026 et fondé sur des enquêtes menées auprès de cadres dirigeants, Yotzov et al. indiquent que 69 % des entreprises aux États-Unis, au Royaume-Uni, en Allemagne et en Australie utilisent une forme ou une autre de technologie d'IA.¹¹ Plus de 90 % des répondants déclarent n'avoir observé jusqu'à présent aucun effet sur l'emploi au sein de leur entreprise, et 89 % aucun effet sur la productivité du travail. Il s'agit ici d'estimations fournies par les dirigeants, et non de mesures directes de la productivité. Au Danemark, Humlum et Vestergaard ne constatent, après deux ans, aucun effet moyen détectable supérieur à 2 % sur le nombre d'heures travaillées ou les revenus, tout en mettant en évidence des transformations dans les tâches et les trajectoires professionnelles.¹² L'adoption est bien réelle, mais sa traduction en valeur économique suppose souvent une évolution de l'organisation.

La difficulté de l'Europe ne tient donc pas seulement à sa capacité d'innovation, mais aussi à la diffusion de l'IA. Début 2026, 43 % des salariés américains déclaraient utiliser l'IA générative dans le cadre de leur travail, contre 26 % en Italie et 36 % au Royaume-Uni. Son utilisation était également moins intensive en Allemagne, en France et en Italie.¹³ Dans la zone euro, 38 % des entreprises estimaient avoir atteint un niveau d'adoption avancé,

¹¹ Yotzov et al., *Firm Data on AI*, NBER Working Paper 34836 (2026).

¹² Humlum and Vestergaard, *Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI*, NBER Working Paper 33777, version révisée en 2026.

¹³ Bick et al., *Differences in AI adoption in Europe and the US* (CEPR, 2026).

tandis que 33 % en restaient à un usage occasionnel ou expérimental.¹⁴ Ces chiffres mettent en évidence une priorité très concrète : accélérer l'intégration de l'IA dans les processus réels, en particulier dans les secteurs où les entreprises européennes disposent déjà d'une forte expertise métier.

Les petits modèles et les modèles multimodaux peuvent présenter un intérêt économique dès lors qu'ils permettent d'atteindre le niveau de qualité requis à moindre coût ou avec une latence plus faible, ou lorsque la maîtrise locale des données est nécessaire. Les systèmes à poids ouverts et les architectures RAG hybrides peuvent eux aussi élargir les possibilités de déploiement. Aucune de ces caractéristiques ne garantit toutefois, à elle seule, la création de valeur : les efforts d'intégration, l'évaluation et la responsabilité restent déterminants.

Les entreprises devraient donc envisager l'adoption de l'IA comme un modèle opérationnel, plutôt que comme l'achat d'un logiciel auprès d'un fournisseur unique. La pile technologique comprend généralement l'accès aux modèles, des mécanismes de récupération appliqués aux connaissances propriétaires, des garde-fous et des dispositifs de gouvernance. C'est dans ce contexte que la politique européenne des données prend tout son sens : le règlement sur la gouvernance des données soutient la mise en place d'espaces européens communs de données dans des secteurs tels que la santé, la finance, l'industrie manufacturière, l'administration publique et les compétences.¹⁵ La stratégie pour l'union des données prévoit de relier ces espaces aux écosystèmes d'IA au moyen de futurs laboratoires de données. Si les systèmes d'IA fondés sur des connaissances sectorielles deviennent une source majeure de valeur, les institutions capables de rendre les données effectivement exploitables compteront tout autant que les capacités liées aux modèles de pointe.

L'industrie manufacturière en fournit une première illustration. Siemens présente son *Industrial Copilot* comme un outil destiné à faciliter la génération de code, le diagnostic et la collaboration entre l'humain et la machine, tandis que Beckhoff commercialise des solutions de vision par

¹⁴ Ferrando et al., [Adopting and investing in AI: evidence from euro area firms in the SAFE](#) (ECB, 2026).

¹⁵ Commission européenne, [Acte sur la gouvernance des données](#) et [Stratégie européenne pour l'union des données](#).

ordinateur pour des applications telles que le contrôle visuel de la qualité.¹⁶ Il s'agit d'exemples fournis par les entreprises elles-mêmes, et non d'évaluations indépendantes, mais ils s'inscrivent dans des domaines où l'Europe dispose d'atouts bien établis : la production à forte valeur ajoutée et les processus complexes mais clairement circonscrits. Dans de nombreuses usines, le besoin porte moins sur le modèle le plus puissant disponible que sur des systèmes fiables, vérifiables, dotés de capacités de vision et capables de fonctionner au plus près des lignes de production.

La finance offre un cas de figure différent. BNP Paribas s'est fixée pour objectif de générer 750 millions d'euros de valeur grâce à l'intégration de l'IA d'ici fin 2026 et a noué un partenariat avec Mistral AI couvrant plusieurs métiers du groupe.¹⁷ Il s'agit là d'un objectif annoncé par l'entreprise, et non d'un résultat effectivement constaté de manière indépendante. Dans son rapport annuel, ING décrit un cadre de gouvernance de l'IA, des équipes dédiées aux risques ainsi qu'une déclaration qualitative d'appétence au risque spécifique à l'IA générative (Generative AI Risk Appetite Statement).¹⁸ La Banque d'Angleterre fait pour sa part état d'un soutien du secteur à une réglementation fondée sur des principes, tout en soulignant les pénuries de compétences et la prudence des fonctions de gestion des risques de deuxième ligne.¹⁹ Dans les services financiers, le déploiement ne dépend donc pas seulement des capacités de l'IA, mais aussi de la possibilité de documenter, superviser et justifier les décisions prises.

La santé constitue un domaine prometteur, mais exigeant, car elle conjugue forte intensité de données, risques cliniques, fragmentation transfrontalière et poids important du secteur public. Le règlement relatif à l'espace européen des données de santé établit un cadre pour l'accès aux données de santé électroniques, leur interopérabilité et leur réutilisation sécurisée.²⁰ Parmi les initiatives de la Commission figurent Cancer Image Europe et les futurs centres de dépistage assistés par l'IA, tandis que MedGemma de Google illustre le développement de modèles multimodaux ouverts

¹⁶ Exemples d'entreprises : Siemens, [Artificial intelligence at Siemens](#); Beckhoff, [TwinCAT 3 Machine Learning](#) (2026).

¹⁷ BNP Paribas, [Rapport intégré 2024](#) (2025), objectif annoncé par l'entreprise.

¹⁸ ING Bank N.V., [Annual Report 2025](#) (2026).

¹⁹ Bank of England, [Summary of AI roundtables](#) (2026).

²⁰ Commission européenne, [Règlement relatif à l'espace européen des données de santé](#) (2025).

appliqués aux textes et aux images médicales.²¹ Ces initiatives se trouvent encore à des stades différents de mise en œuvre et de validation. L'opportunité pour l'Europe réside dans les infrastructures capables d'articuler données, évaluation clinique et réglementation.

Les administrations publiques représentent elles aussi un potentiel important. Une note de l'OCDE souligne que l'IA peut améliorer les services administratifs à condition que les pouvoirs publics développent parallèlement leurs compétences internes, leurs dispositifs de gouvernance et la préparation de leurs agents. Elle indique notamment que Kela, en Finlande, estime à l'équivalent de 38 années de travail à temps plein les économies annuelles permises par la classification et le traitement automatisés de documents, et que la municipalité danoise de Gladsaxe exploite un modèle fondé sur GPT et hébergé sur ses propres infrastructures.²² Là encore, il s'agit d'exemples déclarés et non d'une estimation générale des gains de productivité. La Commission a également proposé de recourir davantage aux normes et aux modèles ouverts dans la commande publique, en faisant de la souveraineté, de l'interopérabilité, de la portabilité et de la durabilité des critères de conception.²³ L'ensemble de ces exemples montre comment la demande publique pourrait soutenir le développement de systèmes auditables et interopérables. Par ailleurs, comme nous le soulignons dans le policy paper de SKEMA Publika *Croissance des start-ups et PME*, faciliter l'accès des entreprises européennes à la commande publique permettrait à l'État et aux collectivités locales de jouer un rôle décisif pour leur ouvrir des débouchés, accélérer leur activité et les aider à changer d'échelle. Utilisée comme un véritable outil stratégique, la commande publique peut ainsi devenir un puissant levier de croissance.

Dans les entreprises comme dans les administrations, l'accès aux modèles est de moins en moins le facteur rare. La difficulté réside davantage dans la capacité à repenser les processus et à répartir clairement les responsabilités. Un modèle peut répondre à une question, mais il ne peut pas, à lui seul, refondre un processus de traitement des dossiers, une fonction achats, un parcours clinique ou une procédure de maintenance. Cela suppose des

²¹ Commission européenne, [L'intelligence artificielle dans le domaine des soins de santé et Centres européens de dépistage avancé alimentés par l'IA](#) ; Google, [MedGemma model card](#) (2026).

²² OCDE, [Building an AI-ready public workforce](#) (2026).

²³ Commission européenne, [Commission boosts open and interoperable digital ecosystems for public administrations](#) (2026).

équipes, des données, des interfaces et la confiance des institutions. C'est autour de ces contraintes que devrait se construire la stratégie européenne de déploiement.

C. GOUVERNANCE, SOUVERAINETÉ ET IA AVANCÉE

L'opposition habituelle entre réglementation et innovation est trop simpliste. Le règlement sur l'IA est entré en vigueur le 1er août 2024 et est devenu, sauf exceptions particulières, généralement applicable le 2 août 2026. Le règlement omnibus sur l'IA, entré en vigueur le 27 juillet 2026, a repoussé au 2 décembre 2027 l'échéance applicable aux systèmes à haut risque visés à l'annexe III, et au 2 août 2028 celle concernant les systèmes d'IA intégrés aux produits réglementés relevant de l'annexe I. La Commission a également publié des lignes directrices et des codes de bonnes pratiques relatifs aux modèles d'IA à usage général. Les obligations de transparence concernant le marquage de certains contenus générés par l'IA s'appliquent depuis août 2026 ; le code de bonnes pratiques correspondant est volontaire et vise à faciliter la mise en conformité.²⁴ L'ensemble de ces mesures dessine un cadre de déploiement de plus en plus concret, même si sa mise en œuvre reste encore inachevée.

Ce cadre gagne également en assise institutionnelle. Le Bureau européen de l'IA accompagne sa mise en œuvre, tandis que le Groupe scientifique sur la législation sur l'IA conseille les autorités sur les incidences et les risques associés aux modèles d'IA à usage général.²⁵ L'Europe pourrait tirer un avantage concret de référentiels documentaires communs, de normes sectorielles, d'une expertise scientifique partagée et d'exigences plus homogènes dans la commande publique. Ces dispositifs peuvent réduire l'incertitude, à condition toutefois de rester proportionnés et applicables dans la pratique.

La commande publique joue ici un rôle particulièrement important. La Public Buyers Community de l'Union européenne met à la disposition des acheteurs

²⁴ Commission européenne, [AI Omnibus enters into force](#) ; [Code de bonnes pratiques de l'IA à usage général](#) ; [Lignes directrices à l'intention des fournisseurs de modèles d'IA à usage général](#) ; [Code de bonnes pratiques sur le marquage et l'étiquetage des contenus générés par l'IA](#) (2025-2026).

²⁵ Commission européenne, [Bureau européen de l'IA](#) et [Groupe scientifique sur la législation sur l'IA](#).

publics des ressources ainsi que des clauses contractuelles types pour l'acquisition de solutions d'IA dignes de confiance.²⁶ La Commission a également plaidé en faveur de normes et de modèles ouverts et fait de la souveraineté l'un des critères de son appel d'offres consacré au cloud souverain. La commande publique peut ainsi orienter la demande vers des systèmes auditable, portables et interopérables : son impact dépendra toutefois de la constance avec laquelle les acheteurs appliqueront ces exigences.

Le plan d'action pour le continent de l'IA prévoit de mobiliser 200 milliards d'euros pour accélérer le développement de l'intelligence artificielle, dont 20 milliards destinés à financer jusqu'à cinq *gigafactories* d'IA, parallèlement aux usines d'IA, au renforcement des capacités des centres de données et à l'union des données.²⁷ EuroHPC présente les usines d'IA comme des dispositifs donnant accès à des capacités de calcul et à un accompagnement adapté aux start-up et aux PME. Ces mesures ne supprimeront pas la dépendance de l'Europe à l'égard des accélérateurs, plateformes cloud ou logiciels non européens, mais elles constituent un programme d'infrastructures d'une ampleur significative. La souveraineté ne doit pas être synonyme d'autarcie. La stratégie pour l'union des données associe une plus grande maîtrise européenne à une ouverture envers des partenaires de confiance, dans des conditions équitables et sûres. Une approche viable consisterait à favoriser des flux de données transfrontaliers fiables, une traçabilité documentée, des composants ouverts réutilisables, un déploiement local lorsque cela est nécessaire et des capacités de calcul suffisantes pour limiter les dépendances critiques.

L'intelligence artificielle générale (AGI) devrait rester un enjeu de long terme, plutôt que devenir le principe structurant de la politique européenne. Un document prospectif publié en juin 2026 envisage plusieurs trajectoires possibles entre l'AGI et des systèmes plus avancés, notamment de vastes collectifs multi-agents. Il s'agit toutefois d'un cadre de réflexion tourné vers l'avenir, et non d'éléments permettant d'affirmer que l'une de ces trajectoires serait imminente.²⁸ La leçon pratique à en tirer est plus circonscrite : dans les systèmes distribués, la répartition des responsabilités

²⁶ Public Buyers Community, [Procurement of AI](#) (2026).

²⁷ Commission européenne, [Plan d'action pour le continent de l'IA](#) (2025-2026).

²⁸ Genewein et al., [From AGI to ASI](#), arXiv:2606.12683 (2026), document prospectif/prépublication.

entre fournisseurs de modèles, composants de récupération et orchestrateurs peut devenir moins lisible. L'Europe devrait donc étendre aux couches d'orchestration les exigences en matière de traçabilité, de journalisation et de signalement des incidents, tout en veillant à proportionner les obligations aux capacités et aux risques effectivement démontrés.

Les outils de gouvernance existants vont déjà dans ce sens. Le Code de bonnes pratiques de l'IA à usage général comprend notamment un formulaire de documentation des modèles. IBM publie, pour Granite, des fiches de modèles ainsi que des informations relatives à la signature cryptographique, illustrant ainsi une approche possible de la traçabilité.²⁹ L'espace européen des données de santé offre un autre exemple sectoriel, avec ses exigences de certification et d'interopérabilité applicables aux systèmes de dossiers médicaux électroniques. Un déploiement digne de confiance reposera sur un ensemble cohérent associant traçabilité des modèles, évaluation, tests sectoriels, rigueur dans la commande publique et interopérabilité transfrontalière.

Néanmoins, d'importants contre-arguments subsistent : l'Europe risque de ne pas investir suffisamment dans les capacités de pointe, les modèles ouverts ne suppriment pas sa dépendance à l'égard des matériels américains ou asiatiques et un écart croissant en matière de qualité ou d'écosystème pourrait réduire les bénéfices tirés de son expertise en matière de déploiement. Ces risques justifient de poursuivre les investissements dans l'IA de pointe. Ils ne signifient pas pour autant que l'imitation des laboratoires américains ou chinois puisse, à elle seule, constituer une stratégie industrielle suffisante.

²⁹ IBM, [Granite 4.1 models](#) (2026).

II. CINQ THÈSES STRATÉGIQUES

A. UNE APPROCHE PAR PORTEFEUILLE PEUT RÉDUIRE LE COÛT MOYEN DE L'INTELLIGENCE

Les tarifs publics font déjà apparaître des écarts considérables entre les modèles haut de gamme et les modèles plus légers.³⁰ Cela ne signifie pas pour autant qu'un modèle de petite taille conviendra nécessairement à une tâche donnée. En revanche, l'évaluation, l'allocation et le routage des modèles peuvent avoir une incidence significative sur les coûts. Les entreprises qui adaptent les capacités des modèles aux exigences de leurs processus pourraient ainsi obtenir de meilleurs résultats que celles qui recourent systématiquement aux solutions haut de gamme.

B. DANS L'IA D'ENTREPRISE, LES DONNÉES PROPRIÉTAIRES NE SONT PAS LE NOUVEAU PÉTROLE : ELLES SONT LE NOUVEAU FONDS DE ROULEMENT

Le fonds de roulement alimente les processus et soutient l'activité. Par analogie, les données prennent de la valeur lorsqu'elles peuvent jouer un rôle comparable. La RAG, la récupération multimodale, les espaces de données et les assistants fondés sur des connaissances sectorielles permettent de transformer le savoir institutionnel en ressources directement exploitables. Le règlement européen sur la gouvernance des données et les initiatives liées à l'union des données revêtent donc une importance particulière : ils peuvent réduire les obstacles à une mobilisation rapide des données, dans des conditions à la fois sûres et conformes au droit.

C. LA MEILLEURE CARTE DE L'EUROPE POURRAIT ÊTRE LE DÉPLOIEMENT SPÉCIALISÉ, PLUTÔT QUE L'IMITATION DES ACTEURS DE POINTE

L'IA de pointe se caractérise par de fortes économies d'échelle, une concentration des capacités de calcul et des écosystèmes fortement polarisés, autant de domaines dans lesquels l'Europe part avec un désavantage. L'économie du déploiement obéit toutefois à d'autres

³⁰ Pages tarifaires publiques : Google, [Gemini Developer API pricing](#) ; Mistral AI, [Pricing](#) (consultées le 26 août 2026).

logiques : la confiance des institutions, la profondeur de l'intégration, l'expertise sectorielle et le poids de la commande peuvent y jouer un rôle plus déterminant. L'Europe pourrait se construire un avantage en coordonnant ses efforts en tant qu'ensemble d'acteurs adoptant l'IA de manière stratégique, à condition de conserver des capacités suffisantes dans l'IA de pointe pour évaluer les meilleurs systèmes et les adapter à ses propres besoins.

D. LA GOUVERNANCE PEUT DEVENIR UNE TECHNOLOGIE DE PRODUCTION

Dans les secteurs où les exigences de confiance sont élevées, la gouvernance ne peut pas être reléguée au rang de simple préoccupation juridique intervenant a posteriori. Documentation, évaluation, certification et maîtrise du comportement des systèmes sont autant de composantes opérationnelles qui conditionnent leur capacité à être déployés à grande échelle. L'architecture de confiance peut donc faire partie intégrante du processus de production, plutôt que rester extérieure à ce dernier.

E. L'ORCHESTRATION DEVIENDRA UN ENJEU ORGANISATIONNEL CENTRAL DE L'IA DÉCENTRALISÉE

À mesure que les capacités se répartissent entre modèles locaux, chaînes de traitement multimodales et systèmes agentiques, la responsabilité peut se trouver fragmentée entre le fournisseur du modèle, le composant de récupération, l'orchestrateur, l'acheteur et l'organisation qui déploie la solution. L'Europe devrait traiter explicitement cette question de répartition des responsabilités. Une attribution claire des responsabilités tout au long d'une chaîne de modèles distribuée pourrait devenir un levier concret de confiance et d'accès au marché.

III. RECOMMANDATIONS

Les thèses exposées précédemment font ressortir quatre priorités institutionnelles.

1. Faire des usines d'IA de véritables structures dédiées au déploiement :

relier les usines d'IA à des laboratoires de données sectoriels, à des dispositifs d'évaluation multilingue, à des services d'accompagnement au déploiement pour les PME et à des formations pratiques. Les capacités de calcul sont essentielles, mais le véritable critère de réussite est la capacité des entreprises et des organismes publics à transformer cet accès en processus opérationnels réellement exploitables.

2. Faire de la commande publique un levier de confiance :

faire de la portabilité, des journaux d'audit, de la documentation des modèles, de la transparence contractuelle et de la traçabilité des sources utilisées pour la récupération de l'information des exigences de référence lorsqu'elles sont pertinentes au regard du risque et du cas d'usage. Cette approche permettrait de réduire les risques pour le secteur public tout en créant une demande pour des systèmes conçus dès l'origine pour être gouvernables et interopérables.

3. Faire passer les entreprises de l'achat d'un modèle unique à la gestion d'un portefeuille de modèles :

définir des règles de routage, des architectures de récupération de l'information, différents niveaux de modèles locaux et distants, ainsi que des critères portant sur le coût, la latence, la confidentialité, la fiabilité et la responsabilité juridique. Dans les secteurs réglementés, cette démarche devrait s'appuyer sur des inventaires de modèles, des déclarations d'appétence au risque et des fonctions de deuxième ligne disposant des compétences nécessaires.

4. Mutualiser les infrastructures de validation dans les secteurs à fortes exigences de confiance :

les systèmes de santé, les universités et les organismes de financement de la recherche devraient considérer les capacités de validation comme une ressource commune. Des jeux de données sectoriels, des benchmarks multilingues, des chaînes d'outils de récupération de l'information et des bancs d'essai pour l'inférence en périphérie peuvent permettre de mener des travaux qu'aucune organisation

n'aurait intérêt à développer seule. La priorité doit aller à des dispositifs d'évaluation rigoureux et réutilisables dans la santé, la finance et les services publics.

CONCLUSION

Les perspectives de l'Europe en matière d'IA ne sauraient se résumer à la capacité de posséder le modèle le plus puissant au monde. Le continent reste en retrait dans la course aux systèmes de pointe à très grande échelle, mais il dispose d'atouts crédibles pour déployer, au sein des entreprises comme des institutions publiques, des systèmes auditable, multimodaux, aux coûts maîtrisés et fondés sur des connaissances sectorielles. L'architecture de politiques publiques qui se dessine à Bruxelles fournit des briques utiles, mais elle ne constitue pas en soi la preuve d'un avantage comparatif. Pour transformer ces fondations en véritable avantage, l'Europe devra disposer de compétences de déploiement, d'infrastructures accessibles, de données exploitables et d'une gouvernance proportionnée. La règle pratique est simple : recourir à des modèles plus petits lorsqu'ils répondent aux besoins de la tâche, à des systèmes multimodaux lorsque le texte ne suffit pas, à un déploiement local lorsque la latence ou la confidentialité l'exigent, et à des dispositifs de contrôle plus stricts lorsque le niveau de risque est élevé.

ANNEXE

A. LE MARCHÉ DE L'IA EN 2026

Famille de modèles	Taille / Échelle	Accès	Auto-hébergement	Usages adaptés
OpenAI : GPT-5.4 mini	Non communiquée	API prioritaire	Poids non publics	Tâches rapides en entreprise, programmation, utilisation d'outils et processus agentiques.
Anthropic : famille Claude Haiku	Non communiquée	API prioritaire	Poids non publics	Charges de travail Claude à faible latence, capacités variables selon les versions.
Google : famille Gemini Flash-Lite	Non communiquée	API prioritaire	Poids non publics	Charges de travail multimodales légères et à grande échelle.
Meta : Llama 4 Scout	17 milliards de paramètres actifs, 109 milliards au total, MoE	Poids ouverts sous licence Llama 4 Community License	Oui	Charges de travail multimodales déployées localement ou hébergées par un partenaire, avec contexte étendu.
Google : famille Gemma	Plusieurs variantes petites et moyennes à poids ouverts	Poids ouverts, licence Apache 2.0	Oui	Déploiement en périphérie, sur ordinateurs portables et stations de travail. Applications multimodales et multilingues.
Mistral : Small 4	119 milliards de paramètres au total / 6,5 milliards actifs, MoE	Poids ouverts, licence Apache 2.0	Oui	Déploiements européens multimodaux, personnalisables, dotés de capacités de raisonnement et agentiques
Hugging Face : SmolLM3	3 milliards de paramètres	Poids ouverts	Oui	Raisonnement local et multilingue en environnement ouvert, et ajustement fin

Exemples d'options pour la constitution d'un portefeuille de modèles (au 26 août 2026)³¹

³¹ Liste indicative et non exhaustive. Documentation officielle : OpenAI, [Models](#) ; Anthropic, [Models overview](#) ; Google, [Gemini models](#) et [Gemma model card](#) ; Meta, [Llama 4](#) ; Mistral AI, [Mistral Small 4](#) ; Hugging Face, [SmolLM3](#). Les modèles, les tarifs et les licences sont susceptibles d'évoluer.

skema
THINK TANK

PUBLIKA

SKEMA Publika

SKEMA Business School, Campus Grand Paris
5 Quai Marcel Dassault – CS 90067
92156 Suresnes Cedex, France

Tél. : +33.1.71.13.39.32
Courriel : publika@skema.edu
Site Internet : www.publika.skema.edu