

Septembre 2025

Intelligence artificielle

Une bombe énergétique ?

Intelligence artificielle

Une bombe énergétique ?

A propos de ce rapport

L'intelligence artificielle connaît aujourd'hui une croissance rapide, portée par des avancées majeures dans le développement de modèles algorithmiques et par la puissance croissante des infrastructures de calcul. Des investissements considérables dans le domaine sont prévus et la demande est importante de la part des utilisateurs à tous les niveaux. L'essor de l'IA n'est pas un phénomène temporaire, et soulève des interrogations croissantes sur son impact environnemental et sa consommation énergétique.

Ce rapport propose une analyse technologique et prospective de la consommation énergétique de l'IA, en retraçant ses évolutions historiques, en identifiant les moteurs de cette consommation, et en évaluant les scénarios futurs possibles. L'étude couvre la consommation de l'IA à travers celle des data centers, mais n'analyse pas la consommation indirecte en amont des data centers, telles que les consommations liées à la fabrication de matériel. Ce rapport examine également les conséquences potentielles de cette trajectoire. Enfin, il explore les leviers d'action disponibles à l'échelle française et européenne pour concilier développement de l'IA, soutenabilité et souveraineté énergétique.

Conditions de diffusion

Cette publication est mise à disposition sous licence libre. La reproduction et la diffusion totale ou partielle des textes de cette publication sont possibles pour toute utilisation non commerciale, en l'état et sans modification.

Comment citer ce rapport : Zhang L., Dupin J. *Intelligence artificielle : une bombe énergétique ?*, Rapport Zenon Research. Septembre 2025.

Où trouver ce rapport

Ressource en ligne / Elle est disponible gratuitement sur zenon.ngo

Auteurs

Lucie Zhang

Analyste

Zenon Research

Sous la direction scientifique
de **Jean-Baptiste Dupin**,
Zenon Research

Relecteurs

Cédric Philibert

Expert énergie climat, membre
du conseil scientifique de Zenon,
chercheur associé à l'Ifri

Claude Tadonki

Enseignant-Chercheur en informatique
à Mines Paris – PSL

Xavier Blot

Professeur en changement climatique
et transition énergétique
à l'emlyon business school



	Résumé exécutif	04
1	Consommation énergétique de l'IA et des data centers	05
	La consommation de l'IA, entre entraînement et inférence	06
	Les data centers, colonne vertébrale de l'IA	07
	Evolution historique des consommation de l'IA et des data centers	12
	Empreinte environnementale de l'IA	14
2	Besoins énergétiques de l'IA et des data centers d'ici 2030	15
	Panorama des scénarios d'évolution	16
	Les conséquences d'une forte demande énergétique de l'IA	22
3	Enjeux stratégiques à l'ère de l'intelligence artificielle	28
	Echelle européenne	29
	Attractivité du territoire français	31
	Références	35

Résumé exécutif

La consommation énergétique des data centers est estimée à 415 TWh dans le monde en 2024, soit 1,3 % de la consommation électrique mondiale aujourd'hui. Si à l'échelle planétaire, cette consommation ne semble pas importante, c'est sur les réseaux électriques locaux que la répercussion est grande. En effet, **la répartition des data centers est loin d'être homogène dans le monde** : ces infrastructures sont largement regroupées en cluster dans des régions spécifiques – Etats-Unis, Asie (Chine), et une poignée de pays en Europe.

L'IA représente environ 15 % du besoin énergétique des data centers aujourd'hui, soit environ 62 TWh, bien qu'il soit difficile d'établir précisément sa consommation. L'augmentation de la taille des modèles, censée garantir de meilleures performances, provoque l'allongement des temps d'entraînement et l'augmentation de la puissance des processeurs utilisés. L'utilisation croissante de l'IA dans les usages quotidiens et économiques entraîne également l'augmentation de sa consommation.

La trajectoire d'évolution de la consommation énergétique de l'IA et des data centers est incertaine, car de nombreux facteurs influent ces derniers. **Les scénarios les plus plausibles d'ici 2030 estiment une consommation entre 200 et 400 TWh d'ici 2030.** L'IA pourrait passer ainsi de 15 % de la consommation des data centers aujourd'hui à 35 à 50 % en 2030. La demande énergétique des data centers pourrait plus que doubler, et passer de 415 TWh à 945 TWh. La tendance globale est ainsi à la hausse, avec différentes amplitudes selon les régions : les data centers en Chine devraient connaître une forte croissance (x3,3 en moyenne), de même qu'aux Etats-Unis (x1,8 à x3,3), et au sein de l'Union Européenne (x2,7 en moyenne). Ces augmentations s'expliquent notamment par les investissements publics et privés dans l'IA qui provoquent l'augmentation de la puissance des data centers (dépassant souvent plusieurs centaines de MW, voire 1 GW).

La croissance de la demande des data centers mettra sous tension les réseaux électriques. En effet, les data centers spécialisés pour l'IA visent dans le même temps à augmenter leur puissance et à être alimentés en électricité décarbonée. Les réseaux sont soumis à des surcharges locales et les gestionnaires doivent anticiper l'augmentation de la demande pour adapter les réseaux. **Les infrastructures de production d'électricité doivent à la fois fournir une quantité importante d'électricité, stable (disponible en continu), à faible coût, décarbonée, et dans un délai rapide.** A court terme, la sécurisation d'un approvisionnement électrique stable et rapide prime sur la décarbonation du mix électrique des data centers, c'est pourquoi beaucoup d'acteurs misent sur le gaz comme source d'énergie primaire pour produire de l'électricité. Néanmoins, cette solution n'est pas non plus infaillible : la construction de centrales à gaz subit des retards sur les chaînes d'approvisionnement. La production d'électricité sur site est une autre solution privilégiée permettant de contourner les délais de raccordement au réseau local. À moyen terme, les entreprises de la tech misent sur le nucléaire et les petits réacteurs modulaires, ainsi que les technologies de stockage et de flexibilité, qui présentent des incertitudes en termes de déploiement et de maturité.

Pour se démarquer dans la course à l'IA, menée sous le prisme de la consommation énergétique, l'Europe cherche à attirer la construction de data centers, grâce à une fiscalité avantageuse comme en Irlande ou une production d'électricité bas carbone comme en France. L'Europe gagnerait également à développer son propre écosystème d'acteurs et doit miser sur le développement d'IA frugale et éthique. La récupération et la valorisation de la chaleur des data centers ainsi que l'exploitation de leur flexibilité sont aussi des enjeux clés pour permettre de considérer ces infrastructures comme des acteurs du réseau, voire des leviers pour accélérer la transition énergétique.

1

Consommation énergétique de l'IA et des data centers

La consommation de l'IA, entre entraînement et inférence

Les data centers, colonne vertébrale de l'IA

Evolution historique des consommations de l'IA et des data centers

Empreinte environnementale de l'IA

La consommation de l'IA, entre entraînement et inférence

L'intelligence artificielle est un ensemble de technologies permettant aux machines d'effectuer des tâches qui nécessitaient historiquement l'intelligence humaine, telles que la prise de décision, la prédiction, la créativité. Les modèles d'intelligence artificielle de nos jours sont d'abord entraînés sur des données (**phase d'entraînement**), avant d'être déployés dans des environnements de production pour traiter et réagir à des nouvelles données similaires (**phase d'inférence ou d'utilisation**).

La phase d'entraînement représente un pic de calcul particulièrement intense et mesurable. En effet, l'entraînement d'une IA peut durer plusieurs semaines sur des infrastructures dédiées et mobilise des ressources informatiques importantes, tandis que la phase d'inférence est plus diffuse et dépend des usages. Entraîner ChatGPT-4 sur plusieurs semaines aurait consommé environ 50 GWh d'électricité (Ludvigsen, 2023), soit la consommation annuelle d'environ 19 600 voitures électriques.

Avec la multiplication des déploiements de l'IA dans les activités économiques ainsi que l'essor des IA génératives et généralistes, **la consommation de la phase d'inférence devient un enjeu majeur.** Des calculs récents indiquent que le coût énergétique d'une requête ChatGPT serait de 0,3 Wh (You, 2025). **Il n'existe pas aujourd'hui de méthode standardisée pour calculer la consommation énergétique d'une IA.** Il s'agit d'une problématique de taille : les entreprises du numérique ne fournissent pas ou ne parviennent pas à fournir de données complètes sur le sujet.

La consommation énergétique d'un modèle IA peut se décomposer comme suit (voir Figure 1 et Figure 2) :

Phase d'entraînement : La consommation est estimée à partir du temps d'entraînement multiplié par le nombre de processeurs et la puissance électrique unitaire du type de processeur employé ainsi que des unités périphériques auxiliaires. En réalité, les processeurs ne sont pas sollicités à leur capacité maximale, c'est pourquoi on tient également compte du taux d'utilisation moyen des équipements.

- Limites : Le type d'infrastructure de calcul et la durée d'entraînement sont les facteurs clés, or ces informations ne sont pas toujours disponibles.

Phase d'inférence : La méthode la plus répandue consiste à multiplier le nombre de requêtes par la consommation unitaire d'une requête.

- Limites : Cette approche fonctionne pour les IA déployées sous forme de services et fonctionnant de manière interactive, comme les LLM¹. Dans d'autres contextes, les requêtes sont automatiques ou réalisées sur une base fréquentielle. L'inférence peut être continue ou bien effectuée sur du matériel embarqué. Il est alors préférable de considérer la consommation moyenne du matériel multipliée par la durée d'utilisation. dépend des caractéristiques intrinsèques au modèle (nombre de calculs –ou FLOP² – effectués pour répondre) et au matériel utilisé (efficacité énergétique, taux d'utilisation qui peut varier dans le temps), dont on ne connaît souvent pas les détails.

¹ *Large Language Model* ou grand modèle de langage. Un LLM est un modèle possédant un grand nombre de paramètres, capable de comprendre et de générer du texte.

² *Floating-point operation* ou opération en virgule flottante, une unité de mesure du calcul numérique, utilisée pour quantifier la charge de travail ou la complexité d'un algorithme. Le FLOPS, floating-point operation per second, désigne quant à lui le nombre de FLOP effectués par seconde : il s'agit d'une mesure de rapidité et performance d'un système informatique.

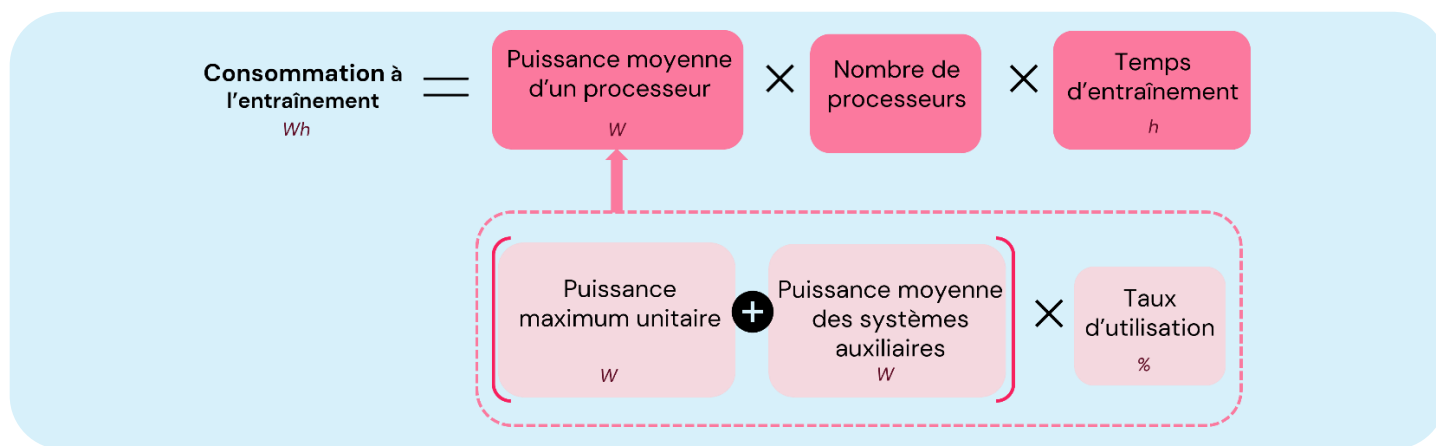


Figure 1 - Estimation théorique de la consommation énergétique de l'entraînement d'une IA

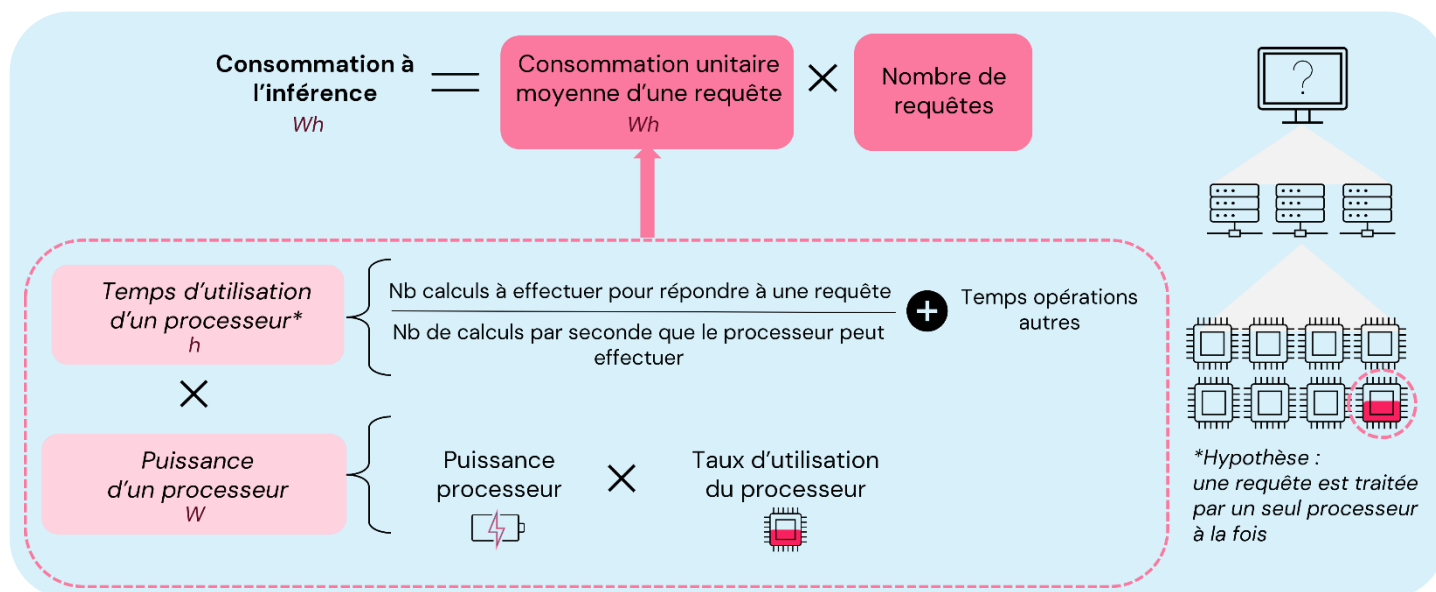


Figure 2 - Estimation théorique de la consommation énergétique de l'inférence d'une IA. Les opérations autres correspondent aux autres mécanismes intervenant dans la réponse à une requête (accès aux données, transferts, vérifications...). Le temps des « opérations autres » est fortement variable.

Les data centers, colonne vertébrale de l'IA

Les data centers servent à stocker, traiter et transmettre les informations numériques. Au cœur des data centers se trouvent les serveurs qui exécutent les traitements informatiques. Ces serveurs sont connectés via une infrastructure réseau interne qui assure la circulation des données. L'ensemble est alimenté par une infrastructure électrique redondante avec des transformateurs, des onduleurs et des générateurs diesel pour garantir la continuité de service en cas de coupure (voir Figure 3). Les data centers génèrent de la chaleur et nécessitent des systèmes de refroidissement pour garantir que la température des serveurs reste située dans leur plage de fonctionnement. Le refroidissement se fait généralement en diffusant de l'air froid dans les salles de serveurs, mais de nouvelles technologies utilisant des fluides diélectriques se développent

pour refroidir directement les équipements. Enfin, des systèmes de sécurité physique et logique protègent les équipements contre les intrusions, les sinistres et les cyberattaques.

Les data centers peuvent être différenciés selon plusieurs catégories. Dans la suite de ce rapport, la mention des data centers englobe l'ensemble des différents types de data centers, sauf mention explicite du contraire.

- Les **data centers d'entreprise** sont exploités par une entreprise pour ses propres besoins et sur site. L'entreprise est responsable du déploiement, de la surveillance et de la gestion du data center.
- Les **data centers en colocation** sont des locaux partagés où plusieurs entreprises peuvent louer de l'espace physique pour installer et exploiter leur propre matériel informatique. Le fournisseur de colocation gère l'infrastructure de base (bâtiment,

électricité, refroidissement, sécurité physique), mais l'entreprise est toujours responsable de ses propres serveurs.

- Les **data centers hyperscale** sont des centres de données de grande taille, conçus pour répondre aux demandes croissantes des services cloud. Ces infrastructures ont typiquement une capacité de 20 MW pouvant aller jusqu'à 100 MW. Ils sont principalement gérés par les géants du cloud et du numérique (Amazon Web Service, Google Cloud, Microsoft Azure...).
- Les **data centers edge** constituent une nouvelle génération de data centers, de taille plus petite et de puissance plus faible (généralement <2 MW). Ils sont situés plus près des utilisateurs finals, ce qui permet de diminuer la latence dans le transfert des données.

Des data centers dits "accélérés" ou spécialisés pour l'IA existent, notamment pour l'entraînement de modèles. Au sein d'un data center il peut également y avoir des "clusters IA", des infrastructures composées de nombreux serveurs interconnectés et conçues pour fonctionner de façon coordonnée pour servir l'IA. Les serveurs spécialisés IA utilisent des processeurs graphiques GPU³. Contrairement aux processeurs traditionnels CPU⁴, les GPU peuvent calculer de manière massivement parallèle, ce qui permet d'accélérer l'entraînement des IA.

Mais en conséquence de leur forte capacité de calcul, les puissances électriques demandées par les GPU sont importantes :

- Un CPU Intel a un TDP⁵ entre 150 et 500 W (Intel, 2025).
- Un GPU de NVIDIA (environ 90 % du marché des processeurs graphiques pour IA en 2025) a un TDP entre 300 W et 1 kW.

³ Graphics Processing Unit

⁴ Central Processing Unit

⁵ Thermal Design Power, quantité maximale de chaleur qu'un processeur est conçu pour dissiper lorsqu'il fonctionne à pleine charge, dans des conditions normales. Il est exprimé en watts (W) et sert principalement à guider la conception des systèmes de refroidissement des composants électriques. Le TDP ne correspond pas exactement à la consommation électrique d'un processeur mais c'est une bonne approximation.

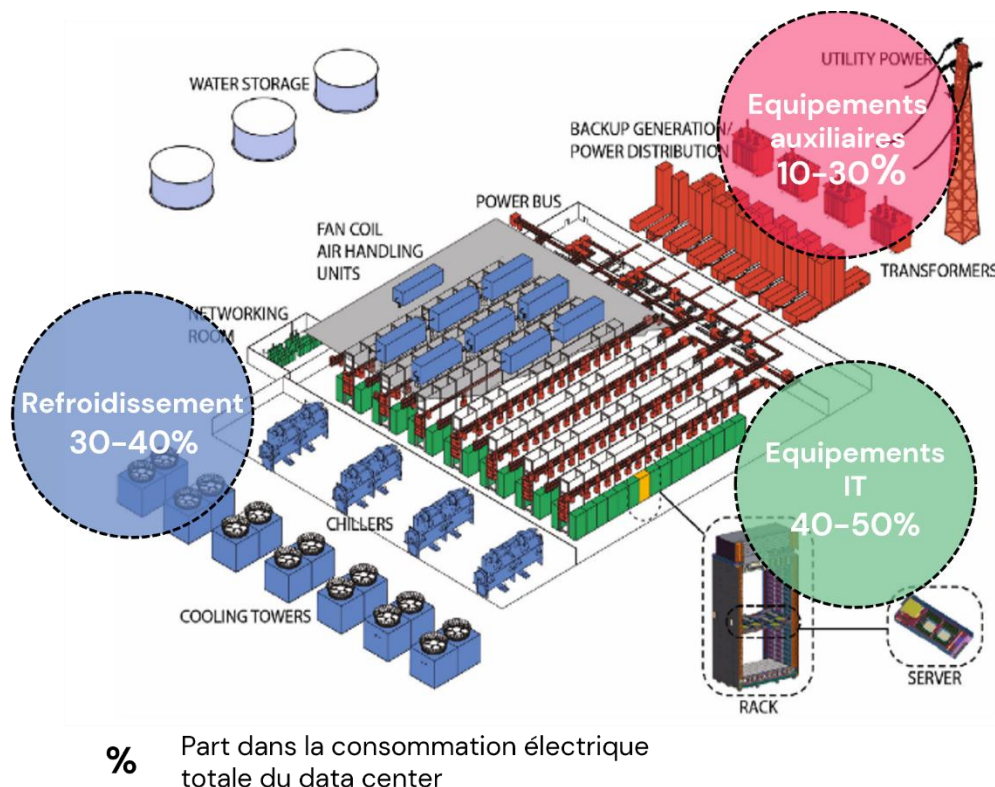


Figure 3 - Composition d'un data center et répartition de la consommation selon les différents postes (adapté de Barroso et al., 2018)

Dans un data center, une part importante de l'énergie consommée ne provient pas des équipements informatiques eux-mêmes, mais des systèmes de refroidissement, qui peuvent représenter jusqu'à 30-50 % de la consommation totale. En particulier, les data centers d'entreprise ont souvent recours à des systèmes de refroidissement basiques (climatisation), tandis que les data centers hyperscale ont recours à l'air extérieur, au refroidissement par liquide ou à des systèmes de refroidissement optimisés pour répondre à la forte densité thermique des racks de serveur. Le PUE (Power Usage Effectiveness) mesure ainsi le rendement électrique global d'un data center :

$$PUE = \frac{\text{Consommation totale du data center}}{\text{Consommation des équipements IT}}$$

Pour autant, un PUE proche de 1 ne signifie pas que la consommation du data center est faible. Le PUE ne rend également pas compte de la localisation géographique du site, laquelle peut limiter les besoins en refroidissement et les consommations d'énergie associées, ni de l'évolution du data center dans le temps (ex : augmentation progressive du nombre d'équipements informatiques installés). Le PUE moyen de l'ensemble global des data centers est

actuellement de 1,5 mais varie en fonction du type et de la taille des data centers. Les data centers hyperscale ont un PUE plus faible que les data centers d'entreprise ou de petite taille et puissance, grâce à une conception optimisée et des technologies de refroidissement avancées.

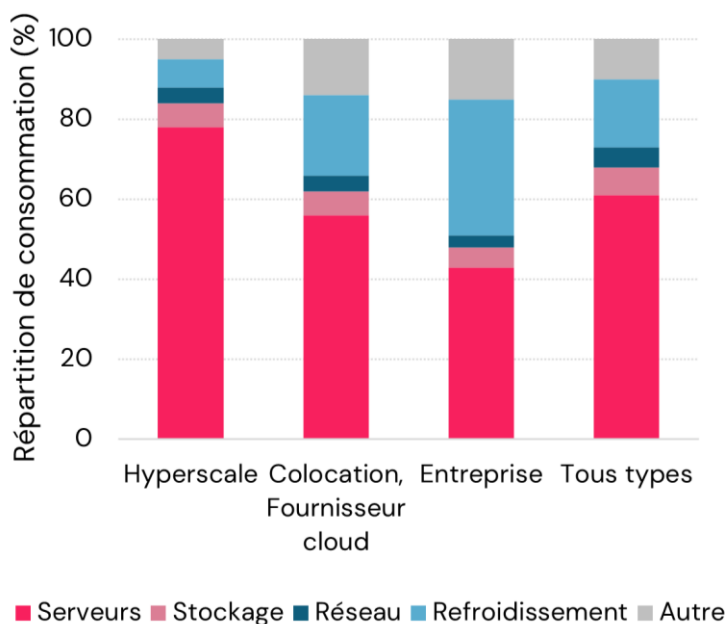


Figure 4 - Répartition des postes de consommation selon le type de data center (IEA, 2025)

La distribution géographique des data centers n'est pas homogène. A l'échelle mondiale, les Etats-Unis concentrent aujourd'hui à eux seuls la moitié (plus de 5 000) des data centers existants (Statista, 2025a). L'Europe constitue également un pôle avec près de 2 500 data centers, concentrés principalement en Allemagne (521), au Royaume-Uni (514), en France (315) et aux Pays-Bas (297). Plus de 400 data centers se trouvent en Chine. **La répartition des data centers en termes de puissance est également hétérogène : concentrés dans les pays développés, environ 97 GW de puissance sont installés⁶ dans le monde en 2024, dont 42 GW aux Etats-Unis, 16 GW en Europe et 24 GW en Chine (IEA, 2025).** Des clusters de plusieurs GW se démarquent dans l'est des Etats-Unis et de l'Asie, ainsi que sur un arc Francfort-Londres-Amsterdam-Paris-Dublin. Ces clusters sont notamment localisés sur des nœuds de connectivité internet (câbles sous-

marins), garantissant une faible latence pour la transmission de données. Les politiques fiscales avantageuses comme en Irlande ou à Singapour attirent aussi ces infrastructures. La présence de nombreux hyperscalers aux Etats-Unis et en Chine peut également s'expliquer par la présence d'un marché interne massif et de géants du cloud nationaux, ainsi que par la disponibilité et le coût de l'énergie. L'Europe constitue un marché fragmenté, sans compétiteur face aux GAFAM et avec un coût de l'électricité plus élevé.

Ces concentrations massives créent des points de tension pour les réseaux électriques locaux et soulèvent aussi des enjeux de souveraineté numérique, à travers une forte dépendance de nombreuses régions du monde aux infrastructures concentrées ailleurs.



Figure 5 – Carte de la répartition mondiale des data centers en puissance (adapté de IEA, 2025). Les clusters de data centers sont définis comme les groupes de data centers situés à moins de 100 km les uns des autres. La puissance représentée est uniquement celle des serveurs installés (puissance informatique) et ne prend pas en compte les autres équipements d'un data center.

⁶ La puissance installée d'un data center correspond à la puissance maximale que la data center peut soutenir au réseau. En réalité, ces infrastructures fonctionnent rarement à leur puissance maximale, et tournent plutôt autour de 60-70 % de cette dernière.

La consommation globale des data centers⁷ en 2024 est estimée à 415 TWh, ce qui représente 1,3 % de la consommation électrique globale. La part de l'IA dans la consommation des data centers est estimée à environ 15 % aujourd'hui, soit 62 TWh à l'échelle mondiale en 2024 (IEA, 2025). Bien qu'il existe des data centers entièrement dédiés à l'entraînement d'IA, il reste en réalité très difficile d'isoler la consommation attribuable à l'IA dans un data center. Cette consommation est approchée par celle des serveurs accélérés à base de GPUs. Néanmoins, cette méthode reste approximative, car l'inférence d'une IA peut être faite sur des serveurs conventionnels, et inversement, les serveurs accélérés peuvent servir à héberger des applications autres que IA, comme des calculs scientifiques.

À l'échelle globale, la consommation des data centers peut sembler négligeable, mais leur expansion peut avoir des répercussions significatives sur les réseaux et ressources locales. Aux Etats-Unis, les data centers représentent, en 2023, 4,4 % de la consommation électrique nationale. Ils exercent une **pression croissante sur**

les réseaux électriques locaux, pouvant contribuer à des tensions sur l'approvisionnement, des hausses de prix, ou encore à des besoins accrus en infrastructures de production d'électricité.

De tels phénomènes sont déjà observables en Irlande, où la consommation des data centers représentait 18 % de la consommation électrique nationale en 2022 (European Commission. Joint Research Centre., 2024). Ce pays compte 82 data centers pour une capacité totale installée de plus de 2 GW⁸, nettement supérieure aux 566 MW des 300 data centers présents en France en 2022 (GIMELEC, 2024 ; RTE, 2025b).

Face à la saturation croissante du réseau irlandais, aux risques de black-out et aux projections indiquant une part de 30 % des data centers dans la consommation électrique nationale en 2030 (CRU, 2025), Eirgrid, l'opérateur du réseau électrique irlandais, a imposé un moratoire sur les projets de data centers dans la région de Dublin depuis 2022 : les nouvelles demandes de connexions pour ces infrastructures ne seront plus acceptées d'ici 2028 (Data Center Dynamics, 2022).

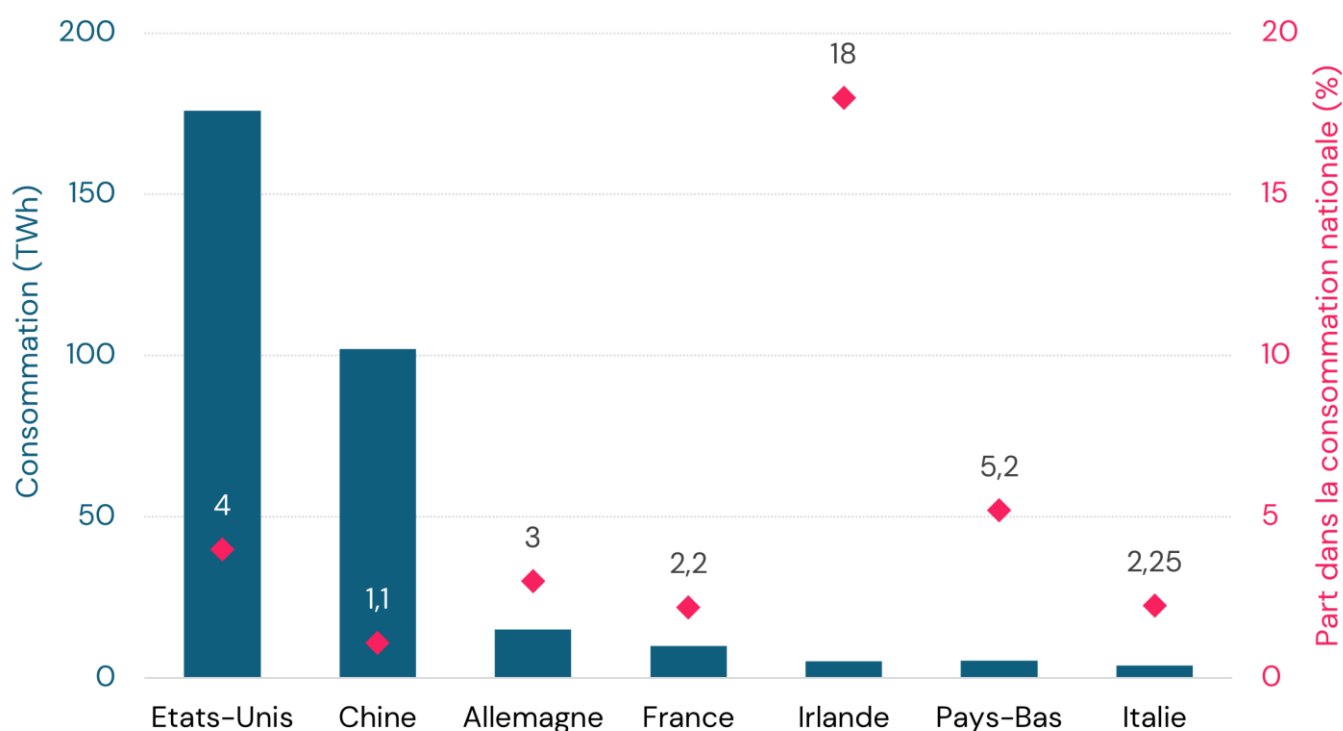


Figure 6 – Consommation des data centers et leur part dans la consommation d'électricité d'une sélection de pays : les Etats-Unis et la Chine constituent le top 2 des pays avec la plus grande consommation des data centers dans le monde, et top 5 des pays avec la plus grande consommation dans l'UE-27. (European Commission. Joint Research Centre., 2024 ; IEA, 2025 ; RTE, 2025b ; Shehabi et al., 2024)

⁷ Les données présentées dans ce rapport n'incluent pas les data centers dédiés à la cryptomonnaie. Ces derniers reposent sur des équipements IT différents des data centers traditionnels. De plus, la collecte de données de consommation de l'industrie de la cryptomonnaie est encore plus limitée.

⁸ L'Irlande s'est en effet dotée d'une fiscalité avantageuse qui a attiré les géants du cloud et leurs data centers hyperscale, tels que Google et Amazon. Son taux d'imposition sur les sociétés est en effet de 12,5 %, contre 20-25 % dans de nombreux pays européens.

Evolution historique de la consommation de l'IA et des data centers

Historiquement, entre 2010 et 2018, la consommation énergétique des data centers dans le monde est restée stable autour de 100 TWh par an (Kamiya & Coroamă, 2025). Pourtant, cette même période a enregistré une forte croissance du trafic de données et de la demande numérique, avec une multiplication par six des charges de travail des data centers. Cette stabilité s'explique par des gains d'efficacité significatifs dans la gestion des infrastructures : la quantité d'énergie nécessaire par unité de calcul a été divisée par quatre pour les serveurs standards grâce à l'amélioration de l'efficacité des processeurs. Les charges de travail ont également basculé vers des infrastructures cloud ou hyperscale ayant des systèmes de refroidissement et d'alimentation plus efficaces (Masanet et al., 2020).

Toutefois, à partir de la période 2018–2020, une hausse de la consommation est observée. En effet, la consommation globale des data centers aurait augmenté d'environ 16 % entre 2010 et 2015, contre environ 50 % entre 2018 et 2022 (IEA, 2025). Plusieurs raisons sont avancées pour expliquer cette augmentation de consommation :

- **L'explosion des usages numériques** liés au cloud et à la digitalisation (télétravail, services numériques, réseaux sociaux...) qui fait augmenter le nombre de serveurs et leurs puissances.
- **Le ralentissement des gains d'efficacité énergétique du matériel informatique** : la consommation d'énergie par tâche informatique⁹ a diminué d'environ 30 % entre 2005 et 2015, et seulement de 10 % entre 2015 et 2023 (IEA, 2025).
- **L'émergence des IA génératives et des LLM** : les investissements privés en IA, qui s'accompagnent généralement de déploiement d'infrastructures de calcul et d'entraînement, s'élevaient à environ 130 Md\$ en 2021, contre seulement 25 Md\$ en 2017.

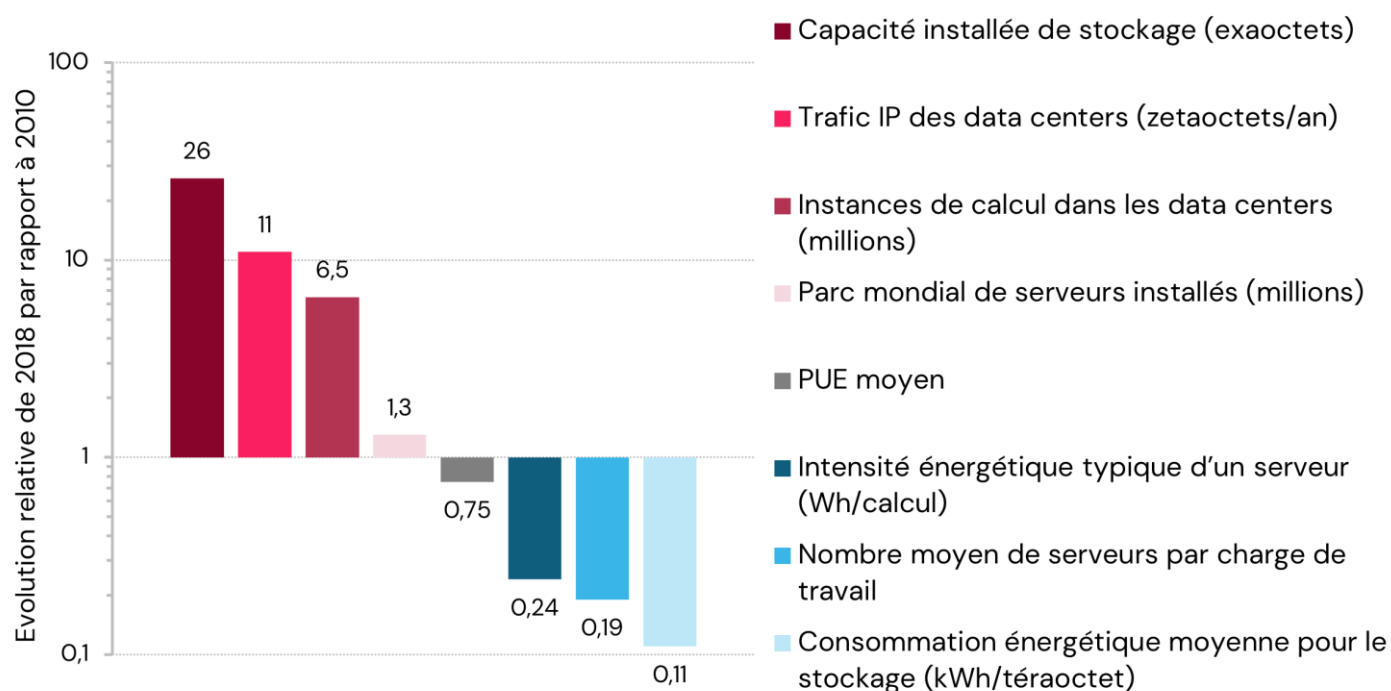


Figure 7 – Evolution relative entre 2010 et 2018 des facteurs influençant la consommation des data centers (Masanet et al., 2020)

⁹ Une tâche informatique est une unité de travail qu'un ordinateur doit exécuter (ex. lancer un programme, effectuer un calcul, exécuter une requête d'IA, sauvegarder un fichier...).

Les estimations de la consommation actuelle des data centers divergent très fortement à partir de 2020, variant entre 200 à plus de 600 TWh pour l'année 2022, reflétant des incertitudes méthodologiques et un manque de données fiables (voir Figure 8).

Les Etats-Unis hébergeant une part importante des infrastructures des principaux géants de la tech, ce pays constitue un environnement clé pour comprendre les dynamiques énergétiques à grande échelle (voir Figure 9). En ligne avec la tendance mondiale, la consommation énergétique des data centers y est restée historiquement stable, mais à partir de 2018, une augmentation notable de cette consommation est observée (+75 % entre 2020 et 2023). Cette hausse semble fortement liée à la montée en puissance des usages liés à l'intelligence artificielle : **environ 22 % de la consommation des data centers aux Etats-Unis en 2023 est attribuée à l'intelligence artificielle, contre moins de 5 % en 2018** (Shehabi et al., 2024).

La croissance actuelle rapide de consommation de l'IA s'explique par une logique d'amplification, tant du côté de l'entraînement que de l'inférence des modèles :

- **Phase d'entraînement** : les modèles les plus récents sont de plus en plus gourmands en calcul. En effet, plus un modèle est gros et entraîné sur un corpus large, plus il est performant, et ce jusqu'à l'atteinte d'un seuil de performance. Cela a pour conséquence de rallonger le temps d'entraînement, et les processeurs capables de réaliser des calculs complexes consomment davantage d'énergie.
- **Phase d'inférence** : l'explosion du nombre d'usages, le déploiement croissant de l'IA dans les activités économiques, l'augmentation du volume de requêtes générées en temps réel et la multiplication des supports générés (images, vidéos), contribuent à faire de l'inférence un poste énergétique désormais aussi critique que l'entraînement.

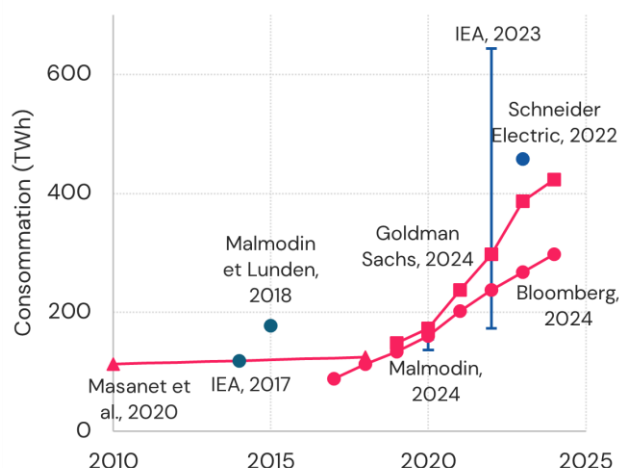


Figure 8 - Estimations de la consommation historique des data centers dans le monde (adapté de Shehabi et al. 2024)

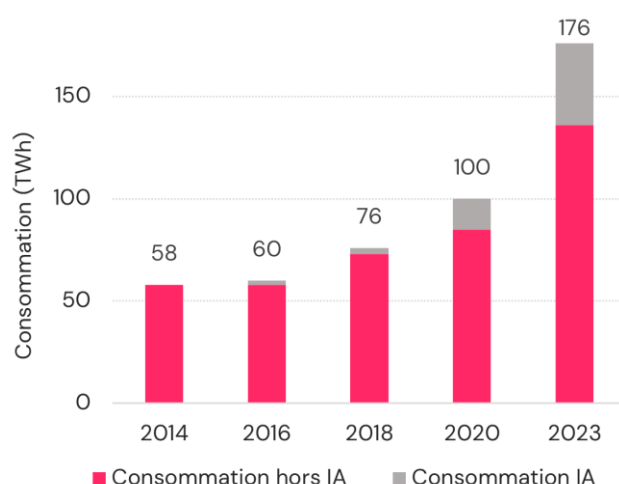


Figure 9 - Estimations de la consommation historique des data centers aux Etats-Unis (adapté de Shehabi et al. 2024)

Empreinte environnementale de l'IA

La consommation électrique de l'IA contribue à son empreinte carbone, notamment dans les zones où le mix énergétique reste majoritairement carboné. **Aujourd'hui, près de la moitié de l'électricité alimentant les data centers provient encore de sources fossiles.** Les émissions indirectes associées à la consommation d'énergie des data centers sont estimées à **180 MtCO₂e dans le monde aujourd'hui** (IEA, 2025), soit environ **0,4 % des émissions globales de CO₂.**

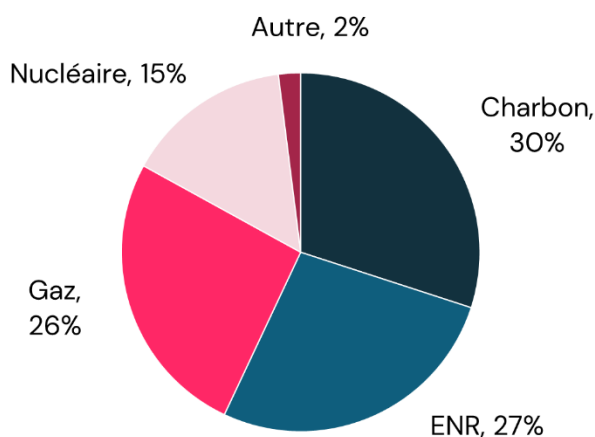


Figure 10 – Mix électrique des data centers dans le monde (IEA, 2025)

Au-delà de la consommation d'électricité, l'empreinte environnementale de l'IA repose sur d'autres facteurs. Par exemple, l'eau est souvent utilisée comme fluide caloporteur pour refroidir les data centers, ce qui contribue à une pression accrue sur les ressources hydriques. En juillet 2025, Mistral AI a publié un bilan environnemental de son modèle Large 2, basé sur le référentiel pour mesurer et réduire l'impact environnemental de l'IA développé par l'AFNOR (AFNOR, 2024; Mistral AI, 2025) ; c'est la première analyse de cycle de vie d'un modèle IA rendue publique. Le bilan tient compte de la phase de conception du modèle, du cycle de vie des équipements et infrastructures utilisées, ainsi que de l'entraînement et l'inférence du modèle. Ainsi, le bilan pour le modèle Large 2

indique que générer une page de texte consomme 45 mL d'eau. A l'échelle d'une requête, ces chiffres sont minimes, mais leur accumulation peut contribuer à augmenter significativement la consommation en eau des data centers et poser problème dans les régions soumises au stress hydrique. Toutefois, les besoins en eau des data centers restent très faibles par rapport à ceux de l'agriculture¹⁰.

L'empreinte de l'IA doit aussi s'évaluer sur l'ensemble de son cycle de vie, de la recherche de modèles à la fabrication et la fin de vie des équipements électroniques. Les émissions liées à la fabrication des processeurs proviennent pour une large part de la consommation d'électricité d'origine fossile. **Les réflexions actuelles sur l'impact environnemental de l'IA se limitent le plus souvent à sa consommation énergétique en phase d'entraînement et d'usage, sans intégrer les étapes amont et aval de son cycle de vie**, or ces dernières mériteraient d'être prises en considération, surtout dans le cas des processeurs conçus spécifiquement pour l'IA (ex : TPU¹¹ de Google). Néanmoins, certains cas sont plus complexes puisque ces processeurs peuvent, en pratique, servir à d'autres applications que l'IA. L'ACV de Mistral révèle ainsi que **l'entraînement et l'inférence du modèle Large 2 représentent 85,5 % des émissions de CO₂e de ce dernier et 91 % de sa consommation en eau.**

L'initiative de Mistral AI est un pas vers plus de transparence environnementale dans l'industrie de l'IA et elle ouvre la voie à une normalisation nécessaire. Cependant, les données de consommation énergétique sont manquantes, et l'entraînement et l'inférence sont agrégées en une seule catégorie, ce qui ne permet pas de visualiser leur impact séparé. Cette méthodologie doit être étendue aux autres acteurs, pour construire un véritable standard permettant de comparer les modèles entre eux.

Ces constats, déjà préoccupants aujourd'hui, soulèvent d'autant plus de questions quant à l'évolution des besoins énergétiques de l'IA et les conséquences qui en découleront.

¹⁰ La totalité des data centers pourrait utiliser environ 8 millions de m³ d'eau par an (Siddik et al., 2021). En comparaison, environ 70 % des prélèvements en eau douce sont destinés à l'agriculture chaque année, soit environ 2 800 Md m³ d'eau en 2021 (Our World in Data, 2025; United Nations, 2024) ; c'est 350 000 fois plus que les besoins des data centers.

¹¹ *Tensor Processing Unit*. Ce sont des accélérateurs d'IA conçus sur mesure et optimisés pour l'entraînement et l'inférence de modèles IA chez Google.



2

Besoins énergétiques de l'IA et des data centers d'ici 2030

Panorama des scénarios d'évolution

Les conséquences d'une forte demande
énergétique de l'IA

Panorama des scénarios d'évolution

D'un point de vue mondial, parmi les dizaines d'études proposant des scénarios d'évolution de la demande énergétique de l'IA, deux conclusions s'imposent :

- Les scénarios prospectifs actuels ne permettent pas d'établir une trajectoire claire d'évolution, même à court terme. **D'ici 2030, les estimations de la consommation globale de l'IA se situent entre 200 TWh et 800 TWh, reflétant une forte incertitude.**
- **Aucun scénario ne projette une stabilisation de la demande** : le besoin en énergie de l'IA est amené à croître.

L'AIE a examiné les méthodes de modélisation ainsi que les dynamiques du secteur, et juge **plausible une augmentation de la consommation de l'IA atteignant 200 à 400 TWh d'ici 2030** (Kamiya & Coroamă, 2025), soit 3 à 6 fois plus que la consommation actuelle attribuée à l'IA. Elle représenterait alors **35 à 50 % de la demande totale des data centers à l'échelle mondiale, contre**

environ 15 % aujourd'hui. Cette évolution s'inscrit dans une tendance plus large de hausse de la consommation des data centers. En effet, au-delà de l'IA, l'explosion du volume des données et leur stockage massif, la poursuite d'une tendance à l'augmentation des usages cloud et le déploiement des services numériques dans les pays en développement constituent d'autres facteurs d'augmentation de la consommation des data centers. Ces derniers pourraient voir leur consommation plus que doubler d'ici 2030, passant de 415 TWh en 2024 à 945 TWh (+127 %). Ils représenteraient alors 3 % de la consommation électrique globale.

Les projections sur la consommation des data centers sont également incertaines, car elles dépendent largement de l'évolution encore imprévisible de la demande en IA et de l'impact des normes, réglementations et contraintes opérationnelles. Des projections à plus long terme que 2030 ne sont donc pas pertinentes au vu des incertitudes autour de la trajectoire d'évolution de l'IA.

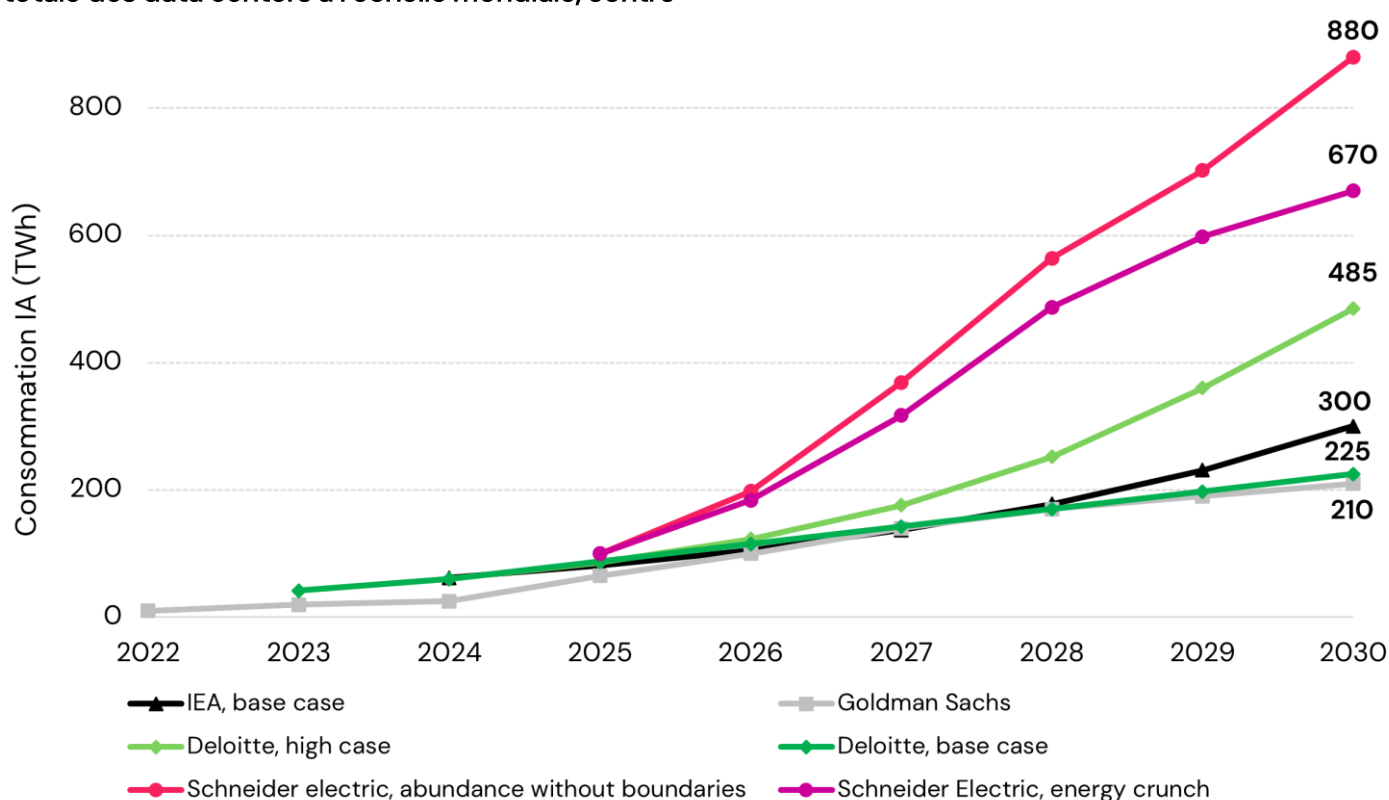


Figure 11 – Sélection de scénarios d'évolution mondiale de la consommation de l'IA (Goldman Sachs, 224 ; IEA, 2025 ; Lorentz et al., 2024 ; Paccou et Wijnhoven, 2024)

	IEA Base case	Deloitte Base case	Goldman Sachs Base case
 Data centers	945 TWh	750 TWh	1050 TWh
 Intelligence artificielle	300 TWh	225 TWh	210 TWh

Tableau 1 – Projections de la consommation des data centers et de l'IA d'ici 2030 au sein d'un même scénario (sélection)

La part de l'IA dans la consommation des data centers se situe autour du même ordre de grandeur (20–30 %), malgré des estimations de consommation globale des data centers plus dispersées. Les 3 scénarios utilisant une approche bottom-up, ces incertitudes peuvent s'expliquer par des différences d'hypothèses concernant l'évolution du stock de serveurs et de leur taux d'utilisation, ainsi que l'évolution des caractéristiques des data centers (PUE, capacités de stockage et transfert d'information). De plus, dans le cas de l'IEA (pas de détail sur les autres scénarios), la consommation des serveurs accélérés sert à approcher celle de l'IA, ce qui ne reflète pas la diversité des usages, ni l'utilisation auxiliaire des infrastructures de support.

De plus, **les trajectoires d'évolution varient selon les régions géographiques**. Les Etats-Unis, la Chine et l'Europe vont rester dominants en termes de consommation des data centers (85 % de la demande mondiale en 2030). D'après la figure 14, les estimations prévoient que les data centers aux Etats-Unis vont connaître une augmentation de consommation de +274 TWh d'ici 2028 (+155 % par rapport à 2023) (Shehabi et al., 2024). D'ici 2030, ce serait +193 TWh (+230 % par rapport à 2023) en Chine (IEA, 2025) et +95 TWh (+172 % par rapport à 2022) au sein de l'Union européenne (European Commission. Joint Research Centre., 2024). Ainsi, si les Etats-Unis continueront à être les premiers en termes de consommation absolue de data centers, la Chine connaîtra une plus forte croissance relative.

La part des data centers dans la consommation électrique de ces régions augmente également, ce qui posera des défis de gestion et planification des infrastructures dans un contexte d'électrification global des usages et de l'industrie.

En effet, les data centers pourraient représenter jusqu'à 12 % de la consommation électrique des Etats-Unis d'ici la fin de la décennie¹² et pourraient doubler leur part dans l'Union Européenne, y compris en France¹³.

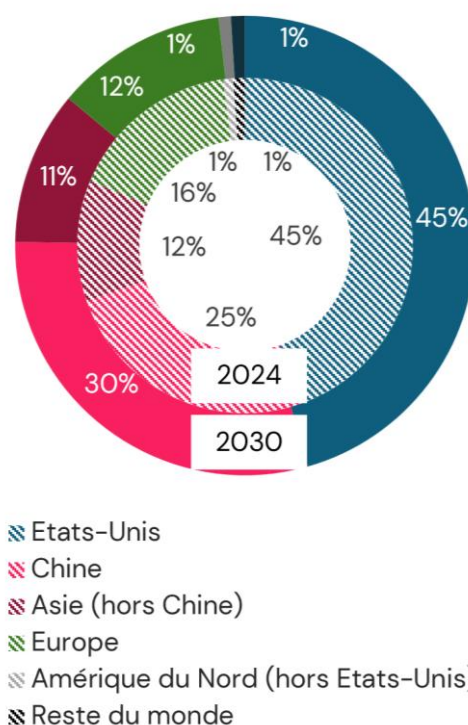


Figure 12 – Répartition de la consommation des data centers par régions en 2024 (hachuré) et 2030 (plein) (IEA, 2025)

¹²La consommation des data centers aux Etats-Unis en 2028 est estimée entre 325 et 580 TWh, représentant entre 6,7 % et 12 % de la consommation électrique future du pays.

¹³ La consommation des data centers en France est estimée à 23–28 TWh en 2030 contre 10 TWh en 2022.

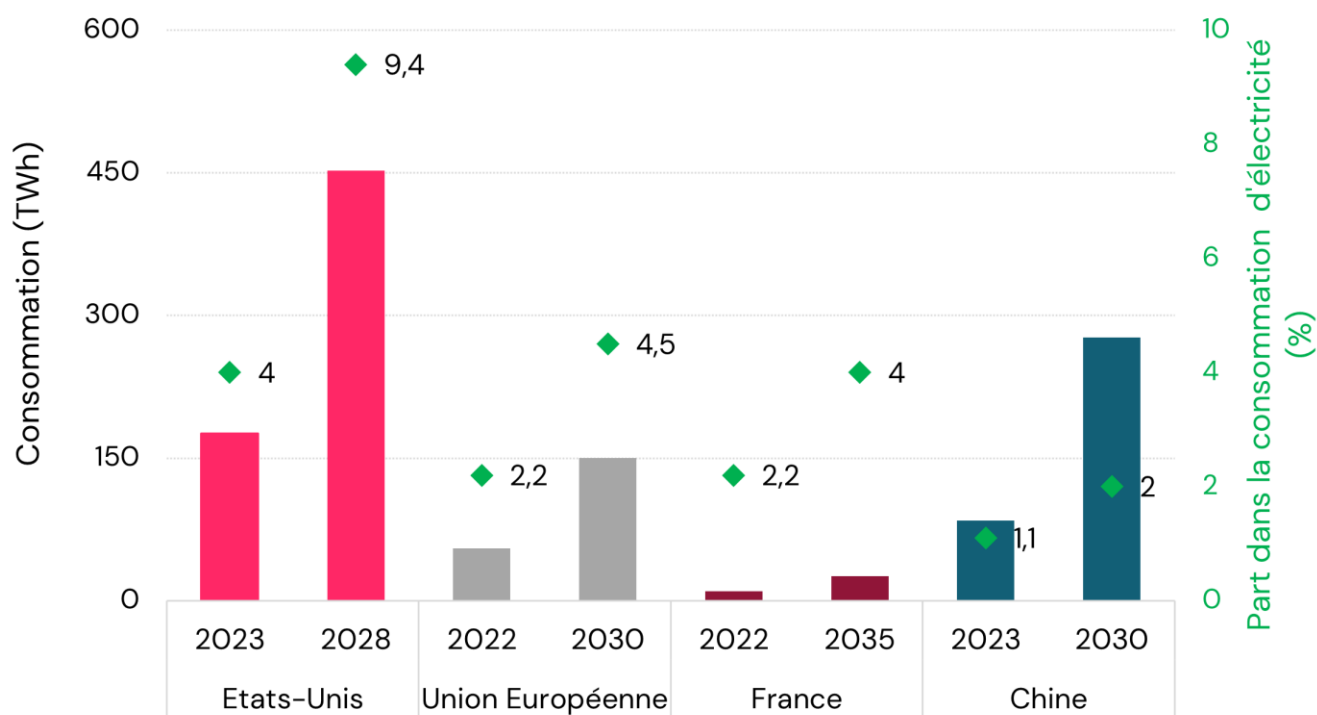


Figure 13 – Évolutions de la consommation des data centers et leur part dans la consommation électrique régionale¹⁴ (European Commission, Joint Research Centre, 2024 ; GlobalData, 2025 ; IEA, 2025 ; RTE, 2025b ; Shehabi et al., 2024 ; McKinsey et al., 2024)

L'absence de consensus sur le niveau d'augmentation de la consommation énergétique de l'IA révèle l'incertitude autour des facteurs qui influencent cette trajectoire :

Méthodes de calcul

La majorité des études s'appuie sur une approche *bottom-up* consistant à estimer la consommation unitaire des équipements puis à multiplier par la base de serveurs installés et un PUE¹⁵ moyen pour arriver à une estimation pour les data centers dans leur globalité. **Cette méthode dépend fortement de la qualité et de la disponibilité des données de base.** Le facteur PUE permet de passer d'une consommation purement informatique à une estimation de la consommation électrique globale prenant en compte des besoins comme le refroidissement des infrastructures. Néanmoins, il n'est pas toujours précisé si les calculs différencient les types de data centers (les data centers hyperscale ayant généralement un PUE plus faible que les data centers d'entreprise), ce qui peut fausser les comparaisons.

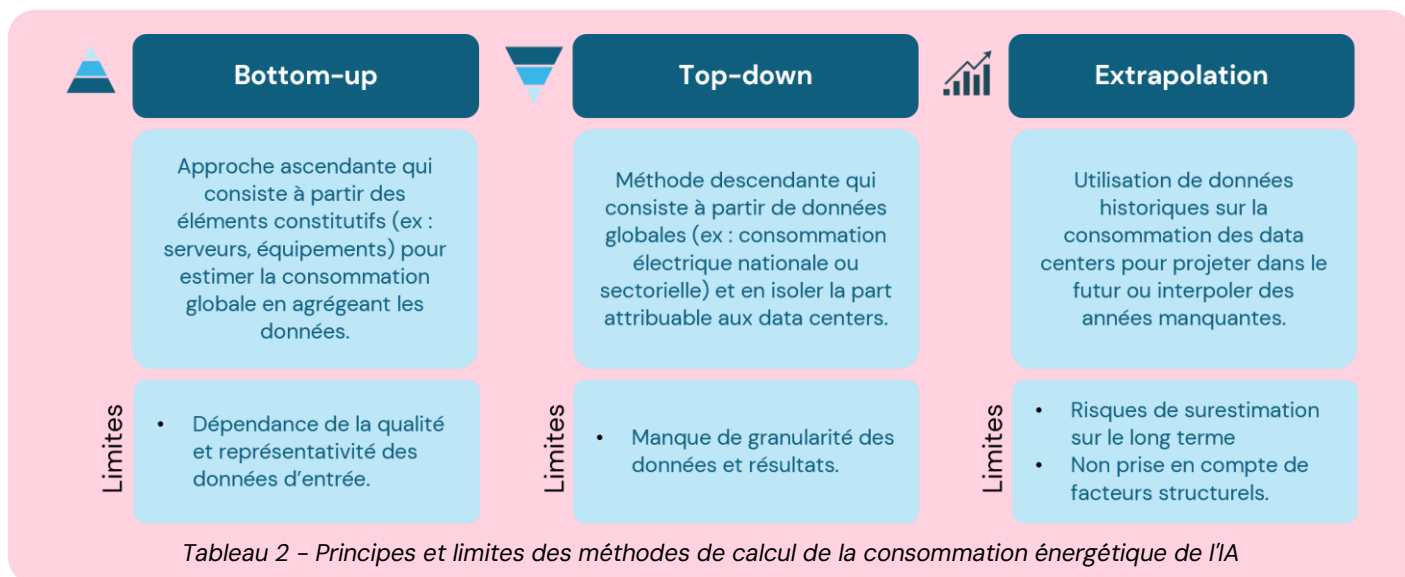
D'autres approches existent, telles qu'une méthode de calcul top-down et l'extrapolation temporelle, mais elles sont moins répandues dans les études faites sur la consommation des data centers et de l'IA (voir Tableau 2).

Facteurs d'évolution de l'IA

Les projections prennent en compte des facteurs quantitatifs tels que l'évolution de la demande en IA, à travers un taux de croissance annuel moyen, et l'évolution de l'efficacité énergétique des infrastructures (amélioration des performances des GPU). Certains scénarios, tels que ceux de Schneider Electric (Paccou et Wijnhoven, 2024), intègrent également des facteurs contextuels tels que le déploiement d'une régulation de l'IA ou bien de possibles tensions géopolitiques entourant la disponibilité des matériaux nécessaires pour fabriquer les processeurs, soulignant l'aspect systémique de l'évolution de l'IA. Plusieurs paramètres affectent ainsi la trajectoire de l'IA et sa demande énergétique.

¹⁴ Dans les cas où seules des fourchettes d'estimation (min-max) sont disponibles, la valeur moyenne de l'intervalle est utilisée pour simplifier la représentation graphique.

¹⁵ Le PUE d'un data center correspond au ratio de sa consommation énergétique totale par la consommation des équipements informatiques seuls.



Taille et complexité des modèles

Les modèles IA déployés sont de plus en plus grands, car les performances d'un modèle s'améliorent (jusqu'à un certain seuil) lorsque l'on augmente sa taille, la quantité de données d'entraînement, ou les ressources de calcul utilisées (loi d'échelle ou *scaling law*). Ainsi, 65 % des améliorations de performance des modèles de langage depuis 2014 peuvent être attribuées aux augmentations d'échelle, contre 35 % attribuables aux innovations algorithmiques (Epoch AI, 2025b). La compétitivité du secteur de l'IA pousse également à grossir les modèles pour pouvoir se distinguer. Le nombre d'opérations nécessaires (FLOP) pour entraîner les modèles notables¹⁶ de machine learning est multiplié par 4,8 chaque année depuis 2010, le nombre de paramètres par 2 et le volume de données d'entraînement jusqu'à 3,7 selon le domaine d'application (Epoch AI, 2024).

La taille des modèles, dont l'amplification contribue à l'augmentation de la consommation à l'entraînement de l'IA, va vraisemblablement stagner ou ralentir progressivement. Les raisons sont majoritairement :

- **Economiques** : Le coût d'entraînement des grands modèles, dont la moitié est consacrée aux équipements informatiques, est multiplié par 2,8 chaque année depuis 2016 (Cottier et al., 2024).
- **Technologiques et algorithmiques** : Le gain de performance des modèles devient marginal à mesure que leur échelle grossit.

Des stratégies portant sur les techniques d'entraînement et les architectures algorithmiques permettent à la fois de réduire la taille des modèles, de garantir de bonnes performances et de diminuer leur consommation énergétique.

Il existe en effet de nouvelles stratégies¹⁷ pour diminuer la durée, les coûts et la consommation énergétique à l'entraînement d'un modèle :

- La **distillation** est une technique qui consiste à transférer les connaissances d'un grand modèle (*teacher*) vers un modèle plus petit (*student*) tout en essayant de conserver le plus de performance possible. Les prédictions et résultats en sortie du grand modèle sont utilisés pour entraîner le modèle élève.
- Le **fine-tuning** consiste à partir d'un modèle pré-entraîné (généralement sur un très grand corpus de données générales) et à le réentraîner partiellement sur un ensemble de données spécifiques à une tâche ou à un domaine, dans le but d'adapter un modèle généraliste à une tâche spécifique.
- La **quantization** vise à réduire la précision des nombres flottants utilisés dans les paramètres des modèles. La quantization conduit également à une baisse du temps d'inférence des modèles, mais peut avoir des incidences sur la performance de ces derniers.

¹⁶ Un modèle notable répond aux critères suivants : 1) il représente une avancée de pointe, 2) il est largement cité (>1000 citations), 3) il a une importance historique, 4) il est largement utilisé.

¹⁷ Des entreprises françaises s'emparent de ces stratégies et s'insèrent dans l'écosystème IA en France, telles que [Pruna AI](#) (optimisation et compression de modèles existants pour améliorer l'efficacité et la durabilité des modèles). [Mistral AI](#) vise également la conception de modèles plus efficaces et durables.

- Le **pruning** consiste à retirer des parties de modèles (neurones ou connexions entre neurones) pour l'alléger.

L'apparition de modèles, non-américains et performants, comme DeepSeek R1, qui s'appuie sur une architecture algorithmique permettant une meilleure efficacité énergétique à l'entraînement et pendant l'inférence, remet en question les stratégies algorithmiques d'un domaine jusqu'ici dominé par quelques géants de l'IA à très forte intensité énergétique.

Demande en IA

La demande croissante en IA accentue la consommation énergétique de la phase d'inférence des modèles et constitue l'un des principaux moteurs de l'augmentation des besoins énergétiques de l'IA. En particulier, **l'adoption de l'IA à grande échelle par les acteurs industriels et entités étatiques laisse entrevoir une généralisation des cas d'usage**. Une enquête de McKinsey souligne que 92 % des cadres interrogés prévoient d'augmenter les dépenses en IA sur les trois prochaines années (McKinsey, 2025). De plus, les plateformes numériques grand public intègrent l'IA dans leurs services de base, ce qui entraîne une invisibilisation de la demande¹⁸ mais une explosion de son volume.

Cette dynamique risque fortement de suivre une trajectoire ascendante. Cette dernière peut être rapide (omniprésence de l'IA dans les activités économiques), ou plus modérée (freins économiques ou sociétaux), voire connaître une spécialisation par secteur, avec un usage intensif dans certains domaines (R&D, automatisation de processus) et marginal dans d'autres.

Box 1 – DeepSeek annonce-t-il la fin des besoins énergétiques démesurés de l'IA ?

DeepSeek R1 est un grand modèle de langage développé par l'entreprise chinoise DeepSeek et rendu public en janvier 2025. Les performances de DeepSeek R1 sont comparables à celles de OpenAI o1 (décembre 2024) (DeepSeek-AI et al., 2025). Les deux modèles sont des modèles de raisonnement, conçus pour raisonner de manière structurée en plusieurs étapes (*chain-of-thought*). Construire et entraîner ce modèle de 670 milliards de paramètres, n'aurait coûté que 6 M\$. A titre de comparaison, GPT-4 possède 1 760 milliards de paramètres et a nécessité environ 41 Md\$₂₀₂₃ pour son entraînement (Epoch AI, 2025a).

Ces faibles coûts d'entraînement et consommations énergétiques sont dus à l'architecture algorithmique et la stratégie d'entraînement de DeepSeek R1 :

- Architecture Mixture of Experts (MoE) : Il s'agit d'un modèle constitué de multiples sous modèles spécialisés. Cela permet d'éviter d'avoir à solliciter un grand modèle généraliste lors de l'entraînement et de l'inférence : seuls 37 milliards de paramètres sont activés par token, c'est-à-dire pour traiter un morceau de texte.
- Multi-head Latent Attention (MLA) : Pour traiter un texte ou une requête, le modèle résume l'information importante sans stocker tous les détails, ce qui permet de ne pas surcharger la mémoire du GPU.
- DeepSeek R1 est issu d'un apprentissage par renforcement, qui repose sur le principe d'essais et erreurs : le système apprend par expérience en fonction de récompenses et pénalités qu'il reçoit.

Néanmoins, les avancées que représente DeepSeek ne signifient pas que la problématique de l'explosion de la consommation de l'IA est résolue. En effet, le processus de réponse à une requête reposant sur le principe du chain-of-thought, les étapes de réflexion supplémentaires à chaque requête génèrent davantage d'unités de texte et peuvent donc engendrer une consommation énergétique importante.

¹⁸ Cette demande en apparence se rapproche plutôt d'une offre inévitable pour l'utilisateur (résumés et traductions de contenus par IA, algorithmes de recommandations personnalisés, etc.).

Améliorations technologiques des équipements informatiques

Historiquement, l'efficacité énergétique (FLOP par seconde par watt) des GPU a doublé tous les deux ans. Or, depuis 2020, la puissance requise pour entraîner les modèles IA de pointe double de façon bien plus rapide tous les 10 mois (Epoch AI, 2025b). **Ainsi, malgré les améliorations significatives de l'efficacité énergétique des GPU, la croissance rapide de la demande énergétique liée à l'IA devrait entraîner une augmentation nette de la consommation énergétique.** La seule efficacité énergétique des processeurs ne permettrait pas de stabiliser les besoins énergétiques, d'autant plus que des effets rebond pourraient survenir et contribuer à augmenter les usages.

De plus, même si les capacités de calcul et d'efficacité énergétique des GPU s'améliorent, d'autres considérations technologiques, telles que les capacités de mémoire des GPU, sont à prendre en compte. Ces dernières progressent lentement : le nombre de calculs par seconde (FLOPS) que les GPU peuvent effectuer double environ tous les 2,4 ans, tandis que la capacité de mémoire (Byte) double tous les 4 ans (Epoch AI, 2025b). La mémoire d'un GPU sert à stocker localement les données du modèle pendant l'entraînement ou l'inférence, pour qu'elles soient rapidement accessibles pendant les calculs. Or, si cette mémoire est insuffisante, le GPU doit aller chercher les données ailleurs, ce qui ralentit fortement les calculs, et peut devenir, à terme, un goulot

d'étranglement dans l'amélioration des performances informatiques des équipements.

Approvisionnement électrique

Le déploiement des projets d'infrastructures énergétiques fait face aujourd'hui à des problèmes physiques et logistiques. Aux Etats-Unis, le temps de raccordement des installations de production d'énergie est passé de <2 ans pour les projets construits entre 2000-2007 à 4 ans pour ceux construits entre 2018-2023; environ 2 600 GW de projets de production et stockage d'énergie sont en attente d'instruction, 95 % de ces projets étant pour de l'éolien, du solaire ou du stockage. (Lawrence Berkeley National Laboratory, 2024). Dans le contexte climatique actuel, les data centers devraient être alimentés en électricité décarbonée, or les projets ENR ne sont pas opérationnels. **Le gaz naturel pouvant être utilisé pour produire de l'électricité, il est souvent considéré comme une énergie de transition pour répondre sur le court terme à la demande des data centers face à la variabilité des ENR.** Cependant, des goulots d'étranglement apparaissent dans les chaînes d'approvisionnement : des pénuries de turbine à gaz aux Etats-Unis ralentissent la construction de nouvelles centrales à gaz sur plusieurs années. **Sur le court terme, des capacités de production d'électricité insuffisantes pourraient donc limiter la construction de data centers plus énergivores.**

Box 2 – Evolution du PUE des data centers

Les data centers les plus efficaces atteignent aujourd'hui des PUE proches de 1,2, et en moyenne 1,56 dans le monde, contre environ 2 en 2011. Ce sont principalement les data centers hyperscale et ceux consacrés à l'IA qui tirent la moyenne mondiale vers le bas. Cependant, les gains énergétiques se stabilisent depuis plusieurs années car les marges de progrès deviennent plus limitées. De plus, une baisse du PUE traduit une meilleure efficacité du data center en tant que bâtiment, mais elle ne reflète pas une baisse de la consommation des serveurs IA. Ainsi, **même avec un PUE proche de 1, les besoins en électricité peuvent continuer à croître fortement si la charge de calcul IA augmente ou si les équipements informatiques utilisés sont énergivores**, ce qui est aujourd'hui le cas.

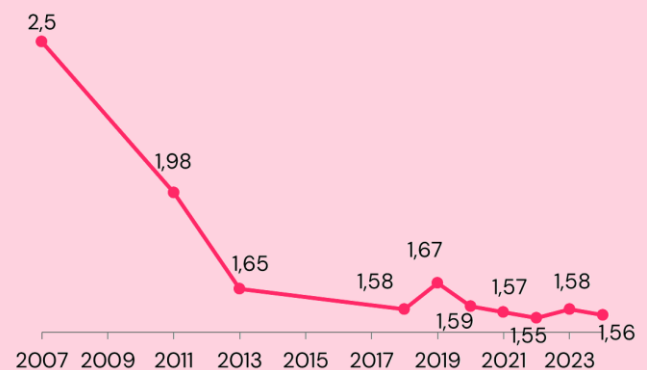


Figure 14 – Évolution du PUE moyen des data centers dans le monde (Statista, 2025b)

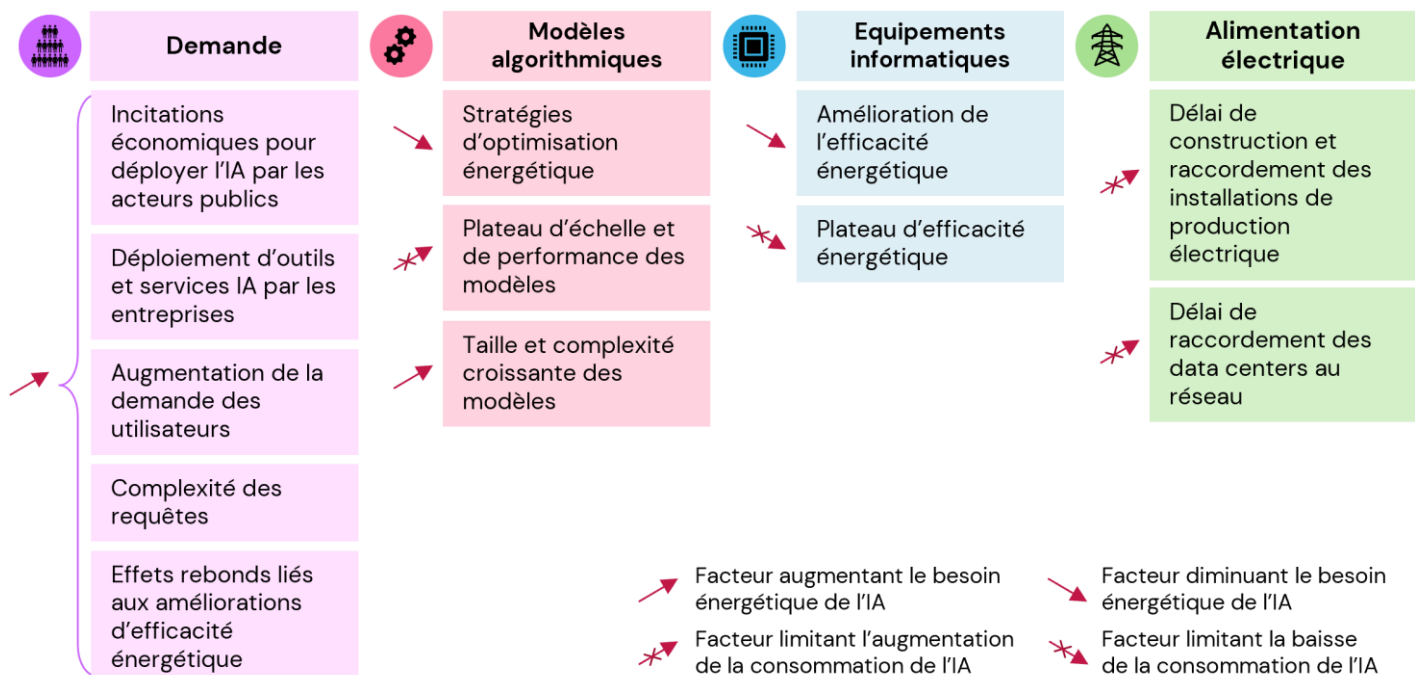


Figure 15 – Synthèse des différents facteurs inflationnistes et déflationnistes de la consommation énergétique de l'IA

Les conséquences d'une forte demande énergétique de l'IA

Mix énergétique des data centers et ambitions climatiques

Le besoin en énergie de l'IA et des data centers nécessite des capacités additionnelles : la puissance appelée par les data centers est amenée à doubler d'ici 2030. D'après l'AIE, les énergies renouvelables, tirées par les investissements des entreprises du numérique, pourraient répondre à 50 % de la demande additionnelle des data centers en électricité d'ici 2030, et le gaz naturel et le charbon devraient couvrir environ 40 % de la demande électrique supplémentaire. La Figure 16 montre que la croissance de la production électrique pour les data centers serait particulièrement forte entre 2025 et 2030. C'est la production solaire photovoltaïque qui contribue le plus à cette croissance, suivie par le gaz naturel, dont l'essor explique la poursuite de l'augmentation des émissions liées aux data centers. Ces dernières seraient probablement plus élevées encore dans un scénario dans lequel le gaz naturel occupe une part plus importante dans la réponse à la demande d'ici 2030.

Tant que leur consommation d'électricité reposera en partie significative sur des sources fossiles comme le charbon ou le gaz, les émissions des data centers continueront d'augmenter. Or, aux Etats-Unis, la croissance de la demande électrique des data centers est encore satisfaite en partie par le charbon et le gaz naturel, ce qui compromet les efforts globaux de décarbonation.

Les data centers sont de plus en plus alimentés par des énergies renouvelables, mais leur approvisionnement dépendra encore des énergies fossiles d'ici 2030, ce qui empêche une réduction drastique des émissions. Les SMR¹⁹ pourraient constituer une solution pour alimenter les data centers en électricité décarbonée, grâce à leur modularité et leur flexibilité d'installation, mais la majorité des projets en cours ne seront fonctionnels qu'après 2030 au mieux.

Cette composition du mix énergétique associée à une augmentation de la demande des data centers entraîne une hausse de leurs émissions de CO₂ : elles atteignent un pic de 320 MtCO₂e en 2030, soit x1,7 par rapport à 2024, puis subissent une légère baisse à 300 MtCO₂e en 2035. Cette diminution coïncide avec l'augmentation de la part du nucléaire dans le mix énergétique des data centers, elle-même liée à l'entrée en service progressive des SMR.

¹⁹ Small Modular Reactors ou petits réacteurs modulaires, conçus en modules standardisés, plus compacts et de plus faible puissance que les réacteurs traditionnels. Les SMR pourraient être fabriqués en série pour une installation plus flexible et rapide près des sites industriels.

Pour que la croissance des usages numériques soit compatible avec les objectifs climatiques, il est nécessaire de décarboner rapidement l'approvisionnement en électricité :

- En **augmentant la part d'énergies renouvelables dans le mix électrique des data center** : il y a un risque de concurrence avec l'électrification d'autres secteurs (transports, industrie). Il faut également développer les technologies de stockage ou de flexibilité pour pallier la variabilité des ENR.
- En déployant des **technologies de captation de carbone (CCS) pour les centrales électriques** dont la source d'énergie primaire est fossile : ces technologies sont coûteuses et peu répandues.

D'autres estimations telles que celles de Deloitte tiennent compte des émissions indirectes liées au cycle de vie complet de l'électricité utilisée, incluant non seulement les émissions directes de combustion, mais aussi celles liées à l'infrastructure, à la chaîne d'approvisionnement et aux fuites de méthane. Ainsi, dans un scénario de forte adoption de l'IA, les émissions des data centers atteindraient 235 MtCO₂e en 2030 (Lorentz et al., 2024), soit une différence de 26 % avec la projection de l'AIE. La demande énergétique globale des data centers estimée par

les deux approches est sensiblement la même (970 TWh en 2030 dans un scénario de forte adoption par Deloitte, 945 TWh d'après l'AIE), et les deux méthodes de comptabilisation carbone étant location-based²⁰, cette différence pourrait s'expliquer par un chiffrage plus optimiste par Deloitte de l'intensité carbone de l'électricité, ou par la prise en compte du scope 3 des émissions par l'AIE.

Stratégies d'approvisionnement énergétique

Pour diminuer les émissions indirectes des data centers et atteindre leurs objectifs climatiques tout en satisfaisant l'augmentation de la demande énergétique de ces infrastructures, les entreprises du numériques, et particulièrement les GAFAM, misent sur plusieurs stratégies combinées.

Recours aux énergies renouvelables

Les entreprises de la tech peuvent revendiquer avoir recours aux énergies renouvelables pour alimenter leurs data centers et activités opérationnelles :

- En construisant ou en exploitant leur propre parc ENR,
- En passant par des contrats d'achat d'énergie renouvelable (PPA) ou des certificats de provenance.

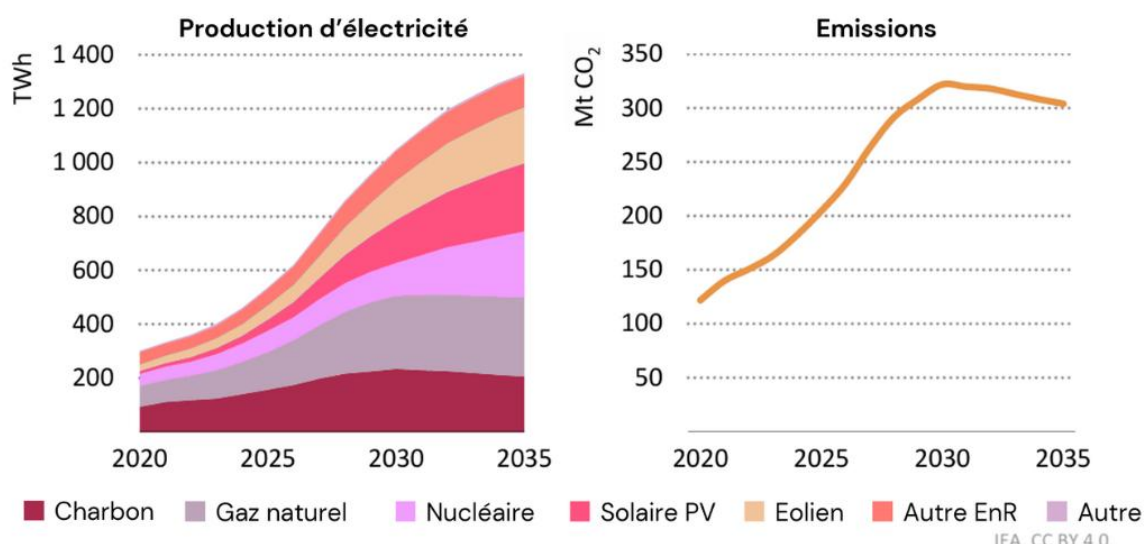


Figure 16 – Evolution du mix énergétique alimentant les data centers et les émissions associées (adapté de IEA, 2025, scénario de base)

²⁰ La méthode location-based de comptabilité carbone évalue les émissions liées à la consommation d'électricité des data centers en utilisant l'intensité carbone moyenne du réseau électrique local ou national sur lequel ils opèrent. Elle reflète donc l'impact réel du mix énergétique régional, indépendamment des contrats ou achats d'énergie verte effectués par l'entreprise. Contrairement à la méthode market-based, elle ne tient pas compte des mécanismes de compensation comme les certificats d'énergie renouvelable.

Cette dernière méthode permet de soutenir la production d'électricité décarbonée mais ne garantit pas une correspondance physique et temporelle entre la production et la consommation d'électricité. Un PPA est en effet un contrat à long terme entre un producteur d'énergie (souvent renouvelable) et un acheteur, par lequel ce dernier s'engage à acheter une certaine quantité d'électricité à un prix fixé, sans forcément consommer directement cette énergie sur site.

C'est une manière pour les entreprises de garantir un approvisionnement vert et de soutenir financièrement la production renouvelable, ainsi que de se protéger des fluctuations de prix. Les GAFAM sont leaders dans l'achat d'énergie renouvelable et des crédits carbone : Google, Amazon, Microsoft et Meta ont représenté 29 % des contrats d'achats d'éolien et de solaire en termes de puissance en 2023 (Catsaros, 2024). Néanmoins, l'électricité achetée via un PPA est injectée dans le réseau public, sans garantie que le data center consomme effectivement cette électricité renouvelable en temps réel. De plus, les installations d'énergie renouvelable sont de plus en plus nombreuses à subir des délais rallongés de raccordement au réseau local, pouvant compromettre les engagements climatiques des entreprises du numérique et leur stratégie énergétique.

Recours à la production d'électricité par le gaz

Le gaz apparaît comme une solution à court terme pour répondre à la montée en puissance globale des data centers et de l'IA, grâce à la flexibilité et la pilotabilité des centrales à gaz.

Aux Etats-Unis, le gaz constitue une énergie abondante et peu coûteuse et représente 40 % du mix électrique du pays (LowCarbonPower, 2025). Goldman Sachs estime que le gaz pourra répondre à 60 % de la demande additionnelle des data centers aux Etats-Unis (Goldman Sachs, 2024), correspondant à une capacité additionnelle de 28 GW. Néanmoins, les nouvelles politiques énergétiques de Trump exhortant à l'exportation du gaz risquent de diminuer les réserves disponibles pour un marché interne et d'engendrer des conflits d'usage avec l'électrification d'autres secteurs. Les tensions présentes dans les chaînes d'approvisionnement logistique ralentissent également la construction de nouvelles centrales.

En Europe, le recours au gaz dépend du mix électrique des pays et de leurs stratégies climatiques. Au même titre qu'aux Etats-Unis, le gaz reste considéré comme une énergie de secours et flexible pour alimenter les data centers. En Irlande, fin 2024, une poignée de data centers en Irlande consomment effectivement du gaz (moins de 5, pour ~1,8 GW), mais la file de projets en attente est considérable : 22 projets supplémentaires (9,3 GW) ont déposé des demandes (Daly, 2024). Les data centers en Allemagne manifestent également un intérêt croissant envers les centrales à gaz, voire à charbon.

En France, à court-terme (2030-2035), le nucléaire et les énergies renouvelables devraient constituer la majorité du mix électrique national (RTE, 2022) : les data centers pourront être alimentés sans recourir massivement au gaz. Néanmoins, sur la prochaine décennie, une augmentation de la demande électrique globale, une faible maturité de solutions de stockage pour pallier la variabilité des ENR, et l'arrêt de certaines centrales nucléaires ainsi que le délai de mise en service de nouveaux réacteurs et de SMR peuvent provoquer le recours au gaz.

Meta a entamé en 2025 la construction d'un nouveau data center en Louisiane, le plus grand qu'il a jamais construit. Ce dernier nécessitera plus de 2 GW de capacité électrique, et l'entreprise locale de production d'électricité Entergy envisage de construire 3 nouvelles centrales à gaz pour répondre à sa demande énergétique (Palmer, 2025).

Exemple de stratégie énergétique des géants du numérique

Recours au nucléaire

Les centrales nucléaires fournissent une production électrique stable et bas carbone, adaptée aux data centers car ces derniers fonctionnent sans interruption. **Les GAFAM se tournent en particulier vers les SMR qui pourraient être construits plus rapidement que les centrales nucléaires classiques et implantés près des sites industriels**, bien que les technologies de SMR ne soient pas encore matures – TRL 4-5 aux Etats-Unis (Académie des Sciences, 2022) – et que **les premières mises en service ne soient pas prévues avant 2030.**

Google a noué un partenariat avec l'entreprise de SMR Kairos Power pour bénéficier de plus 500 MW d'électricité bas-carbone d'ici 2035, soit 4 réacteurs de 140 MW. La mise en service d'un premier réacteur est prévue en 2030, avec une montée en régime jusqu'en 2035 (Kairos Power, 2024; Sfen, 2024b).

Microsoft a passé un contrat d'achat d'électricité de 20 ans avec l'entreprise Constellation, afin de bénéficier du redémarrage du réacteur 1 de la centrale nucléaire Three Mile Island, en Pennsylvanie, d'ici 2028. Le réacteur 1 de la Centrale Three Mile Island a été fermé en 2019 pour des raisons économiques (coûts de production trop élevés et faible compétitivité par rapport au gaz de schiste); il possède une capacité de 835 MW (Sfen, 2024a).

Recours à la géothermie

Les systèmes géothermiques conventionnels, ou systèmes hydrothermaux, sont **matures (TRL 9) mais limités par des emplacements géographiques spécifiques**. Par exemple, la plupart des ressources géothermiques aux Etats-Unis se trouvent dans l'Ouest du pays ; le pays disposerait de 40 GW potentiels de géothermie conventionnelle et en exploite actuellement 3,7 GW (Sweeney & Gordon, 2025). Les data centers pourraient s'adapter à cette contrainte géographique : la phase d'entraînement d'un modèle d'IA ne nécessite pas que les data centers soient proches des utilisateurs finaux, donc les centres de données pourraient être implantés près des sites de géothermie. La plupart des data centers hyperscale se trouvent aujourd'hui en Virginie, à l'Est du pays, or la chaleur géothermique est locale. L'électricité produite par géothermie peut être transportée mais subira des pertes, donc elle est surtout injectée dans les réseaux électriques locaux.

Google s'est associé à la startup Fervo en 2023 pour développer un projet de géothermie dans le Nevada apportant de l'électricité décarbonée sur le réseau local, dans le cadre de son objectif d'utiliser une énergie sans carbone en continu (24/7) pour ses opérations. Google a également signé en 2025 son premier PPA géothermique en Asie, à Taïwan, avec l'entreprise Baseload Capital. Le PPA porte sur 10 MW de puissance géothermique ajoutée au réseau d'ici 2029 (Google, 2025).

Production sur site

Décarboner l'électricité qui approvisionne les data centers n'est pas la priorité principale : il s'agit avant tout d'être capable de les alimenter de façon continue et suffisante. Or avec l'allongement des délais de raccordement des installations aux réseaux électriques, une alternative est de **connecter directement les centrales de production d'énergie aux data centers, sans passer par le réseau public (approvisionnement behind the meter)**. Cette stratégie est encore marginale chez les entreprises du cloud, qui ont principalement recours aux PPAs.

Augmenter la flexibilité des data centers

Améliorer la flexibilité des data centers est une stratégie clé pour faciliter l'intégration des énergies renouvelables et limiter la pression sur le réseau électrique. Les data centers gagneraient à devenir acteurs du réseau électrique et à ajuster leur activité en fonction de la disponibilité d'électricité bas carbone. Cela consiste par exemple à **réduire leur consommation lors des pics de consommation (mécanisme d'effacement)** ou à **décaler des charges non critiques comme l'entraînement de modèles**, ou encore à mobiliser des batteries. Aux Etats-Unis, le réseau électrique pourrait être capable d'absorber 100 GW de charge flexible sans nécessiter d'investissements majeurs dans l'expansion du réseau ou dans la production d'électricité, en jouant sur la flexibilité de certaines charges électriques, dont celles des data centers, qui seraient décalées vers des périodes où la demande globale est plus faible (Norris et al., 2025).

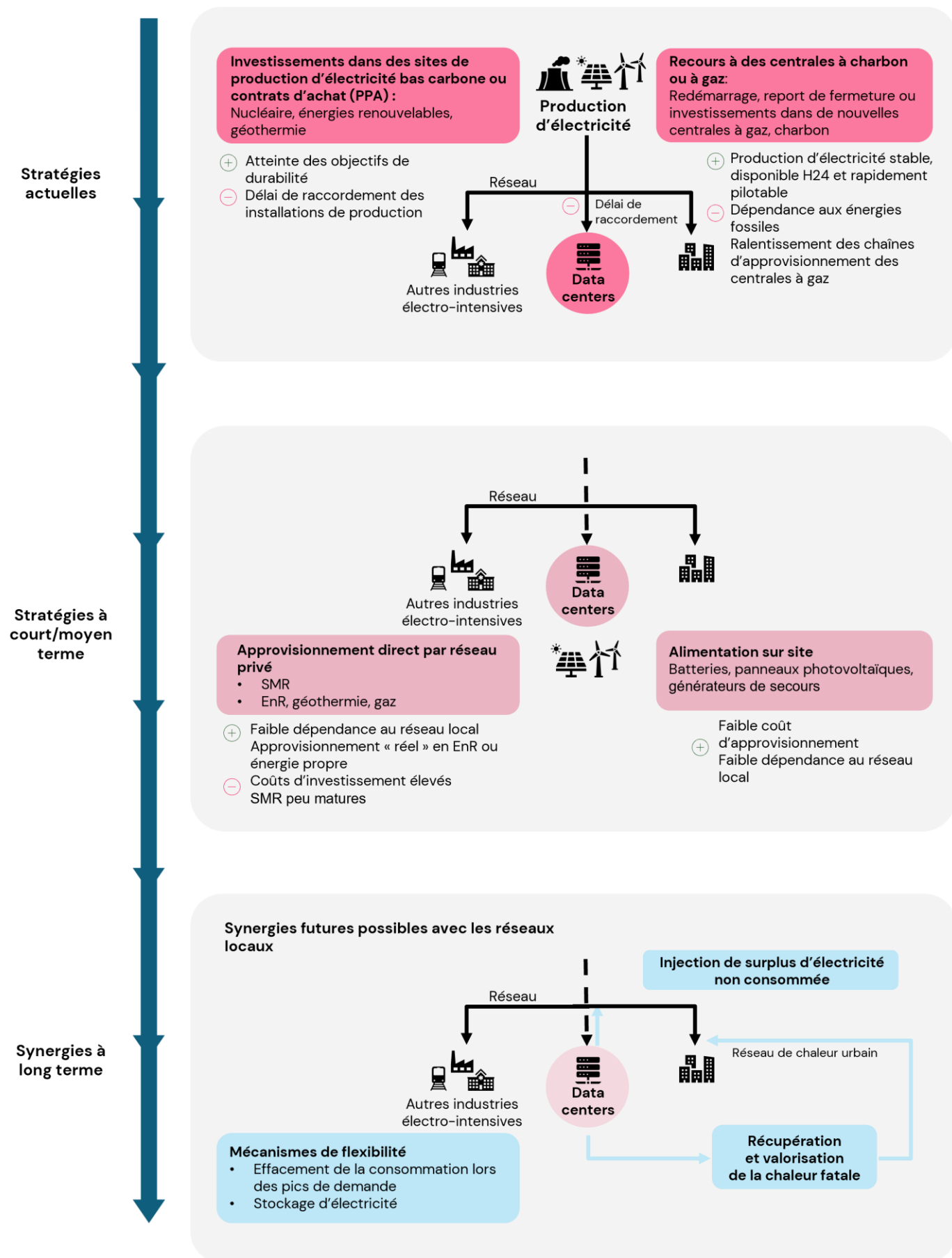


Figure 17 – Stratégies actuelles et futures d’approvisionnement énergétique et synergies possibles avec les réseaux locaux

Contrairement à d'autres usages électriques, la demande des data centers est fortement concentrée spatialement (les data centers hyperscale planifiés ont souvent une puissance de plusieurs centaines de MW, voire supérieure à 1 GW). Cette concentration crée des contraintes majeures sur les réseaux locaux :

- **Risques (déjà existants) de congestion et surcharge des réseaux** nécessitant l'ajout de nouvelles capacités de raccordement et de transport.
- **Besoin d'anticipation** de la part des gestionnaires de réseaux pour pouvoir absorber la demande. Les délais de raccordement de data centers aux réseaux sont notamment plus longs que leur temps de construction.
- **Risques de concurrence** avec d'autres besoins d'électrification sur le même territoire.

Ainsi, l'implantation des data centers dépend autant de la disponibilité d'une production électrique abondante et décarbonée que de la capacité des réseaux locaux de transport et de distribution à absorber une charge très concentrée.

Si la demande énergétique des data centers suscite des préoccupations croissantes, notamment en matière de capacité réseau, d'émissions et de concurrence avec d'autres usages électriques, les prévisions actuelles envisagent des dynamiques de croissance différentes selon les régions du monde.

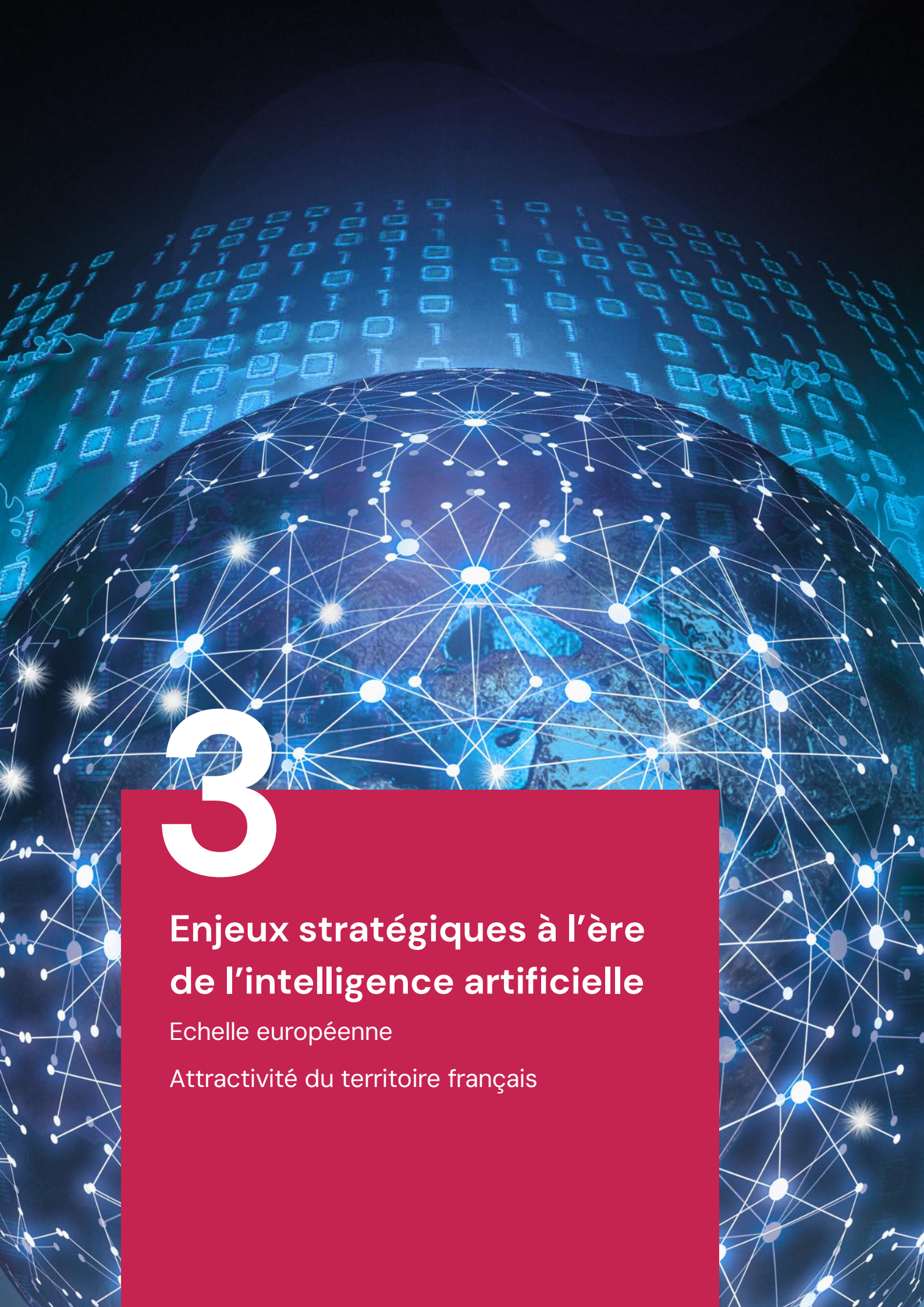


Figure 18– Évolutions des capacités installées (GW) de data centers en 2024 (cercles pleins) et en 2030 (cercles pointillés). (IEA, 2025)

Les puissances appelées par les data centers sont amenées à être multipliées par 2,3 à l'échelle mondiale, par 2,8 en Chine et par 2,4 aux États-Unis, et seulement par 1,6 à l'échelle européenne. Il est important de prendre en considération le fait que **l'évolution de la consommation énergétique des data centers et de l'IA, ainsi que leurs conséquences, ne se produira pas au même rythme selon les régions.**

En effet, en ce qui concerne l'impact sur les réseaux locaux et les infrastructures de production, les États-Unis sont en quelque sorte en avance de phase, car ils concentrent déjà le plus grand nombre de data centers. C'est la raison pour laquelle on observe actuellement une forte mobilisation des GAFAM pour sécuriser leur approvisionnement énergétique. En Europe, des tensions commencent à apparaître (notamment en Irlande), mais le phénomène reste encore limité à certains territoires, car le déploiement des data centers y est plus progressif. La Chine, de son côté, investit massivement dans de nouvelles capacités : la consommation de ses data centers pourrait donc croître très rapidement, toutefois, les données disponibles sont partielles et incertaines. Il est donc difficile d'évaluer précisément l'ampleur de ce développement.

Les émissions associées dépendent largement de l'évolution du mix électrique, qui varie selon les régions. L'Europe avance plus vite dans sa transition vers un mix bas carbone que les États-Unis, ce qui atténue l'empreinte carbone des data centers sur son territoire. Les politiques publiques jouent aussi un rôle déterminant. La régulation concernant la construction et l'implantation des data centers est relativement limitée aux États-Unis et en Chine, tandis qu'elle est beaucoup plus contraignante en Europe. Cela contribue à ralentir l'expansion des data centers sur le continent, même si cela peut aussi favoriser une meilleure maîtrise de leurs impacts environnementaux. Or **si l'Europe entend jouer un rôle de premier plan dans l'IA, elle devra définir une stratégie spécifique pour permettre le développement d'une infrastructure numérique compétitive tout en respectant ses objectifs de durabilité.**



3

Enjeux stratégiques à l'ère de l'intelligence artificielle

Echelle européenne

Attractivité du territoire français

Echelle européenne

La souveraineté d'un pays ou d'une région en matière d'IA peut être approchée par sa capacité à maîtriser et orienter les infrastructures, les technologies, les données et les usages liés à l'IA, sans dépendre de puissances ou d'acteurs extérieurs. La présence de data centers sur le territoire considéré peut également contribuer à une souveraineté en matière d'IA, en garantissant un accès local à des capacités de calcul et une soumission au droit régional ou national. Toutefois, une souveraineté technologique pleine nécessite un contrôle des technologies, des modèles IA et des infrastructures critiques (semi-conducteurs, data centers).

Il convient également de noter que le développement de l'IA est aujourd'hui étroitement lié à la puissance de calcul mobilisée, dont l'augmentation se traduit par une forte consommation d'énergie. Ainsi, les projections de besoins énergétiques futurs des data centers donnent l'impression que l'Europe est en perte de vitesse par rapport à la Chine et aux États-Unis, en termes de capacité brute. Néanmoins, plutôt que de viser une croissance brute des capacités, l'Europe pourrait chercher à se différencier par l'efficacité énergétique des infrastructures et la sobriété des modèles.

L'Europe peut se démarquer dans la course à l'IA :

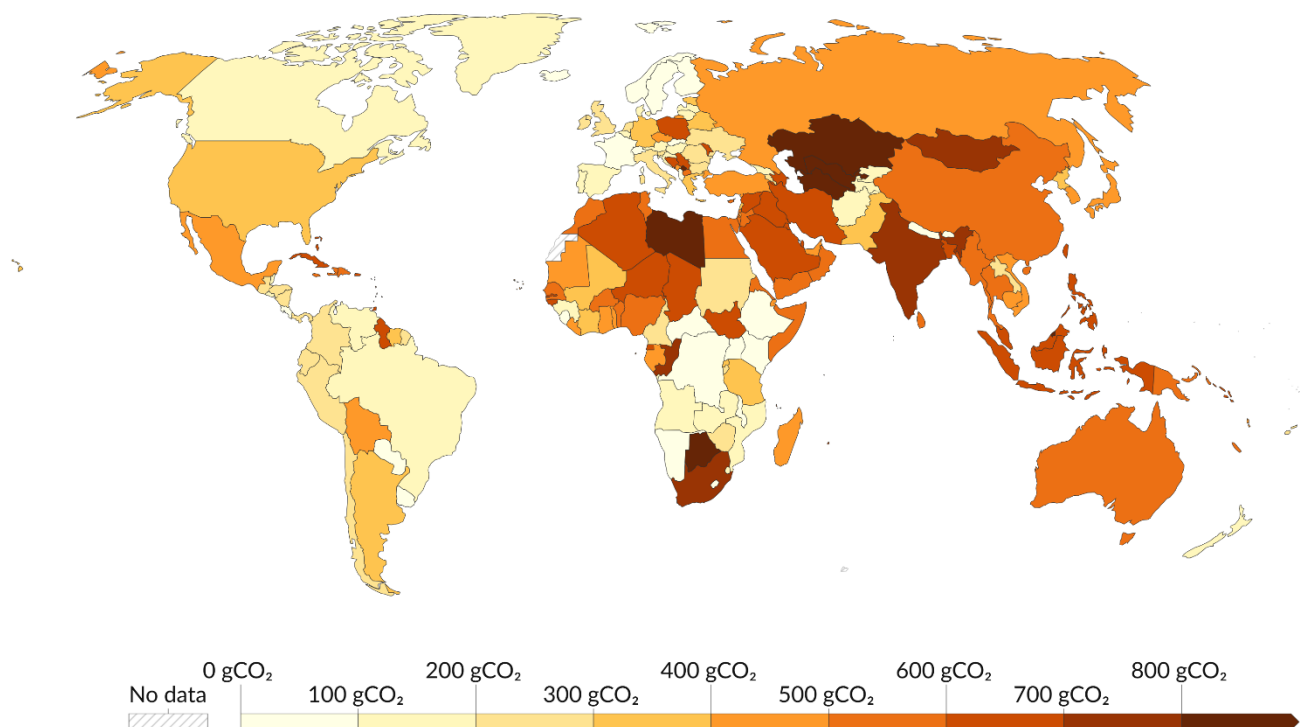
- En attirant la **construction de data centers**, notamment pour l'entraînement d'IA,
- En développant son **propre écosystème** d'acteurs de l'IA,
- En exploitant les **synergies possibles entre les data centers et les réseaux d'électricité et de chaleur locaux**, pour valoriser les ressources énergétiques et accélérer la transition énergétique des collectivités.

Plusieurs facteurs influencent l'implantation de data centers sur le territoire européen :

- Les **tarifs de l'électricité** pour les industries électro-intensives constituent un facteur déterminant pour la rentabilité et l'implantation des data centers. Or ces tarifs de l'électricité aux États-Unis (0,083 USD/kWh) sont environ deux fois moins élevés que le tarif industriel européen moyen (0,150 USD/kWh pour l'UE ¹⁴²¹ et le Royaume-Uni) (Patel et al., 2024).
- Pour atteindre leurs objectifs climatiques, les entreprises du numérique peuvent être incitées à localiser²² leurs data centers dans des **régions où l'électricité est à la fois abordable et décarbonée**. Alors que l'intensité carbone moyenne de l'électricité en 2024 était de 384 gCO₂e/kWh aux États-Unis, elle était de 281 gCO₂e/kWh en Europe et 237 gCO₂e/kWh en Union Européenne (Ember et al., 2025).

²¹ Autriche, Belgique, Danemark, Finlande, France, Allemagne, Grèce, Irlande, Italie, Luxembourg, Pays-Bas, Portugal, Espagne, Suède

²² D'autres facteurs interviennent également dans le choix de localisation des data centers, comme la fiscalité et les incitations économiques, la disponibilité du foncier, la stabilité économique et politique, la confidentialité des données, la présence de main d'œuvre et compétences locales pour l'entretien technique des infrastructures. La non-congestion des réseaux joue également un rôle important.



Data source: Ember (2025); Energy Institute - Statistical Review of World Energy (2025)

OurWorldinData.org/energy | CC BY

Figure 19 – Intensité carbone en gCO₂e/kWh d'électricité produite dans le monde en 2024 (adapté de Ember et al., 2025)

Face aux Etats-Unis et à la Chine, l'Europe ambitionne de devenir un leader mondial de l'IA grâce à sa stratégie *AI Continent Action Plan*

reposant sur 5 axes clés et 200 Md€ d'investissements (dont 50 Md€ de financements publics) :

Infrastructures et capacités de calcul

Création de 13 "AI factories" dans le cadre du programme EuroHPC pour renforcer la puissance de calcul et assurer l'indépendance technologique de l'UE, avec pour objectif de tripler la capacité des data centers d'ici 5 à 7 ans. **Actuellement, l'Europe ne représente que 4 % de la capacité mondiale de calcul pour l'IA**, et les infrastructures ne sont pas conçues pour l'entraînement de grands modèles d'IA (Doan et al., 2025).

Disponibilité de données

Améliorer l'accès à des données fiables et volumineuses avec la création de "laboratoires de données" et d'un marché intérieur de données.

Adoption de l'IA

L'Europe vise le développement d'algorithmes et l'adoption de l'IA dans les secteurs stratégiques, particulièrement dans l'industrie.

Compétences

Renforcer les compétences et les talents grâce à des programmes de bourses pour attirer talents et experts internationaux.

Réglementation

L'IA Act est entré en application début 2025 pour encadrer les usages de l'IA. Il s'agit du "premier règlement complet sur l'IA établi par un organisme de réglementation important" (EU Artificial Intelligence Act, s. d.). L'UE pourrait ainsi **se différencier, dans la course à l'IA, en instaurant un cadre pour une IA éthique et transparente**. En imposant des exigences fortes sur la qualité des données, la transparence, l'explicabilité des modèles, etc., l'IA Act peut pousser à développer des solutions technologiques dans lesquelles l'Europe peut se spécialiser. Cela ne garantit pas un leadership technologique immédiat, mais peut construire un avantage stratégique durable.

Cette stratégie ne tient cependant pas compte des enjeux énergétiques liés à l'IA. Or cette dernière pose indéniablement un dilemme énergétique en Europe, entre mix électrique, coût de l'électricité et robustesse du réseau électrique.

Attractivité du territoire français

La stratégie française pour l'IA est axée sur la **souveraineté technologique grâce à l'attraction de talents et le déploiement d'infrastructures de pointe.**

Face au projet Stargate²³ des Etats-Unis, 109 Md€ d'investissements (principalement privés) dans l'IA ont été annoncés en France par Emmanuel Macron au Sommet de l'IA à Paris en février 2025. La France dispose de plusieurs atouts et met en place plusieurs stratégies pour devenir un acteur majeur de l'IA²⁴.

Mix électrique

Le mix électrique de la France est dominé par le nucléaire (67,1% de la production électrique 2024) et les renouvelables (27,8 %) (RTE, 2025a) et il est largement décarboné, ce qui en fait un atout pour attirer les data centers. EDF a lancé en début d'année un appel à manifestation d'intérêt aux entreprises du numérique prêtes à installer leurs data centers sur des espaces fonciers déjà raccordés au réseau électrique : quatre sites totalisant 2 GW de puissance sont déjà prêts à l'emploi, et deux autres suivront d'ici 2026 (EDF, 2025). **Néanmoins, le raccordement des**

infrastructures au réseau subit un délai d'attente qui se prolonge, à cause de l'augmentation du nombre de demandes de raccordement et de la puissance de raccordement demandée. La Commission de Régulation de l'Énergie (CRE) a approuvé en mai 2025 la procédure "fast track" proposée par RTE dans le but d'accélérer le raccordement, d'ici 2028-2029, des sites de consommation de grande puissance (entre 400 MW et 1 GW) au réseau très haute tension (400 kV). Cela permet d'améliorer l'attractivité du réseau électrique français vis-à-vis des data centers mais aussi des industries électro-intensives (CRE, 2025).

Infrastructures et capacités de calcul

En 2022, près de 300 data centers étaient implantés en France (RTE, 2025b). La puissance installée des data centers était de 683 MW en 2024, avec une forte concentration de la puissance en Île de France (EY & France Data Center, 2025). Une augmentation du nombre d'infrastructures dans ces régions peut mener à des tensions sur les réseaux locaux. **Les data centers gagneraient donc à être davantage répartis sur le territoire, en particulier pour l'entraînement de modèles IA, moins contraints par les délais de communication.** En plus de l'identification de 35 sites prêts à accueillir des centres de données dans plusieurs régions en France, de nombreux investissements pour la construction de data centers sont prévus, dont un campus de 1 GW dédié à la recherche et l'innovation en IA en France.

²³ Initiative annoncée en début de 2025, visant à bâtir les centres de données nécessaires aux Etats-Unis pour porter la prochaine génération d'IA. 500 Md\$ d'investissements privés sont prévus sur 4 ans, grâce au soutien d'[OpenAI](#), [Oracle](#) (spécialiste du cloud) et [SoftBank](#) (géant japonais des investissements).

²⁴ Un acteur majeur de l'IA peut se définir de façon informelle comme une entité exerçant une influence significative sur le développement, déploiement ou la gouvernance de l'IA, à l'échelle internationale.

DES ZONES PRÊTES POUR L'IMPLANTATION
DE CENTRE DE DONNÉES BAS CARBONE POUR
L'IA À TRAVERS LA FRANCE

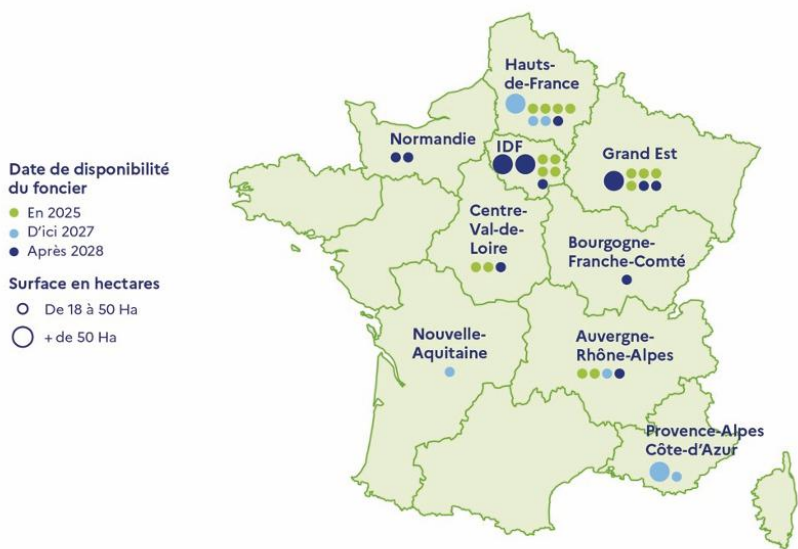


Figure 20 – Carte des régions prêtes à l'implantation de data centers en France (Elysée, 2025)

La France possède également trois supercalculateurs, utilisés pour les tâches nécessitant des grandes puissances de calcul, dont le développement d'IA. Un futur supercalculateur est prévu d'ici 2026, avec une puissance de calcul de 1 Exaflops, soit un milliard de milliard d'opérations flottantes par seconde.

Réglementation

Le projet de loi de simplification de la vie économique vise à faire entrer les projets de construction de gros data centers dans la catégorie "Projet d'Intérêt National Majeur", pour accélérer les procédures administratives (accélérer l'adaptation des règles d'urbanisme local, sécuriser les dérogations à la protection des espèces, garantir un raccordement rapide au réseau électrique, si nécessaire). Le Royaume-Uni adopte une stratégie similaire sous la forme de la création d'AI Growth Zones²⁵.

Supercalculateur	Jean Zay	Adastra	Jules Verne	Alice Recoque ²⁶
Puissance de calcul	126 Petaflops	74 Petaflops	1 Exaflops	1 Exaflops
Année de mise en service	2019	2023	2025-2026	2026

²⁵ Les AI Growth Zones au Royaume-Uni sont des zones géographiques stratégiquement désignées, visant à transformer des territoires en centres d'excellence pour l'IA. Le site candidat devra prouver pouvoir être alimenté à hauteur de >500 MW à horizon 2030, en plus d'autres critères reposant sur la faisabilité technique, la faisabilité de la livraison du site, les impacts locaux et le niveau de soutien nécessaire de la part du gouvernement.

²⁶ Pionnière de l'informatique française ayant contribué dans les années 1960 à 1980 à la conception et l'architecture des ordinateurs, ainsi qu'à leur miniaturisation.

Ecosystème et attractivité de talents

La France possède un écosystème dynamique composé de startups telles que Mistral AI²⁷ et Hugging Face²⁸, d'organismes de recherche publique (INRIA, CNRS), d'universités et d'organismes assimilés. En particulier, **le marché des outils et services IA à destination de cas d'usages concrets est particulièrement dynamique**. Quelques acteurs émergent dans la

fourniture de cloud et d'infrastructure de calcul pour l'entraînement d'IA, mais ce secteur reste dominé par les GAFAM. Parmi les créateurs de modèles d'IA avancés, Mistral AI et une poignée d'autres entreprises se démarquent. Le marché de la conception et des puces destinées à l'IA est dominé par NVIDIA et les fabricants taiwanais et sud-coréens : il n'y a actuellement pas d'acteur local d'envergure capable de rivaliser avec les leaders mondiaux.

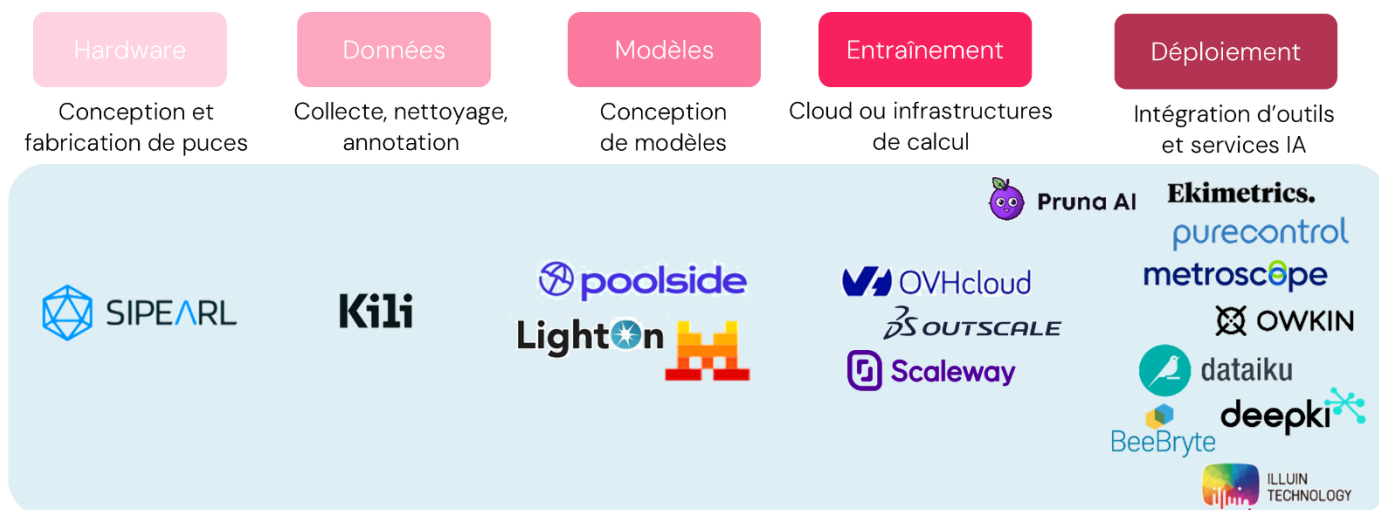


Figure 21 – Panorama non exhaustif des entreprises positionnées sur la chaîne de valeur de l'IA.

Des acteurs non-industriels (associations, ONG, gouvernement) émergent également dans l'optique d'évaluer et d'encadrer l'impact de l'IA.

Récupération et valorisation de la chaleur fatale

La récupération et la valorisation de la chaleur générée par les data centers est un levier stratégique qui permet de valoriser cette ressource énergétique et réduire la dépendance aux énergies fossiles. La directive européenne sur l'efficacité énergétique impose aux data centers de plus de 1 MW de valoriser leur chaleur fatale à partir de fin 2025 (Parlement Européen & Conseil de l'Union Européenne, 2023). La chaleur fatale peut être utilisée pour chauffer les locaux ou produire de l'électricité, elle peut également être injectée dans un réseau de chaleur urbain. **Le potentiel maximal de récupération de chaleur fatale issue des data centers en France était estimé à 6,9 TWh en 2020, et pourrait s'élever jusqu'à 12,2 TWh en 2030 (ADEME et al., 2023).** Des freins techniques, couplés à des frais

d'investissements élevés, constituent néanmoins une barrière à la valorisation de la chaleur fatale. Parmi les problématiques rencontrées, il y a la gestion de la variabilité offre-demande (production de chaleur constante par les data centers vs. demande saisonnière de chauffage), des régimes de température trop bas (35–65°C) par rapport aux réseaux urbains (75–90°C) ou encore l'absence de proximité avec les réseaux de chaleur.

²⁷ Startup française fondée en 2023 développant des LLM open source, dont son agent conversationnel [Le Chat](#).

²⁸ Entreprise franco-américaine développant des outils (plateforme de partage de modèles, bibliothèques open-source) pour utiliser l'intelligence artificielle. Hugging Face est également à l'initiative du [AI Energy Score](#), un outil pour évaluer et communiquer la consommation énergétique des modèles d'intelligence artificielle sur des tâches spécifiques.

Attractivité de la France à l'échelle européenne

De nombreux paramètres entrent en compte pour jauger de l'attractivité d'un pays en matière d'IA et de data center. La présence de ces derniers sur le territoire national est un point clé pour permettre au pays de se positionner dans la course à l'IA. Du point de vue des réseaux, l'Europe et la France disposent d'un réseau de transport maillé et robuste. Néanmoins, les hubs historiques européens de data centers connaissent une saturation croissante des réseaux, comme à Dublin, Francfort ou Amsterdam, ainsi qu'une augmentation des délais de raccordement des infrastructures. L'implantation des data centers

risque ainsi de basculer vers les pays dont les réseaux sont moins congestionnés. Au sein même de la France, les points de congestion sont là où se concentre déjà beaucoup la demande (Île de France, Marseille), d'où l'intérêt de l'identification de nouvelles zones d'implantation de data centers (Figure 20). En comparaison avec les autres pays européens, la France bénéficie de l'avantage d'une électricité décarbonée (contrairement à l'Irlande par exemple). Les lenteurs administratives freinent néanmoins l'installation des data centers, en comparaison avec l'existence d'une réglementation plus souple au Royaume-Uni ou d'une fiscalité plus attractive en Irlande.

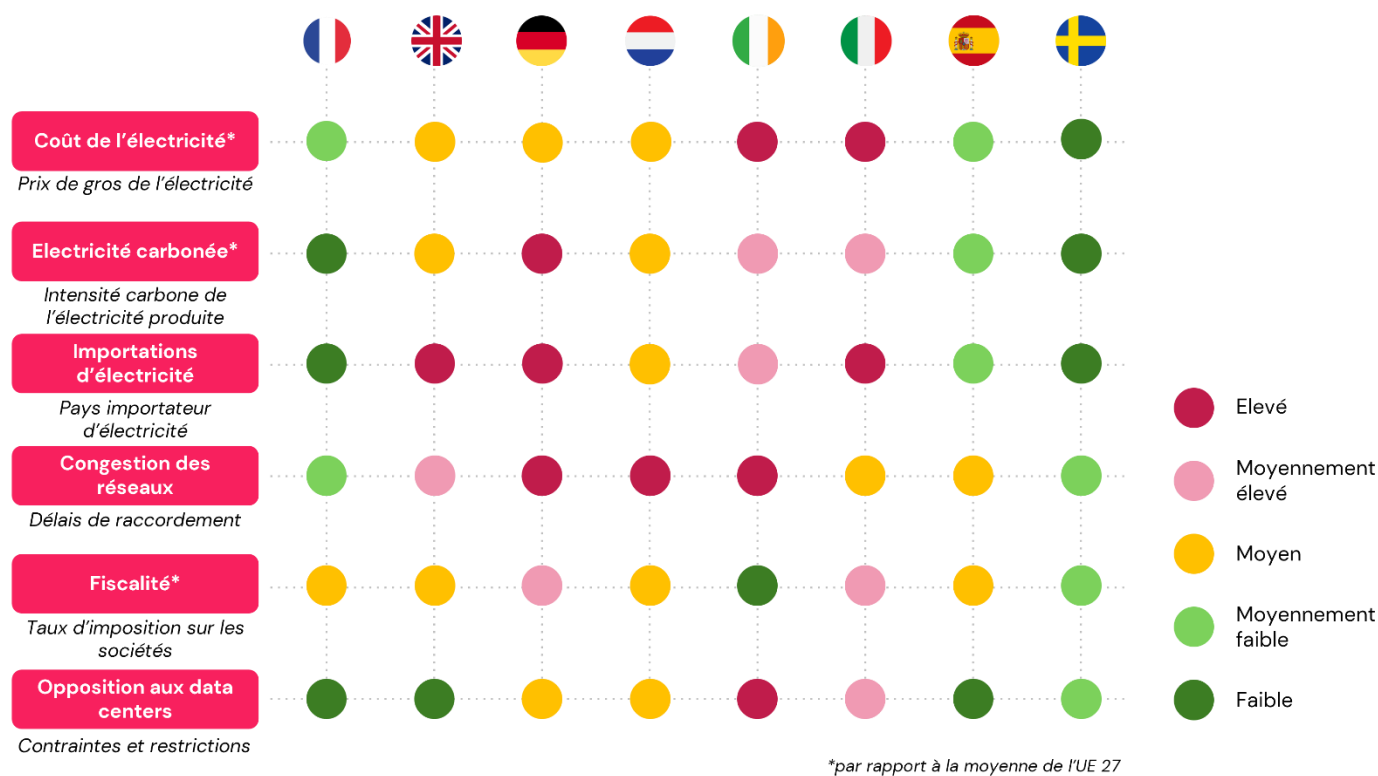


Figure 22 – Attractivité de la France et d'une sélection de pays européens pour l'implantation des data centers selon des critères relatifs au mix électrique, à la disponibilité des réseaux et aux politiques publiques. Ce tableau qualitatif et non exhaustif s'appuie sur diverses études et bases de données : Brijs et al., 2025; Ember et al., 2025; Eurelectric, 2025; EY & France Data Center, 2025 ; Swiss Energy-Charts, 2025 ; Tax Foundation Europe, 2025.

Les pays sélectionnés sont ceux hébergeant des hubs historiques de data centers : Allemagne, France, Pays-Bas, Irlande – qui représentent également le top 4 des pays en termes de consommation de data centers dans l'UE 27 – ainsi que le Royaume-Uni. L'Italie, la Suède et l'Espagne complètent le top 7 des consommations de data centers au sein de l'UE 27.

Le coût de l'électricité considéré est le prix spot moyen sur le marché de gros pour l'année 2024. L'intensité carbone de l'électricité produite dans un pays peut contribuer à diminuer l'empreinte carbone des data centers et aider les entreprises du numérique à atteindre leurs objectifs de durabilité. Les capacités de production d'électricité s'appuient sur les chiffres d'exportations et d'importations d'électricité effectuées par les pays concernés avec leurs voisins sur l'année 2024. Les pays qui n'importent pas d'électricité disposent d'un excédent structurel de production et donc d'une marge pour accueillir de nouvelles consommations. L'attractivité de la fiscalité d'un pays est mesurée par le taux d'imposition des entreprises, et peut être utilisée comme un levier favorable à l'implantation des data centers dans un pays donné. Enfin, l'analyse qualitative de l'état de congestion des réseaux et du soutien des politiques publiques à l'IA et l'implantation de data centers est basée sur Brijs et al., 2025. Les pays dont les réseaux sont déjà fortement sous tension régulent la construction de data centers, dont le délai de raccordement se rallonge. Certains pays comme la France, le Royaume-Uni ou la Suède mettent en place des stratégies nationales ambitieuses en matière d'IA.

Références

- Académie des Sciences., 2022. [Communiqué de presse – Énergie nucléaire : Quelle place pour les SMR?](#)
- ADEME et al., 2023. [Valorisation sur son territoire de la chaleur fatale issue des data centers.](#)
- AFNOR, 2024. [Un référentiel pour mesurer et réduire l'impact environnemental de l'IA.](#)
- Barroso et al., 2018. [The Datacenter as a Computer: Designing Warehouse-Scale Machines.](#)
- Brijs et al., 2025. [The Power of Compute : The Effects of Data Center Growth on Belgium's Energy System.](#)
- Catsaros, O., 2024. [Corporate Clean Power Buying Grew 12% to New Record in 2023.](#)
- Cottier et al., 2024. [How Much Does It Cost to Train Frontier AI Models?.](#) Epoch AI.
- CRE, 2025. [La CRE approuve la procédure de raccordement « fast track » pour raccorder les sites de consommation de grande puissance au réseau très haute tension \(400 kV\) de RTE.](#)
- CRU, 2025. [Review of Large Energy Users Connection Policy.](#)
- Daly, H., 2024. [Data centres in the context of Ireland's carbon budgets.](#)
- Data Center Dynamics, 2022. [EirGrid says no new applications for data centers in Dublin until 2028.](#)
- DeepSeek-AI et al., 2025. [DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning.](#)
- EDF, 2025. [EDF accompagne les entreprises du numérique pour le développement de nouveaux centres de données en France.](#)
- Elysée, 2025. [Faire de la France une puissance de l'IA.](#)
- Ember et al., 2025. [Carbon intensity of electricity generation.](#)
- Epoch AI, 2024. [Data on Notable AI Models.](#)
- Epoch AI, 2025a. [Data on AI Models.](#)
- Epoch AI, 2025b. [Machine Learning Trends.](#)
- Eurelectric, 2025. [Electricity Price.](#)
- European Commission., Joint Research Centre, 2024. [Energy consumption in data centres and broadband communication networks in the EU.](#)
- EY, & France Data Center, 2025. [Baromètre 2025—Etude d'impact économique, social et environnemental de la filière des data centers en France.](#)
- GIMELEC, 2024. [Contribution des data centers français à la flexibilité du système électrique.](#)
- GlobalData, 2025. [China Power Market Trends and Analysis by Capacity, Generation, Transmission, Distribution, Regulations, Key Players and Forecast to 2035.](#)
- Goldman Sachs, 2024. [Generational Growth: AI, data centers and the coming US power demand surge.](#)
- Google, 2025. [Our first geothermal energy deal in Asia.](#)
- IEA, 2025. [Energy and AI.](#)
- Kairos Power, 2024. [Google and Kairos Power Partner to Deploy 500 MW of Clean Electricity Generation.](#)
- Kamiya, G., Coroamă, V., 2025. [Data Centre Energy Use: Critical Review of Models and Results.](#)
- Lawrence Berkeley National Laboratory, 2024. [Queued Up : Characteristics of Power Plants Seeking Transmission Interconnection.](#)
- Lorentz et al., 2024. [Powering AI.](#)
- LowCarbonPower, 2025. [Mix Électrique de États-Unis 2024.](#)
- Ludvigsen, K., 2023. [The carbon footprint of GPT-4.](#)
- Masanet et al., 2020. [Recalibrating global data center energy-use estimates.](#)
- McKinsey, 2025. [AI in the workplace : A report for 2025.](#)
- McKinsey, 2024. [The role of power in unlocking the european AI revolution.](#)
- Mistral AI, 2025. [Our contribution to a global environmental standard for AI.](#)
- Norris et al., 2025. [Rethinking Load Growth—Assessing the Potential for Integration of Large Flexible Loads in US Power Systems.](#)
- Our World in Data, 2025. [Annual freshwater withdrawals.](#)
- Paccou, R., Wijnhoven, F., 2024. [Artificial Intelligence and Electricity.](#)
- Palmer, I. D., 2025. [Gas, Nuclear, Renewables Battle Over Power For Meta's New Data Center.](#)
- Parlement Européen, & Conseil de l'Union Européenne, 2023. [Directive \(UE\) 2023/1791 du Parlement européen et du Conseil relative à l'efficacité énergétique.](#)
- Patel et al., 2024. [AI Datacenter Energy Dilemma – Race for AI Datacenter Space.](#)
- RTE, 2022. [Futurs énergétiques 2050 : Les chemins vers la neutralité carbone à horizon 2050.](#)
- RTE, 2025a. [Synthèse—Bilan électrique 2024.](#)
- RTE, 2025b. [Data centers : 11 chiffres sur leur essor en France et leurs besoins en électricité.](#)
- Sfen, 2024a. [Constellation Energy et Microsoft vont relancer le réacteur de Three Mile Island.](#)
- Sfen, 2024b. [Google et Kairos Power s'engagent pour déployer une flotte de SMR.](#)
- Shehabi et al., 2024. [United States Data Center Energy Usage Report.](#)
- Siddik et al., 2021. [The environmental footprint of data centers in the United States.](#)
- Statista, 2025a. [Data centers worldwide by country 2025.](#)
- Statista, 2025b. [Data center average annual PUE worldwide from 2007 to 2024.](#)
- Sweeney, R. W., Gordon, N., 2025. [Geothermal Energy and U.S. Competitive Advantage : Drill, Baby, Drill.](#)
- Swiss Energy-Charts, 2025. [Les échanges commerciaux aux frontières en 2024.](#)
- Tax Foundation Europe, 2025. [2025 Corporate Income Tax Rates in Europe.](#)
- United Nations, 2024. [UN World Water Development Report.](#)
- You, J., 2025. [How much energy does ChatGPT use?.](#)

A propos de Zenon research

Zenon Research est une association loi 1901 et un think tank indépendant. Ses travaux, fondés sur la science, sont dédiés aux technologies pour un monde bas carbone.

Contactez-nous



[/zenon-research](#)



www.zenon.ngo



contact@zenon.ngo

Pour plus d'informations sur ce sujet

Lucie Zhang

Analyste

Zenon Research

lucie@zenon.ngo

Zenon Research souhaite remercier les experts suivants pour leurs connaissances et leurs contributions sur les différents aspects de la consommation énergétique de l'intelligence artificielle et des data centers :

Cédric Philibert – Ifri
Claire Sagnol – Ekimetrics
Claude Taddonki – Mines Paris – PSL
Devon Swezey – Google
Emilien Boucaud – Ekimetrics
Floriane Fay – Google
Marlène de Bank – The Shift Project
Nicolas Lhuillier – RTE
Noëllie Viennet – APL Data Center
Olivier Lebois – RTE
Roblex Nana Tchakouté – Mines Paris – PSL
Samuel Rincé – GenAI Impact
Vincent Poncet – Google
Xavier Blot – emlyon business school

Le contenu de ce rapport représente les opinions de Zenon Research et ne doit pas être considéré comme représentant celles des relecteurs ou des organisations auxquelles ils sont affiliés.

