INSTITUT Montaigne

Infrastructures numériques

Un plan décisif

MARS 2025

Think tank de référence en France et en Europe, l'Institut Montaigne est un espace de réflexion indépendant au service de l'intérêt général. Ses travaux prennent en compte les grands déterminants économiques, sociétaux, technologiques, environnementaux et géopolitiques afin de proposer des études et des débats sur les politiques publiques françaises et européennes. Il se situe à la confluence de la réflexion et de l'action, des idées et de la décision.

RAPPORT - Mars 2025

Infrastructures numériques

Un plan décisif

Les rapports de l'Institut Montaigne proposent des analyses exhaustives, issues d'une réflexion collégiale et ont vocation à identifier des solutions de long terme.

Note d'éclairage
Se situer et rendre intelligible notre environnement

Note d'enjeux
Poser des constats et identifier des problématiques

Note d'action
Formuler des recommandations opérationnelles

Opération spéciale
Sonder, chiffrer, expérimenter

Rapport
Analyser et proposer collégialement des solutions de long terme

Avant-propos

Notre compétitivité et l'efficacité de nos moyens de communication dépendent des infrastructures numériques. Elles sont le socle de notre organisation moderne. Les Américains en ont fait un vecteur de conquête économique, des technologies les plus avancées aux usages du quotidien. Les Chinois en usent comme d'un instrument de puissance politique et exploitent sans vergogne les divergences occidentales. Pendant ce temps, d'autres pays adoptent des méthodes de guerre dite hybride de plus en plus intrusives.

Les annonces fracassantes de l'Administration Trump ou celles, plus mesurées, du Sommet pour l'action sur l'IA témoignent de cette compréhension intrinsèque. La France, pour sa part, tarde à définir son positionnement, oscillant entre renoncements technologiques, régulation et velléités de souveraineté sur l'ensemble de la chaîne de valeur.

Ce rapport vise à mobiliser la puissance publique sur la nécessité d'une stratégie lucide et réaliste en matière d'infrastructures numériques de traitements de données et de réseaux de télécommunication. Le triptyque « Financements, Énergie et Talents », à l'origine de la force de frappe américaine, nous est aussi accessible, grâce à notre énergie bon marché et à l'excellence de nos ingénieurs et chercheurs.

Nous proposons un plan stratégique qui permettrait à la France de sécuriser des usages souverains prioritaires, maîtriser ses dépendances, demeurer un acteur majeur du traitement de données et du calcul intensif et valoriser l'excellence technologique de son infrastructure de réseaux. Ces actions immédiatement mobilisables supposent une adaptation de l'environnement réglementaire national et européen, visant à le rendre plus favorable aux entreprises.

Il y a urgence pour concentrer nos efforts, capitaliser sur nos atouts, rester un acteur de rang mondial en excellant sur des usages réservés. Il en va de notre liberté.

Marie-Pierre de Bailliencourt,
directrice générale de l'Institut Montaigne

Les annonces franco-européennes lors du Sommet de l'Action de l'IA de février 2025 témoignent d'un engagement financier inédit : plus de 300 Mds € investis, dont 109 Mds € pour la France et 200 Mds € pour l'Union européenne, pour renforcer les infrastructures numériques. Mais si ces investissements traduisent une ambition affirmée, ils ne suffisent pas à garantir une véritable maîtrise stratégique de nos infrastructures. **Les *data centers* ne sont qu'un maillon de la chaîne, et leur développement doit s'inscrire dans une vision globale et cohérente, intégrant l'ensemble des briques technologiques essentielles à la souveraineté numérique allant des circuits intégrés aux applications logicielles, les réseaux, le *edge*, le *cloud*, l'IoT et le calcul haute performance**.

Pour la France, cela appelle une approche ciblée et pragmatique. Plutôt que de chercher à combler systématiquement notre retard dans certains secteurs dominés par les États-Unis ou la Chine, nous devons concentrer nos efforts sur des segments où nous avons un avantage compétitif ou un levier stratégique. **Notre expertise en gestion du risque de cyberattaque, notre capacité industrielle à produire des infrastructures énergétiquement sobres, et nos capacités en matière de télécommunications – notamment en 5G – et *edge computing* constituent des atouts que nous pouvons structurer en un écosystème cohérent**. Cela passe par des choix stratégiques assumés, sans dispersion des moyens et en concentrant les efforts là où un *leadership* européen reste atteignable.

Le développement de l'IA et des futurs usages numériques nécessite une anticipation des infrastructures sous-jacentes, qu'il s'agisse de puissance de calcul, de stockage, ou du maillage géographique des *data centers* et réseaux. **Ce rapport structure cette réflexion en identifiant les ressources critiques et en formulant des recommandations constituant des éléments clés pour une planification stratégique autour des infrastructures numériques, avec des objectifs mesurables et**

un cadre programmatique clair. Les récentes annonces internationales ont confirmé que les infrastructures numériques ne se cantonnent pas à un sujet technologique mais constituent le socle de la puissance et de la souveraineté économiques du XXIe siècle.

Dans un contexte budgétaire contraint pour la France et l'Europe, la question centrale n'est pas seulement de savoir combien investir, mais avec quelle feuille de route. Quels usages devons-nous impérativement maîtriser à l'avenir et comment les sécuriser ? À cet égard, le *edge computing* représente une opportunité stratégique pour reprendre le contrôle sur les données sensibles des utilisateurs européens, en évitant qu'elles ne soient systématiquement captées et exploitées par des infrastructures étrangères. Contrairement au *cloud*, où la domination des géants américains et chinois est déjà actée, le *edge* reste un territoire à conquérir. En 2031, plus de la moitié des données seront traitées directement en périphérie du réseau, soit en mode « *edge* »[1]. Selon la Fondation Linux[2], le *edge computing* pourrait générer quatre fois plus d'activités que le *cloud* et traiter 75 % des données mondiales, ce qui en fait un levier stratégique majeur pour structurer un continuum souverain réseaux-cloud-edge-IoT.

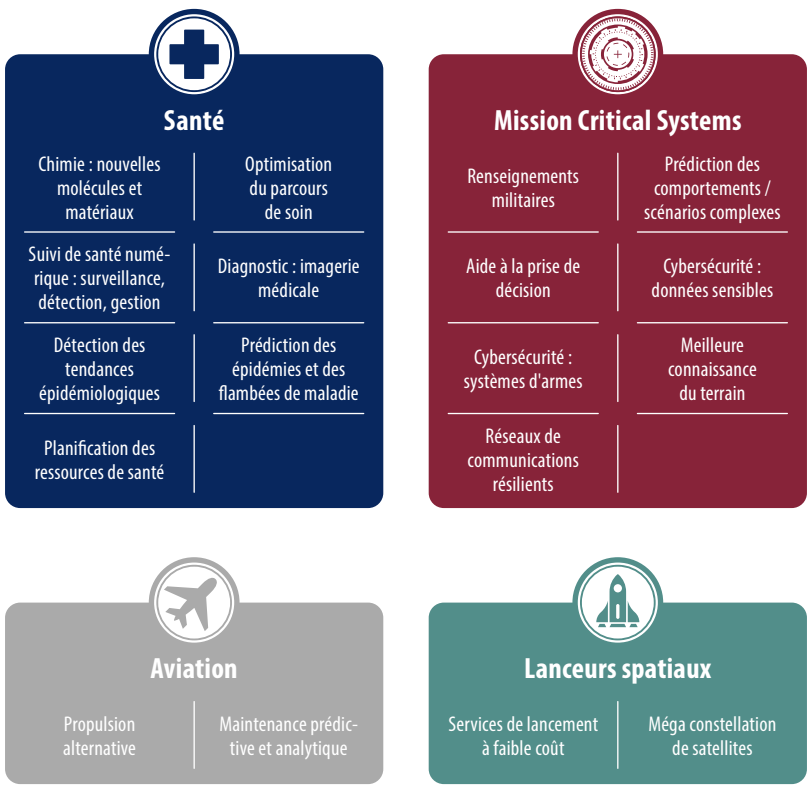
L'Europe dispose déjà d'atouts pour bâtir cette souveraineté numérique : des réseaux 5G privés performants, une énergie décarbonée et compétitive, et des ingénieurs hautement qualifiés. Mais pour concrétiser cette ambition, il est impératif d'adresser l'enjeu central de l'accès aux données de qualité, essentielles à l'entraînement des modèles d'IA et au déploiement de solutions numériques efficaces. **En d'autres termes, la souveraineté sur les usages critiques passe par la souveraineté sur l'ensemble des infrastructures sous-jacentes : sans contrôle sur les réseaux, le *cloud*, le *edge* et l'IoT, il est impossible de garantir l'autonomie stratégique sur les services numériques de demain**.

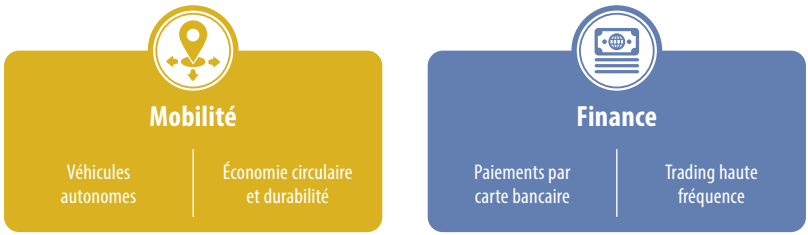
[1] *https://www.rtinsights.com/edge-computing-set-to-dominate-data-processing-by-2030/.*

[2] *A. Joshipura, directeur général des réseaux au sein de la Fondation Linux (2019). Discours lors de l'Open Networking Summit.*

L'Institut Montaigne propose une première catégorisation d'usages prioritaires pour lesquels déployer cette infrastructure en trois axes : (i) les domaines dans lesquels l'Europe dispose d'un savoir-faire reconnu à l'échelle mondiale ; (ii) les secteurs nécessitant une autonomie stratégique, que ce soit au niveau national ou européen ; et (iii) les champs où une accélération est impérative pour consolider ou renforcer un avantage compétitif.

Graphique n° 1 • Cas d'usages critiques pour lesquels il faut une infrastructure edge-cloud-Iot-réseaux souveraine





Trois priorités stratégiques : le développement d'infrastructures numériques de traitement de données en France, l'exportation de l'excellence des infrastructures numériques françaises de réseau et une adaptation du cadre normatif européen aux nouvelles dynamiques technologiques mondiales.

Au niveau national, investir dans des infrastructures de traitement de données est une priorité stratégique pour combler le retard de la France et de l'Europe face aux États-Unis et à la Chine. Dans un contexte où le marché mondial des supercalculateurs croît de 40 % par an avec une véritable « course à la puissance de calcul », de nombreux pays européens, au premier rang desquels le Royaume-Uni, prévoient de multiplier par 20 leur capacité de calcul d'ici 2030. Les supercalculateurs exascale publics comme Jupiter et Alice Recoque porteront certes la part européenne de la puissance mondiale à 19 % en 2026, mais sans investissements supplémentaires, cette part pourrait chuter à 5 % en 2030. Pour combler ce fossé, il est indispensable d'investir sans plus tarder, et d'associer les acteurs privés européens qui détiennent près de 10 fois moins de GPU que les acteurs américains aujourd'hui. Parallèlement, outre les investissements conséquents nécessaires, il est tout aussi crucial de planifier la construction de *data centers* sur le territoire national pour héberger ces capacités de calcul, tout en tenant compte des intérêts économiques et sociaux de l'ensemble des acteurs concernés. Or, la France est confrontée à deux obstacles majeurs en la matière. D'une part, le système de raccordement électrique, basé sur une logique de « premier arrivé, premier servi », favorise des pratiques spéculatives :

certains acteurs bloquent des capacités réseau sans intention réelle de construire, ralentissant les projets stratégiques. D'autre part, les délais administratifs restent excessifs, malgré la volonté politique affichée lors du Sommet de l'Action de l'IA pour les réduire.

Toujours au niveau national, la compétitivité et l'excellence des infrastructures numériques de réseau est à préserver et à exporter à l'international, car elles sont le socle sur lequel repose tout projet d'infrastructures numériques de traitement de données. En particulier, la 5G publique comme privée représente un levier stratégique important pour renforcer la compétitivité des entreprises françaises, mais aussi pour créer des offres souveraines faisant appel à du *edge computing*. Pourtant, en France, son adoption reste limitée, notamment dans les environnements industriels en raison des coûts induits et d'une absence de perception claire sur la valeur ajoutée des usages. Exploiter ce potentiel nécessite de mettre en place des solutions de type *Platform as a Service* (PaaS) pour commercialiser à moindre coût des fonctionnalités 5G avancées « sur étagère ». C'est comme cela que les États-Unis ont pu compenser l'absence d'acteurs télécoms sur leur marché, au profit des *hyperscalers*, qui sont désormais bien positionnés sur les infrastructures numériques de réseau.

Au niveau européen, un changement d'échelle s'impose pour se doter de moyens normatifs nécessaires à une compétition internationale non biaisée. Sur le plan réglementaire, la fragmentation du marché européen constitue un désavantage majeur, avec 60 opérateurs actifs en Europe contre seulement 4 aux États-Unis. Cette disparité est accentuée par des différences structurelles : par exemple, T-Mobile intègre les télécommunications dans les infrastructures numériques, ce qui n'est pas le cas en Europe. De plus, l'absence de réciprocité dans l'accès aux marchés publics aggrave cette situation. Alors que d'autres pays favorisent systématiquement leurs acteurs nationaux ou régionaux, l'Europe se prive de la possibilité de promouvoir des achats véritablement souverains, ce qui revient à affaiblir ses propres capacités industrielles et stratégiques.

Recommandations de politiques publiques

Recommandation 1

Construire une offre souveraine cloud-réseau-edge-IoT « bout-en-bout » au niveau français et européen pour des usages aux dépendances maîtrisées.

Recommandation 2

Entamer, dès aujourd'hui, *a minima*, en France, la construction de 6 supercalculateurs exaflopiques additionnels afin de proposer à l'Europe une capacité de calcul de 9 exaflops.

Recommandation 3

Construire une réelle planification étatique en matière d'approvisionnement électrique pour mailler le territoire français en *data centers* de grande capacité en anticipant les usages futurs.

Recommandation 4

Capitaliser sur le lancement des 35 sites clés en main pour raccourcir les délais de construction de *data centers* dont l'intérêt économique et social est démontré en simplifiant les procédures administratives.

Recommandation 5

Lancer un projet « commando » pour développer des formations continues rapprochant les métiers de l'infrastructure numérique de réseaux de ceux de l'infrastructure numérique de traitement de données.

Recommandation 6

Accélérer le déploiement de la 5G en milieu industriel, au moins sur les projets *greenfield*, en ciblant résolument les besoins des entreprises utilisatrices (TPE-PME-ETI).

Recommandation 7

Sécuriser les nœuds critiques de distribution des câbles terrestres par une politique d'enfouissement raisonnée des câbles terrestres et aériens.

Recommandation 8

Valoriser l'atout stratégique que représentent les câbles sous-marins français *via* une stratégie intégrée combinant surveillance renforcée, investissements ciblés en Outre-mer et influence accrue dans les instances internationales.

Recommandation 9

Adapter et simplifier le cadre normatif européen pour renforcer notre compétitivité et favoriser les consolidations d'acteurs à l'échelle mondiale.

Table des matières

Les infrastructures numériques, invisibilisées au service des usages qu'elles permettent, jouent pourtant un rôle aussi structurant que les routes ou les entrepôts logistiques depuis l'Empire romain. À l'époque, les voies romaines – bien qu'à vocation militaire au départ – et les entrepôts logistiques ont façonné le commerce, les échanges et la puissance de l'Empire en créant un réseau stratégique pour relier les territoires. Aujourd'hui, les infrastructures numériques – des *data centers* aux réseaux de fibres optiques en passant par les satellites – remplissent une fonction similaire dans le monde virtuel : elles connectent les individus, les entreprises et les États, soutenant l'économie, l'innovation et la sécurité. Mais comme pour les routes ou les entrepôts, leur développement ne peut être laissé au hasard. **Une approche stratégique est essentielle pour garantir leur robustesse, leur résilience et leur capacité à accompagner les grandes transformations économiques et sociétales, ainsi que les futurs usages, encore imprévisibles, qui redéfiniront nos modes de vie et de travail**. C'est d'autant plus crucial que les infrastructures numériques sont aujourd'hui présentes en quantité insuffisante sur le territoire français et ne font pas l'objet de planification.

L'infrastructure représente la structure essentielle sans laquelle la technologie ne peut se déployer, comme le montre l'exemple de l'informatique. Dans ce secteur les premières avancées se concentraient sur le développement des machines elles-mêmes (le *hardware*) : les câbles sous-marins du XIX^{e} siècle[3], les lignes télégraphiques et téléphoniques, les premiers ordinateurs centraux comme le CDC 6600[4], ou encore les premiers *data centers*[5] étaient des infrastructures matérielles destinées à traiter des informations. **Ces équipements matériels constituaient la base physique sur laquelle reposait toute possibilité de**

[3] *Le premier câble sous-marin construit reliait Douvres et Calais en 1851, pour transmettre des messages en morse.*

[4] *Ordinateur créé par le Control Data Corporation des États-Unis en 1964 pour permettre des calculs complexes dans la recherche scientifique.*

[5] *L'un des premiers fut construit par IBM dans les années 1960 pour abriter les premiers ordinateurs centraux, appelés mainframes.*

transmission, de stockage et de traitement des données. Le *hardware* représentait donc la structure essentielle qui rendait l'essor de la technologie possible. Le *software*, soit la partie applicative, est apparu plus tardivement, avec pour rôle de donner des instructions aux machines et d'exploiter les capacités offertes par le matériel. Les applications ont progressivement pris de l'importance avec l'essor d'Internet, offrant une interface entre l'utilisateur et la machine et ajoutant ainsi de la valeur à la puissance brute des infrastructures matérielles. **Le *software* a ainsi permis d'adapter les infrastructures aux besoins humains, transformant les réseaux physiques en réseaux d'information exploitables et accessibles. Ce que l'on appelle la « fin des couches » désigne cette évolution où un même acteur peut, s'il le souhaite, désormais gérer plusieurs fonctions critiques simultanément : calcul, intelligence artificielle (IA), *cloud* et connectivité**.

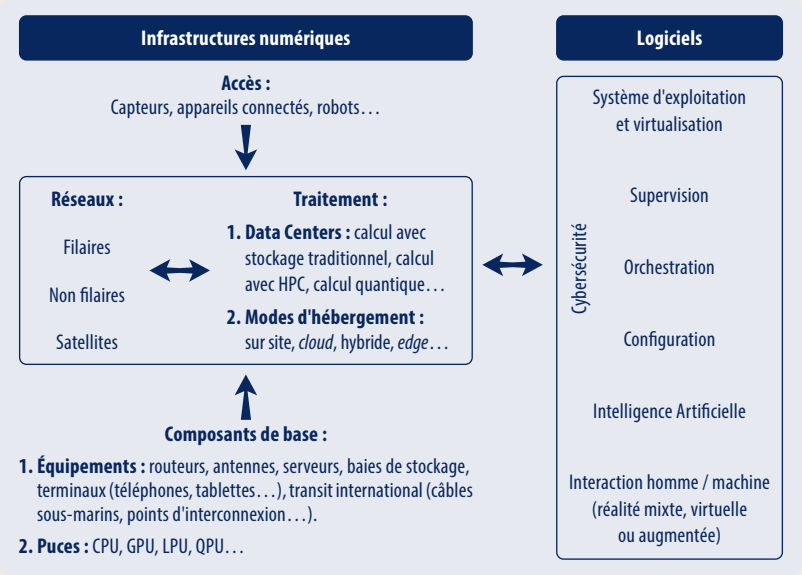
Aujourd'hui, la distinction classique entre *hardware* et *software* ne suffit plus à décrire les infrastructures numériques. Un continuum s'est établi entre le « pure *hardware* » et le « pure *software* », avec un vaste champ intermédiaire difficile à catégoriser. **Ce champ s'est développé grâce à la notion de « virtualisation », qui permet de découpler les ressources physiques de leurs usages**. Par exemple, une machine virtuelle permet de diviser un serveur physique en plusieurs environnements indépendants, optimisant ainsi l'utilisation des ressources. Cette virtualisation a complexifié la structure des infrastructures numériques, désormais organisées en couches : matérielle, logique et applicative. La notion de « virtualisation du réseau » *(software defined networking)* fait référence au processus qui consiste à simplifier la complexité technique sous-jacente du *hardware*, en utilisant des couches logicielles pour proposer des services de connectivité standardisés et flexibles. **Le *cloud computing* illustre cette zone intermédiaire : les solutions des *hyperscalers*[6], massives et homogènes, mutualisant un nombre colossal d'applications et d'utilisateurs, se**

[6] *Entreprise gérant des infrastructures numériques massives, évolutives et hautement automatisées pour offrir des services de cloud, de calcul ou de stockage à grande échelle, comme Amazon Web Services, Microsoft Azure ou Google Cloud.*

rangent du côté des infrastructures car elles fournissent des ressources virtualisées et standardisées à grande échelle, indépendamment des applications spécifiques qu'elles hébergent. À l'inverse, les solutions décentralisées de type *edge computing* sont bien plus applicatives, car elles visent à rapprocher les capacités de calcul des utilisateurs ou des objets connectés, répondant ainsi à des besoins spécifiques en termes de latence, de traitement localisé des données et d'autonomie par rapport aux centres de données centraux.

Nous définirons donc les infrastructures numériques par leur fonction, soit l'ensemble des systèmes matériels et immatériels qui permettent le stockage, la transmission et les capacités de calcul. En effet, la notion d'infrastructure numérique doit être comprise dans ses différentes dimensions matérielles, logicielles et organisationnelles.

Graphique n° 2 • Représentation visuelle de la chaîne de valeur des infrastructures numériques



La cartographie en page précédente, qui n'a pas vocation à être universelle mais suffisante pour mener notre analyse, servira de cadre de référence tout au long de ce rapport pour permettre au lecteur de maintenir une perspective d'ensemble simplifiée.

L'accès aux infrastructures numériques se fait au plus proche de nos usages : il se compose de terminaux (ordinateurs, smartphones, tablettes et autres équipements intelligents) par lesquels les utilisateurs accèdent et interagissent avec les infrastructures numériques. Sont également incluses des briques d'IoT (Internet des objets)[7]. Les technologies de capteurs englobent des dispositifs qui détectent et mesurent des informations physiques ou environnementales (comme la température, l'humidité, la lumière, le mouvement, etc.) et transmettent ces données pour traitement et analyse. En complément, les robots et appareils connectés participent également à cet écosystème, jouant un rôle croissant dans l'automatisation et la gestion des infrastructures numériques. Les robots industriels, domestiques ou de service sont équipés de capteurs et de logiciels avancés qui leur permettent de recueillir, traiter et transmettre des informations en temps réel.

Il existe une forte porosité entre les infrastructures de réseau et de traitement. Les infrastructures numériques de réseau désignent les systèmes qui assurent le transport, l'échange et la connectivité des données, incluant les câbles sous-marins, les réseaux fibre optique, les antennes mobiles (5G/6G), les routeurs et les satellites de communication. Les infrastructures numériques de traitement regroupent quant à elles les équipements et centres permettant de stocker, traiter et analyser des données, tels que les *data centers*, les supercalculateurs, les serveurs de calcul en *edge computing*, et les plateformes *cloud*.

[7] *Réseau d'objets physiques recueillant des données en temps réel et intégrant des technologies de capteurs, de logiciels et d'autres technologies permettant de se connecter et d'échanger des données avec d'autres appareils et systèmes sur Internet.*

Les infrastructures numériques de réseau sont à la fois publiques et privées. Au niveau public, elles se composent des réseaux d'accès mobile (4G, 5G, 6G) et fixe (FTTH et cuivre), de la technologie Open RAN[8], des réseaux de transport IP et Transmission Optique et du cœur de réseau *(core network)* et plateformes de services associés. Au niveau privé, elles sous-tendent des usages industriels, et regroupent la fibre mutualisée professionnelle (FTTE) et dédiée jusqu'au local de l'entreprise en question (FTTO).

Encadré n° 1 • Focus sur les différentes technologies de réseau public

1 **Les réseaux d'accès mobile – 2G/3G/4G/5G permettant de raccorder les smartphones / tablettes via une technologie sans fil aux sites radio de l'opérateur telecom, constitués d'équipements d'accès mobile**. Chaque opérateur dispose en France de 20 à 30 000 sites radio en propre ou partagés suivant les zones et les opérateurs et répartis sur l'ensemble du territoire.

2 **Les réseaux d'accès fixes – Raccordement des logements et entreprises *via* le réseau cuivre et la fibre FTTH (Fiber-To-The-Home) aujourd'hui, puis totalement fibre d'ici 2030 en France**. Chaque utilisateur raccordé à la fibre se situe à 10/20 km maximum du site de l'opérateur sur lequel se trouve les équipements d'accès fixe. Les débits à l'accès sur le réseau FTTH varient de quelques centaines de Mbits/s jusqu'à 10 Gbits/s.

[8] *Architecture de réseau d'accès radio qui standardise les interfaces entre les différentes parties du réseau, permettant d'utiliser des équipements matériels et logiciels de fournisseurs différents. Par exemple, le japonais Rakuten Mobile a construit un réseau mobile 4G et 5G entièrement basé sur l'architecture Open RAN, en intégrant des composants provenant de multiples fournisseurs comme Altiostar (logiciel RAN), Nokia (antennes), et Intel (processeurs).*

3 **Le réseau de transport IP et Transmission Optique :** ces réseaux permettent la transmission de données au niveau régional et national, ainsi que le raccordement aux autres opérateurs à l'international *via* des points d'interconnexion. Ils sont hiérarchisés sur plusieurs niveaux (accès, régional/national) et sont constitués de routeurs chinois et américains et d'équipements de transmission optique *Wavelength Division Multiplexing* (WDM) – une technologie qui permet de transmettre plusieurs canaux de données simultanément sur une même fibre optique avec des longueurs d'ondes différentes pour chaque signal – supportant des débits de 10 Gbits/s à plusieurs centaines de Gbits/s voire 1 Terabit/s pour les toutes dernières générations.

4 **Le cœur de réseau et les plateformes de services :** ce sont des solutions logicielles et matérielles relativement centralisées dans le réseau permettant de contrôler l'accès aux réseaux, d'acheminer les messages et communications voix / données.

Les infrastructures de traitement regroupent les capacités de calcul et de stockage des données. Elles incluent les lieux physiques où les données sont conservées et produites grâce à des ressources de calcul de plus en plus puissantes, principalement les *data centers*, dont les serveurs constituent l'unité de base. **Ces centres intègrent désormais des technologies avancées comme les supercalculateurs, conçus pour accélérer la génération et le traitement des données**. Un supercalculateur est une architecture informatique composée de serveurs de calcul spécialisés, connectés par un réseau ultra-rapide (souvent 4 liens à 400 ou 800 Gb/s) et équipés de systèmes thermiques performants pour gérer leur densité élevée[9]. Ces systèmes sont intégrés à

[9] *Dans ce contexte, la notion de « densité » fait référence à la concentration de composants informatiques (processeurs, GPU, mémoire, etc.) dans un espace physique restreint.*

des architectures centrées sur les données, capables de traiter des flux massifs (plusieurs To/s), et reposent sur des environnements logiciels optimisés pour exécuter des calculs en parallèle sur des milliers, voire des centaines de milliers, de processeurs (CPU, GPU ou plus récemment TPU chez Google). **Les données peuvent ensuite être hébergées selon différentes modalités, localement ou « *on-premise* », dans un cloud géré à distance, de manière hybride ou « en mode *edge* » à la frange du réseau, soit à proximité des utilisateurs ou des appareils connectés pour réduire la latence**[10].

Les logiciels vont ensuite exploiter ces données avec différents outils, que sont les systèmes d'exploitation, la supervision, l'orchestration[11], la configuration, l'intelligence artificielle, la cybersécurité et la virtualisation. **Parce que ces outils sont désormais pleinement intégrés au *hardware*, les infrastructures numériques sont aujourd'hui considérées comme « intelligentes » par nature**. Par exemple, des logiciels spécifiques aux puces comme les microprogrammes *(firmware)* pour les CPU, des solutions de cybersécurité pour les réseaux comme les pare-feu, et des outils d'intelligence artificielle pour le traitement des données comme les plateformes de *machine learning*.

Les infrastructures numériques sont au cœur de nos sociétés modernes : elles ne se limitent pas aux entreprises technologiques ou aux experts du secteur, mais impactent directement notre quotidien, que ce soit par l'accès à Internet, la sécurité de nos données ou le fonctionnement de nos services essentiels. Elles sont devenues une composante incontournable de notre autonomie collective, en garantissant notre capacité à

[10] *Délai entre l'envoi d'une requête et la réception de la réponse correspondante dans un système de communication ou de traitement. Elle se mesure généralement en millisecondes et reflète la vitesse à laquelle les données transitent, influençant directement la réactivité des applications et des services numériques.*

[11] *Processus qui consiste à automatiser et coordonner des tâches, services ou applications pour qu'ils fonctionnent ensemble de manière fluide dans un système complexe. Par exemple, Kubernetes orchestre le déploiement, la mise à l'échelle et la gestion des conteneurs dans des environnements cloud.*

produire, échanger et protéger les informations stratégiques. Le projet Stargate[12], présenté par Donald Trump lors de son investiture, illustre de manière probante que l'infrastructure numérique est aujourd'hui le nerf de la guerre, car elle conditionne à la fois la puissance économique, technologique et stratégique des nations. Elle mobilise des acteurs essentiels à plusieurs niveaux : le financement, avec des investissements massifs comme ceux de SoftBank, la convergence des technologies *cloud*, *edge*, réseau et IoT portée par des entreprises comme Oracle, et surtout les usages, illustrés par des applications d'IA avancées telles que OpenAI. Ce projet montre clairement que maîtriser les infrastructures numériques ne se limite pas à construire des réseaux, mais englobe les financements comme l'intégration technologique, et que la création de valeur se situe en grande partie au niveau des applications. **C'est cette maîtrise globale qui détermine aujourd'hui la compétitivité et l'autonomie des États, dans un monde où la dépendance aux technologies numériques ne cesse de croître. Le développement et la résilience des infrastructures numériques est une priorité qui nous concerne tous**.

Aujourd'hui, les infrastructures numériques sont confrontées à plusieurs défis majeurs, à commencer par l'absence d'une stratégie étatique claire pour définir les priorités de développement et identifier les composantes de la chaîne de valeur qui doivent impérativement rester sous contrôle souverain. S'ajoutent à cela un retard préoccupant dans le déploiement des infrastructures de traitement, une faible capacité à exporter nos solutions de réseau, et un cadre juridique européen imparfait.

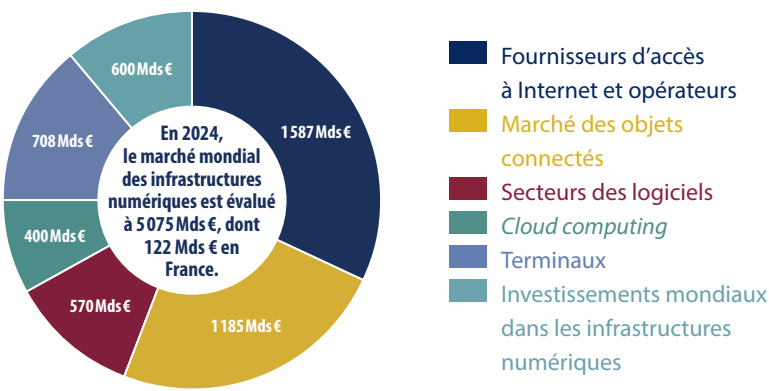
[12] *CNN, C.Duffy, 21 janvier 2025 « Trump announces a $500 billion AI infrastructure investment in the US », https://edition.cnn.com/2025/01/21/tech/openai-oracle-softbank-trump-ai-investment?cid=ios_app.*

Ce rapport analyse les enjeux des infrastructures numériques et propose des réponses concrètes pour y faire face. Il met l'accent sur la nécessité d'une stratégie industrielle claire pour leur gestion, leur développement et leur adaptation aux besoins futurs. L'IA, désormais partie intégrante de ces infrastructures, génère des volumes croissants de données, y compris synthétiques, qui plaident pour une stratégie de gestion de la donnée. Ce défi implique de développer des infrastructures capables de distinguer les données fiables des données biaisées ou inutilisables. Dans un contexte où le retour en arrière n'est plus une option, il est urgent de s'organiser pour anticiper ces besoins et garantir une maîtrise à la fois technique et stratégique de ces évolutions.

1 Dans un secteur désormais dominé par les grands acteurs du *cloud*, la puissance publique peine à structurer une stratégie cohérente

En 2024, le marché mondial des infrastructures numériques est évalué à 5 075 Mds €, dont 122 Mds € en France. Il est dominé par les fournisseurs d'accès à Internet (FAI) et les opérateurs (1 587 Mds €[13]), ainsi que par le marché des objets connectés (1 185 Mds €[14]). Les secteurs des logiciels (570 Mds €[15]), du *cloud computing* (400 Mds €[16]) et des terminaux (708 Mds €[17]) affichent également une forte croissance. Les investissements mondiaux dans les infrastructures numériques, déjà massifs à hauteur de 276 Mds $, devraient atteindre 600 Mds $ d'ici 2035[18].

Graphique n° 3 • Marché mondial des infrastructures numériques en 2024



Le marché des infrastructures numériques, autrefois dominé par les opérateurs télécoms et les fournisseurs d'accès à internet (FAI), évolue désormais au profit des *hyperscalers*, historiquement centrés sur le *cloud*. Le *hardware* ne représente plus que 15 % du chiffre d'affaires mondial des télécommunications, contre 60 % pour les applications et 15 % pour les terminaux. La comparaison entre opérateurs télécoms et *hyperscalers* est éloquente : les trois principaux opérateurs mondiaux, AT&T, Verizon, China Telecom, génèrent un chiffre d'affaires combiné de près de 380 Mds € contre 1 795 Mds € pour les GAFAM. Les huit principaux opérateurs mondiaux génèrent un bénéfice net combiné de 79 Mds €, bien en deçà des 293 Mds € réalisés par les *hyperscalers*. Ces derniers bénéficient également d'un levier financier plus favorable, avec une dette nette de 40 Mds € et une rentabilité supérieure, contre 95 Mds € de dette nette pour les opérateurs télécoms. Les *hyperscalers* investissent également dans les infrastructures historiques des télécoms, notamment les câbles sous-marins et les réseaux, pour garantir le bon fonctionnement d'Internet. Par ailleurs, des écarts subsistent entre opérateurs américains et européens : des acteurs comme AT&T et Verizon ont historiquement consacré davantage d'investissements à leurs infrastructures, affichant des ratios dette/chiffre d'affaires élevés (120-130 %), contre 60-70 % pour leurs homologues européens, qui misent davantage sur la diversification[19].

[13] *Xerfi, « L'industrie mondiale des équipements télécoms », décembre 2021 et « Les opérateurs Telecom », 2023.*

[14] *Statista Market Insight, « Internet Of Things », Monde- Europe – France, novembre 2023.*

[15] *GrandViewResearch, « Software Market Size », 2023.*

[16] *Gartner Market Statistics, Public Cloud Services, Worldwide, 2020-2026, 3Q22 Update.*

[17] *Gartner, Dépenses IT mondiales, 2023.*

[18] *Ballard, Barclay, « Investment in Digital infrastructure shows no sign of slowing », données Liberium, McKinsey et JPMorgan, juin 2023.*

[19] *Les Annales des Mines, septembre 2024, Laurent Benzoni, Convergence des infrastructures numériques: un point de vue économique.*

Dans ce contexte, deux grandes dynamiques se dessinent. D'une part, les *hyperscalers* adoptent des stratégies de verticalisation sur l'ensemble de la chaîne de valeur, intégrant transport, traitement et accès aux données. D'autre part, les FAI ajustent leur modèle économique, soit en se spécialisant, soit en diversifiant leurs offres, souvent en partenariat avec ces mêmes *hyperscalers*.

Les *hyperscalers* investissent massivement dans leurs propres infrastructures de transport des données, telles que les câbles sous-marins et les réseaux privés, pour deux raisons principales : réduire les coûts à long terme et maîtriser la performance ainsi que la qualité de service. Ces infrastructures, bien que coûteuses à déployer, permettent de diminuer significativement les frais d'exploitation en supprimant les paiements d'interconnexion à des opérateurs tiers, proportionnels au volume de données échangées. En parallèle, elles garantissent une maîtrise totale des performances – latence réduite, bande passante optimisée, meilleure disponibilité –, des critères essentiels pour répondre aux besoins de marchés comme le *cloud* ou le *streaming* vidéo. **Cette stratégie de verticalisation s'inscrit dans une logique plus large de construction d'« écosystèmes complets » où les consommateurs trouvent l'ensemble des services dont ils ont besoin chez un fournisseur unique**. Microsoft, par exemple, associe son *cloud* Azure à des solutions professionnelles comme Dynamics 365 et à des plateformes sociales telles que LinkedIn, consolidant un environnement intégré de services internet avec des barrières à l'entrée plus fortes pour les acteurs locaux.

Pour se positionner sur le marché, les acteurs régionaux ont mis en place des stratégies de différenciation. Ils peuvent développer des solutions hyper-spécialisées adaptées à des secteurs stratégiques tels que la santé, l'aérospatial ou la défense. Ces approches permettent de contourner la concurrence directe avec les *hyperscalers* en offrant des services sur mesure répondant à des besoins spécifiques.

Usages | Aviation | Automobile | Défense | Santé | Agriculture

Transport des données

Fournisseurs d'accès Internet et opérateurs

Acteurs exploitant les infrasturctures réseaux (câbles, antennes, satellites…) et commercialisant des services d'accès Internet aux particuliers et aux entreprises. Inclut les réseaux fixes, mobiles et satellites.

Monde : **1 556 Mds €**

COMCAST · T-Mobile · China Telecom · Starlink

France : **31 Mds €**

orange · SFR · Lyca Mobile

Équipementier

Fournisseurs d'équipements de télécommunication, pas d'acteur français.

Monde : **190 Mds €** | France : **7 Mds €**

Huawei · Cisco · Nokia · Hewlett Packard Enterprise · Lenovo

Constructeurs de réseaux

Inclut le génie civil, l'installation et la maintenance. Expertise des acteurs du BTP.

Monde : **58 Mds €** | France : **9 Mds €**

Bouygues · Axione · SPIE

Traitement des données

Hébergement *Cloud*

Un marché dominé par les GAFAM, qui représentent 71 % du total, et dont la part augmente (67 % en 2020, 71 % en 2022). Les offres IaaS et PaaS sont majoritaires. En France, des offres de *cloud* souverain se développent (Numspot, SENS et Bleu).

Monde : **392 Mds €**

aws · Google Cloud · Azure

France : **10 Mds €**

Scaleway · OVHcloud

Constructeurs de *Datacenters*

Mise à disposition des *datacenters* à l'achat, location, colocation. La France dispose d'un bon niveau d'expertise, d'une énergie verte et de hubs numériques majeurs.

Monde : **216 Mds €** | France : **5 Mds €**

Gestionnaires hébergeurs

Lenovo · Data4 · Equinix · Digital Realty

Équipementiers hors FAI (serveurs…)

Bull · DELL EMC · IBM

Microprocesseurs

Monde : **108 Mds €** | France : **N/A**

Producteurs-Fondeurs

TSMC · UMC · Samsung Foundry

Fournisseurs

NVIDIA · intel · ST · AMD · SIPEARL

Accès aux données

Objets connectés

Comprend les équipements associés (capteurs, routeurs dédiés), plateformes, connectivité (Lora, Sigfox), et les activités aux particuliers et aux entreprises (intégration, maintenance).

Monde : **1 170 Mds €**

Cisco · Dell · Qualcomm

France : **15 Mds €**

Linky · orange Business

Terminaux

Téléphones, ordinateurs et tablettes. Principaux inducteurs du coût carbone et numérique.

Monde : **700 Mds €** | France : **8 Mds €**

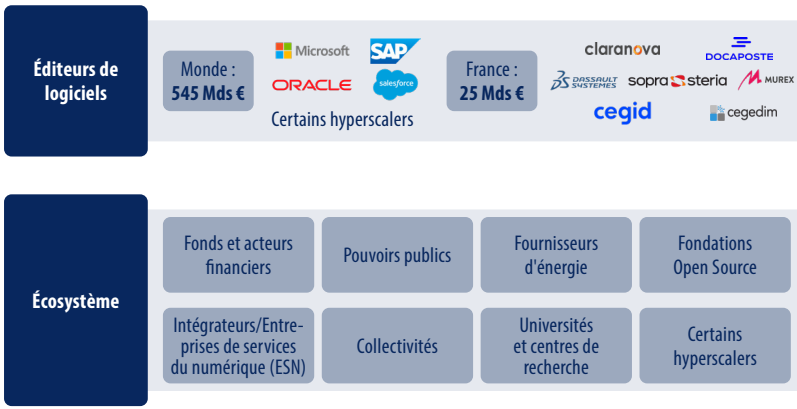
Apple · Samsung · Xiaomi · Lenovo

Informatique industrielle

Inclut les automates, les robots, les cobots, les machines 3 axes et les logiciels associés.

Monde : **140 Mds €** | France : **12 Mds €**

DENSO · Mitsubishi Electric · ABB



1.1. UN MARCHÉ DES INFRASTRUCTURES NUMÉRIQUES EN PROFONDE MUTATION QUI FAIT PESER DES RISQUES IMPORTANTS SUR LES ACTEURS FRANÇAIS ET EUROPÉENS

a. La domination des *hyperscalers* consolide désormais le marché des infrastructures numériques au détriment des acteurs historiques

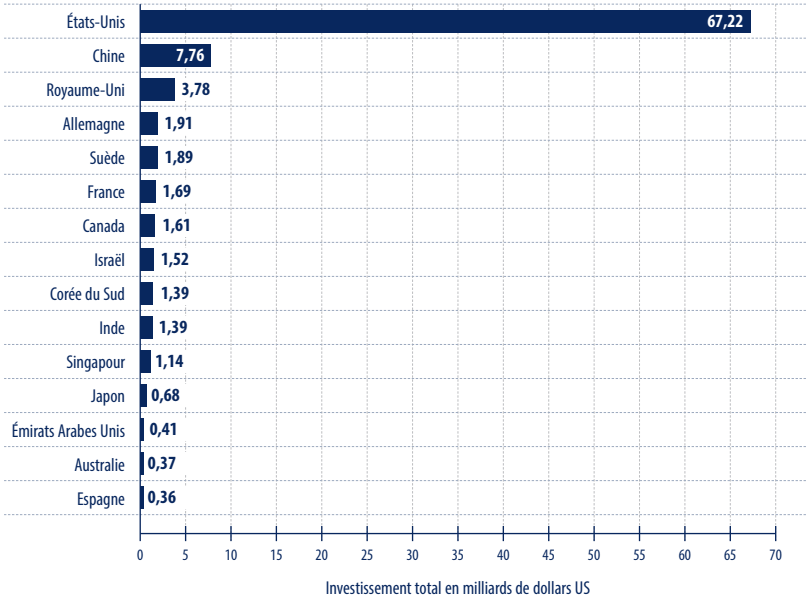
Les cinq entreprises avec les capitalisations les plus élevées au monde sont américaines et dominent le marché du *cloud computing*. Chacune de ces capitalisations dépasse les 2 000 Mds $ en 2024[20], soit l'équivalent de la valeur du PIB de la France en 2023[21]. Les futurs services *cloud* ont vocation à intégrer les modèles d'IA, ce qui s'inscrit dans un contexte marqué par les investissements massifs des *hyperscalers* dans leurs centres de données. À titre d'exemple, Microsoft prévoit

[20] *Au 24 mai 2024, il s'agit de Microsoft (3 064 Mds $), Apple (2 830 Mds $), Nvidia (2 218 Mds $) et Alphabet (2 109 Mds $), suivi de près par Amazon à 1 972 Mds $. LVMH arrive en 20e position à 426 Mds $.*

[21] *Pour l'année 2023, le PIB de la France en valeur était de 2 805,7 Mds € d'après les Comptes Nationaux trimestriels au premier trimestre 2024 de l' Insee.*

d'investir 80 Mds $ en 2025[22]. Par ailleurs, les entreprises américaines ont consacré près de 68 Mds $ à l'IA en 2023, soit 8 fois plus que leurs homologues chinoises et plus de 40 fois plus que les entreprises européennes[23].

Graphique n° 4 • Investissements privés dans l'intelligence artificielle par zones géographiques



Source: Stanford University, Human-Centered Artificial Intelligence, 2024 AI Index Report.

[22] *L'Usine Digitale, J.Bergounhoux, 6 janvier 2025,« Intelligence artificielle : Microsoft va investir 80 milliards de dollars dans ses data centers en 2025 », https://www.usine-digitale.fr/article/intelligence-artificielle-microsoft-va-investir-80-milliards-de-dollars-dans-ses-data-centers-en-2025.N2224910.*

[23] *Stanford University, Human-Centered Artificial Intelligence, 2024 AI Index Report.*

Cette tendance a vocation à s'accélérer dans un contexte où le marché du *cloud computing* devrait bénéficier d'une croissance de 20,4 % en 2024 pour un total de 679 Mds € [24], 70 % de cette croissance étant captée par Amazon Web Services (AWS), Microsoft Azure et Google Cloud. Ces trois acteurs représentent en outre 80 % de la croissance des dépenses en applications et en services de *cloud* public en France en 2021, AWS ayant capté 46 % de cette croissance, Microsoft Azure 17 % et Google Cloud 8 % [25]. **Ainsi, les infrastructures de traitement de données évoluent dans un marché désormais consolidé au profit d'acteurs mondiaux**.

Pour consolider de telles positions, les hyperscalers ont investi dans chaque maillon de la chaîne de valeur des infrastructures numériques pour y construire des avantages concurrentiels :

- Sur le marché des câbles sous-marins, les *hyperscalers*, devenus copropriétaires, puis propriétaires majoritaires, contrôlent aujourd'hui près de 10 % des 559 câbles sous-marins [26] à l'échelle mondiale. Ces entreprises utilisent 70 % de la bande passante de ces infrastructures numériques, et pourraient ainsi devenir propriétaires de près de 50 % de ces câbles.

- Les *data centers* font l'objet d'investissements conséquents, avec des montants atteignant 10 Mds $ [27], comme en témoignent les récents projets d'AWS dans l'Ohio et en Espagne. Microsoft prévoit d'investir 80 Mds $ [28] au cours de l'exercice fiscal 2025 pour développer de

[24] *Cabinet d'études Gartner, 13 Novembre 2023, « Gartner Forecasts Worldwide Public Cloud End-User Spending to Reach $679 Billion in 2024 », https://www.gartner.com/en/newsroom/press-releases/11-13-2023-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-679-billion-in-20240.*

[25] *Autorité de la Concurrence, Avis de juin 2023.*

[26] *TeleGeography, A.Mauldin. 27 juin 2024, « A (Refreshed) List of Content Providers' Submarine Cable ».*

[27] *Ohio Governor Mike DeWine, 16 décembre 2024, « Governor DeWine Announces $10 Billion Investment Plan from Amazon Web Services in Greater Ohio ».*

nouveaux centres de données spécialisés dans l'IA. Parallèlement, les *hyperscalers* investissent directement dans des sources d'énergie pour alimenter leurs infrastructures numériques. Par exemple, Oracle[29] a obtenu les autorisations nécessaires pour construire un centre de données alimenté par des réacteurs nucléaires modulaires (SMR). De même, Google s'est associé à Kairos Power[30] pour installer des SMR, visant une capacité de 500 MW d'ici 2035. Ces initiatives reflètent une tendance croissante des géants technologiques à investir dans des solutions énergétiques innovantes pour répondre à la demande accrue en calcul liée à l'IA. Ainsi, les *hyperscalers* étendent leur stratégie de standardisation à toute la chaîne de valeur, y compris l'approvisionnement en transformateurs. Ces dispositifs, qui ajustent la tension pour optimiser le transport et la distribution d'énergie, pourraient leur permettent de sécuriser leur propre réseau électrique à terme. Si des SMR sont connectés au *grid* et que leurs surplus sont réinjectés sur le marché, cette maîtrise de l'approvisionnement renforce leur avantage concurrentiel face aux *utilities*.

➞ Dans le domaine du calcul, les *hyperscalers* ont sécurisé leur approvisionnement en microprocesseurs grâce à des investissements massifs intégrés à leurs activités de R&D[31]. Par exemple, Meta a consacré 40 Mds $ à la R&D en intelligence artificielle pour l'année 2024, principalement orientés vers la conception de processeurs et le développement d'infrastructures d'entraînement de modèles et d'inférence, en partenariat étroit avec Nvidia[32]. De manière notable, les *hyperscalers* investissent désormais des montants conséquents, de l'ordre

[28] *B.Smith,Vice Chair & President of Microsoft, 3 janvier 2025, « The Golden Opportunity for American AI ».*

[29] *ICT Journal, Y.Chavanne, 17 Septembre 2024, « Oracle se tourne vers le nucléaire modulaire pour ses besoins en calcul IA ».*

[30] *International Data Corporation, A.Salmeron, 21 Novembre 2024, « Can Nuclear Power Fuel Hyperscalers' Energy Transition? ».*

[31] *Le PDG d'AWS déclarait : « Ensemble, nous continuons à innover pour faire d'AWS le meilleur endroit pour exécuter des GPU Nvidia ».*

[32] *Silicon, « Les hyperscalers renforcent leurs recherches et datacenters pour l'IA », avril 2024.*

de plusieurs milliards de dollars, dans le développement de puces personnalisées et propriétaires[33] afin de réduire leur dépendance à des acteurs comme Nvidia et de maîtriser leur chaîne d'approvisionnement. Par exemple, AWS a développé sa propre puce pour l'entraînement de ses modèles, nommée Trainium. Pour l'inférence, Google a conçu ses propres unités de traitement tensoriel (TPU), optimisées pour accélérer les charges de travail d'intelligence artificielle. Microsoft a également introduit deux puces personnalisées : Cobalt, un processeur basé sur l'architecture ARM Neoverse N1, et Maia, un accélérateur d'IA conçu pour exécuter des charges de travail d'IA dans le *cloud*.

- En matière de *cloud*, les *hyperscalers* ont mis en place une stratégie d'intégration des services de leurs partenaires commerciaux à leurs plateformes *cloud*. Les ventes de logiciels d'entreprise *via* des *marketplaces cloud* devraient ainsi augmenter de 40 % par an d'ici à 2028[34], pour une valeur pouvant atteindre 85 Mds $ contre 16 Mds $ en 2023. De manière notable, l'adoption des solutions de type platform-as-a-service (PaaS) devrait accélérer avec un taux de croissance annuel composé (CAGR) de 19,3 %[35] de 2022 à 2028, pour atteindre les 173 Mds $ en 2028. Le PaaS est un modèle de *cloud computing* qui offre une plateforme prête à l'emploi pour développer, déployer et gérer des applications, tout en éliminant la nécessité de gérer ou de posséder l'infrastructure sous-jacente. En d'autres termes, le matériel *(hardware)*, le système d'exploitation (OS), les conteneurs et les environnements, sont déjà inclus et pris en charge par le fournisseur.

[33] *LeMonde informatique, A. Ghoshal, décembre 2024, « Les puces personnalisées se multiplient chez les hyperscalers ».*

[34] *Canalys, août 2024, « Hyperscaler cloud marketplace sales to hit US$85 billion by 2028 ».*

[35] *Grand View Research, « PaaS Market Size, Share & Trends Analysis », 2022.*

- Enfin, les *hyperscalers* concurrencent de plus en plus les acteurs historiques (Denso, ABB) sur le marché de l'IoT, en raison de la « virtualisation » de l'espace numérique, c'est-à-dire la capacité à abstraire les ressources physiques pour les transformer en services logiciels, permettant ainsi une gestion centralisée, une scalabilité accrue et une interopérabilité simplifiée entre les dispositifs connectés et les infrastructures *cloud*, dans un contexte où « *software eats the world*[36] ».

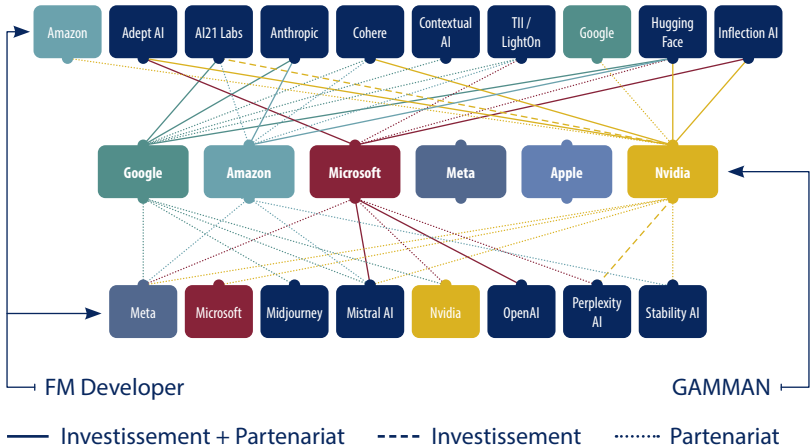
En parallèle d'une stratégie de croissance externe, les *hyperscalers* ont mis en place des partenariats adaptés aux modèles économiques des fournisseurs de modèles de fondation d'IA (*Foundation models* ou FM en anglais). Ces accords consistent à offrir un accès à leur infrastructure *cloud* pour entraîner les modèles, en échange de l'utilisation exclusive de cet environnement par le fournisseur. Pour les développeurs de modèles de fondation d'IA, ce type de partenariat est souvent plus intéressant qu'un investissement direct, car le *cloud* est indispensable pour accéder à la puissance de calcul nécessaire à l'entraînement des modèles, et pour faciliter leur diffusion ensuite[37]. En avril 2024, la *Competition and Market Authority*[38] avait déjà identifié 90 partenariats de ce type.

[36] *The Wall Street Journal, Marc Andreessen, août 2011, « Why Software Is Eating The World ».*

[37] *Autorité de la Concurrence, 28 juin 2024, Avis sur le fonctionnement concurrentiel du secteur de l'intelligence artificielle générative.*

[38] *Competition and Market Authority, Avril 2024, AI Foundation Models Update paper, AI Foundation Models Update paper.*

Graphique n° 5 • Focus sur les partenariats entre *hyperscalers* et modèles de fondation IA



Notes pour le lecteur :

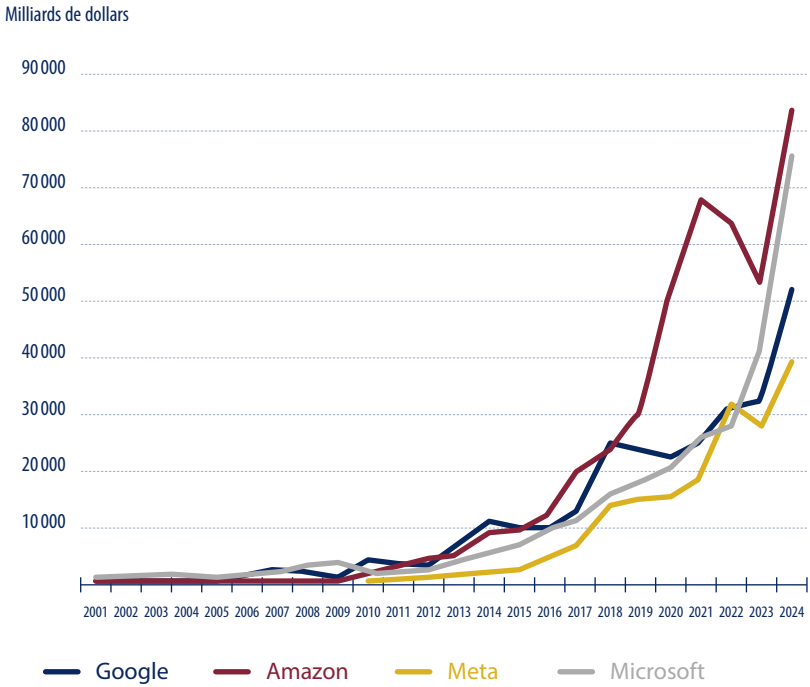
- Meta et Apple envisagent actuellement une alliance dans le domaine de l'IA[39], qui prendrait la forme de l'intégration du modèle d'IA générative de Meta au sein d'Apple Intelligence, le nouveau système d'IA d'Apple destiné à ses appareils.
- La sortie récente du modèle chinois DeepSeek 3, un modèle *open source* comptant plus de 600 milliards de paramètres, basé sur une architecture MoE *(Mixture of Experts)*, qui désigne une approche d'entraînement distribuée permettant d'activer dynamiquement seulement une fraction des paramètres pour chaque tâche. Ce modèle dépasse les performances des modèles fondamentaux américains tout en ayant nécessité beaucoup moins de ressources pour son entraînement : 2028 GPU H800 (une version bridée des H100, limitée notamment en bande passante et puissance de calcul, afin de respecter les restrictions imposées par les exportations américaines). Cela illustre que la disponibilité massive de GPU n'est pas toujours nécessaire, et qu'une optimisation de l'architecture peut significativement réduire les coûts d'entraînement.

[39] *The Wall Street Journal, juin 2024, "Apple, Meta Have Discussed an AI Partnership".*

Ainsi, plus les services des *hyperscalers* sont consommés par les entreprises utilisatrices, plus les données générées par les utilisateurs le sont dans leurs écosystèmes, renforçant ainsi leur consolidation sur l'infrastructure et sur les services associés.

La consolidation des *hyperscalers* tout au long de la chaîne de valeur se fait au détriment des acteurs historiques, incapables de suivre l'ampleur de leurs investissements. Amazon, Google et Microsoft ont, chacun, dépensé en 2024 plus que l'ensemble des trois principaux opérateurs mobiles américains (AT&T, T-Mobile et Verizon) réunis. Amazon et Microsoft affichent les plus hauts niveaux de CAPEX hors Chine, tandis que Google et Meta figurent parmi les sept premières entreprises mondiales en la matière. Ces géants surpassent désormais les investissements des leaders traditionnels des secteurs automobile, énergétique, des semi-conducteurs et des télécommunications. En 2024, Amazon, Google, Meta et Microsoft ont engagé 251 Mds $ en CAPEX, soit une hausse de 62 % par rapport à 2023, creusant encore l'écart avec les acteurs historiques. **Les trois principaux fournisseurs de *cloud* ont dépassé le billion de dollars de CAPEX cumulé depuis le début du siècle – 1 000 Mds $ –, et en intégrant Meta, ce total atteint 1,19 billion $, dont 406 Mds $ investis sur les deux dernières années**.

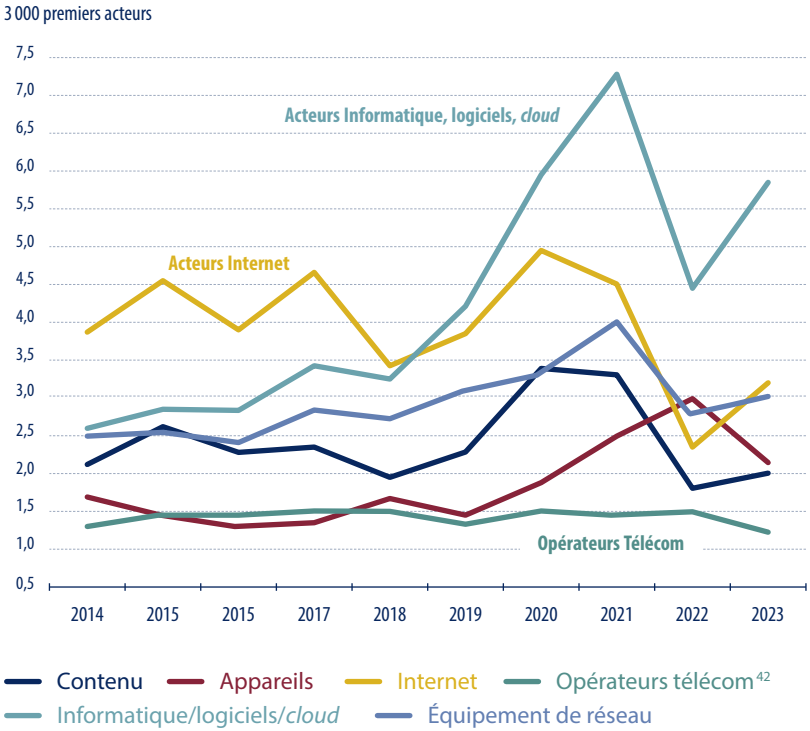
Graphique n° 6 • Dépenses annuelles en capital (CAPEX)



Source : platformonomics.com

Graphique n° 7 • Valorisation des entreprises par rapport à leurs revenus[40]

(Capitalisation boursière / chiffre d'affaires, monde, 2014-2023, en multiples du chiffre d'affaires[41], 500 premiers acteurs par catégorie)



Source : LSEG, Arthur D. Little.

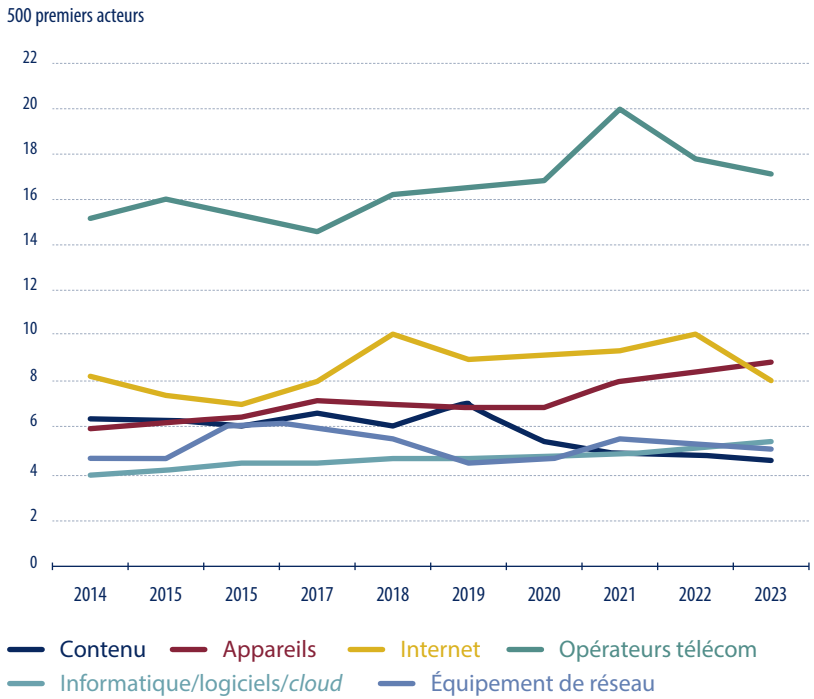
[40] *Panel de 2 855 entreprises. Par secteur, sélection par leur chiffre d'affaires des 500 premières entreprises de plus de 1 million d'euros de chiffre d'affaires en 2023.*

[41] *En euros constants en utilisant le taux de change annuel moyen pour 2023.*

[42] *La capitalisation boursière de T-Mobile US au NASDAQ et le chiffre d'affaires ont été inclus.*

Graphique n° 8 • Intensité des investissements des acteurs de l'écosystème numérique mondial[43]

(CAPEX / chiffre d'affaires, monde, 2014-2023,
en milliards d'euros[44], 500 premiers acteurs par catégorie)



Source : LSEG, Arthur D. Little.

[43] *Panel de 2 855 entreprises. Par secteur, sélection par leur chiffre d'affaires des 500 premières entreprises de plus de 1 M € de chiffre d'affaires en 2023.*

[44] *En euros constants en utilisant le taux de change annuel moyen pour 2023.*

b. Les fournisseurs d'accès à internet français ont opté pour la spécialisation dans les infrastructures dites « actives » et la diversification servicielle

Les FAI dont la taille n'est pas suffisante pour couvrir l'ensemble de la chaîne de valeur des infrastructures numériques peuvent donc opter pour trois stratégies : adapter leur modèle économique, privilégier la croissance organique ou établir des partenariats stratégiques ciblés.

Pour faire face aux investissements massifs nécessaires à l'expansion de la connectivité, les opérateurs historiques ont vu leurs marges se réduire. Pour générer des liquidités, de nombreux FAI ont ouvert leur capital à des investisseurs minoritaires et externalisé leurs infrastructures réseau dites « passives » (pylônes, antennes) à des entreprises spécialisées, les TowerCo[45], ou à des fonds d'investissement. Par exemple, SFR et Bouygues ont récemment cédé 2 000 antennes à l'américain Phoenix Tower International. Parallèlement, de nouveaux acteurs comme Valocîme proposent des loyers plus élevés pour récupérer le contrôle de ces infrastructures. Cette stratégie controversée conduit les TowerCo à déplacer leurs installations, ce qui fragilise la couverture mobile dans certaines zones rurales et augmente les risques de coupures de service.

[45] *Une towerco, pour tower company, est une entreprise qui possède des tours de télécommunication et qui les loue à ses clients (opérateurs, partenaires, etc.) pour améliorer leur réseau télécom. Le terme français est « opérateurs d'infrastructures de télécommunication ».*

Encadré n° 2 • Infrastructures dites « passives » et infrastructures dites « actives »

Les infrastructures dites « passives » représentent les éléments physiques ou structurels nécessaires pour supporter les réseaux, mais qui n'interviennent pas directement dans le traitement ou le transport actif des données. Ce sont essentiellement des ressources matérielles, comme la fibre, les câbles, les pylônes ou encore les baies de brassage, soit un meuble appelé « rack centralisé » où sont regroupés tous les câbles et équipements réseau d'un bâtiment pour les connecter entre eux.

Les infrastructures dites « actives » désignent les équipements électroniques et logiciels qui permettent de gérer, d'acheminer ou de traiter les données circulant sur le réseau. Ces infrastructures rendent le réseau opérationnel. Elles regroupent les équipements réseau (routeurs, commutateurs, multiplexeurs), les antennes relais 4G et 5G, les serveurs informatiques de stockage ou encore les logiciels de supervision réseau.

Les infrastructures passives sont ainsi le socle physique sur lequel reposent les infrastructures actives, lesquelles donnent vie au réseau en assurant le transport et le traitement des informations. Par exemple, une fibre optique (passive) a besoin d'un routeur ou d'un commutateur (actif) pour transmettre les données.

Les infrastructures réseau sont aujourd'hui fragmentées en deux catégories : les infrastructures passives, qui assurent le transport physique des données, et les infrastructures actives, comme les stations de base (BTS, NodeB, eNodeB, gNodeB), qui nécessitent une alimentation électrique pour gérer le trafic. Les FAI conservent un

pouvoir de négociation avec les équipementiers (Huawei, Nokia, Ericsson) pour équiper leurs réseaux, notamment grâce à la standardisation croissante des équipements, qui permet d'utiliser des technologies ouvertes et interopérables, réduisant ainsi leur dépendance à un fournisseur unique.

Les FAI ont aussi stimulé leur croissance organique en s'étendant progressivement sur deux marchés stratégiques : l'IoT industriel et l'IoT grand public.

- **Le déploiement encore limité de la 5G industrielle en France a permis aux FAI de prendre pied sur le marché en pleine croissance de l'IoT industriel**. Ils diversifient ainsi leurs offres en proposant des dispositifs connectés, des plateformes logicielles et des applications adaptées aux besoins des entreprises. Leur implication précoce dans la définition et la standardisation de technologies IoT, telles que LoRaWAN – une innovation française née à Grenoble, aujourd'hui sous contrôle d'entreprises américaines – et NB-IoT, leur confère un avantage stratégique sur ces segments.

- **En parallèle, la vente ou la location de terminaux de type « IoT grand public » représente pour les FAI une porte d'entrée stratégique pour attirer de nouveaux clients vers leurs offres de forfaits mobiles et Internet**. Cette activité a généré des revenus significatifs, atteignant 3,6 Mds € en 2023, soit une part importante des 37,6 Mds € du marché de détail, avec une croissance annuelle de 2,5 %[46]. Certains FAI vont plus loin en développant leurs propres gammes de terminaux : c'est le cas d'Orange avec sa ligne Orange Neva[47], conçue et vendue en exclusivité par l'opérateur. Cela leur permet également de capitaliser sur la montée en puissance de ce

[46] *ARCEP, « Observatoire des marchés des communications électroniques, résultats provisoires 2023 », mai 2024.*

[47] *Orange, gamme Orange Neva, voir : https://boutique.orange.fr/nouveautes/orange-neva/.*

segment, avec des solutions conçues pour répondre aux exigences de différents secteurs. Bouygues Telecom, par exemple, propose des solutions de connectivité IoT sur mesure, incluant des réseaux privés adaptés aux environnements industriels et des services de gestion des objets connectés *via* des plateformes dédiées.

Enfin, pour répondre aux exigences croissantes de diffusion massive de contenu, les FAI ont développé des partenariats stratégiques ciblés avec des fournisseurs de contenu, des *hyperscalers*, et des spécialistes de l'infrastructure réseau.

- Face à la demande de rapidité et de faible latence, les relations traditionnelles de Peering[48] et de Transit[49] **ont progressivement évolué vers des échanges directs au profit des plateformes de contenu. Désormais, pour optimiser la diffusion, les fournisseurs de services installent directement leurs serveurs de cache (CDN)[50] aux points d'interconnexion, intégrant ainsi leur propre bloc d'infrastructures au sein des réseaux des FAI**. Ces accords permettent de réduire la latence pour l'utilisateur final en rapprochant le contenu des lieux de consommation et également de réduire la quantité de données à transporter dans le réseau de l'opérateur.

- Les négociations autour de ces partenariats stratégiques impliquent aujourd'hui des accords commerciaux et techniques complets entre FAI, *hyperscalers* et fournisseurs de contenu. **Ces accords couvrent des aspects variés : localisation des points d'interconnexion, coût, qualité de service, et efficacité des échanges de trafic**. La croissance rapide des plateformes numériques a ainsi amené certains acteurs spécialisés, tels que les transitaires de contenu, à

[48] Accord entre deux opérateurs pour s'échanger directement le trafic de leur réseau respectif.

[49] Configuration dans laquelle l'opérateur est client d'un autre opérateur.

[50] Le cache est un processus de stockage du contenu statique d'un site Web sur plusieurs serveurs du réseau.

adapter leurs offres en hébergeant directement du contenu ou en construisant leurs propres infrastructures de CDN pour maintenir leur positionnement. Ces ajustements permettent aux FAI de coopérer avec les *hyperscalers*, sans tenter de les concurrencer directement.

→ **Parallèlement, les FAI ont diversifié leurs offres en développant de nouveaux services pour leurs clients B2B**. Certains FAI ont tenté de sortir de leur cœur de métier traditionnel, notamment avec des services bancaires (comme Orange Bank) ou des offres de contenu multimédia, telles qu'Orange Content et ses services de vidéo à la demande (VOD). Toutefois, ces initiatives n'ont pas encore généré les résultats escomptés. En revanche, ils ont obtenu de meilleurs résultats en se spécialisant sur des marchés de niche, comme la fourniture de services dits « *managés* » – soit des services comprenant la gestion externalisée des infrastructures réseau et des solutions clés en main adaptées aux besoins spécifiques des entreprises – ou des solutions de *cloud* hybride, qui combinent des environnements *cloud* publics et privés. Des entreprises comme Gigalis, par exemple, proposent ce type de services, en s'appuyant sur une expertise locale et des partenariats étroits avec les collectivités territoriales. Ces collaborations permettent de répondre aux besoins spécifiques des acteurs régionaux, notamment dans les domaines de la gestion de données sensibles ou des réseaux critiques. Les solutions régionales, telles que celles proposées par le réseau fibre Altitude Infra ou les initiatives d'Axione, se développent également en misant sur des services personnalisés, une proximité accrue avec les utilisateurs et une prise en charge des exigences réglementaires locales.

1.2. FACE AUX NOUVEAUX RISQUES INDUITS POUR LES ACTEURS, LA NÉCESSITÉ DE CHOISIR LES USAGES D'INFRASTRUCTURES NUMÉRIQUES À MAÎTRISER SUR LE TERRITOIRE FRANÇAIS

Les infrastructures numériques françaises ont été développées depuis le milieu du XIXe de manière incrémentale, avec des ajustements successifs au fil de l'eau pour renforcer notre maîtrise technologique[51]. Cette approche par « solutions temporaires » trahit toutefois une absence de vision stratégique qui soit véritablement ancrée dans la maîtrise des aspects techniques et l'adaptation au rythme de l'innovation en matière de nouvelles technologies.

Conscient de cet écueil, le Gouvernement, dans sa feuille de route sur la décennie numérique 2024-2030[52], propose la construction d'infrastructures numériques « durables, sûres et efficaces » en se concentrant sur deux volets : le déploiement d'infrastructures de télécommunications et le renforcement de l'autonomie stratégique européenne.

Concernant le déploiement d'infrastructures de télécommunications, la stratégie s'inscrit dans la continuité des politiques volontaristes – dont l'efficacité fait débat – qui ont été mises en place dès les années 2000 avec les délégations de service public, puis en 2013 avec le plan Très Haut Débit, puis en 2018 avec le *New Deal Mobile*, en y ajoutant un volet sur les câbles sous-marins. En effet, ses trois axes prioritaires se concentrent sur la finalisation du plan Très Haut Débit de 2013 avec l'objectif de généraliser la fibre pour 2025, d'améliorer la qualité de la couverture mobile sur le territoire dans la lignée du *New Deal Mobile* de 2018 et de positionner la France comme un « hub d'interconnexion » notamment pour déployer les câbles sous-marins.

[51] *Architecture informatique capable de réaliser au moins un milliard de milliards (10^{18}) d'opérations par seconde, soit un exaflop. Les supercalculateurs exaflopiques, comme Frontier ou Aurora, incarnent cette puissance, permettant de traiter des simulations complexes à une échelle inédite.*

[52] *Direction générale des Entreprises, Mars 2024, Feuille de route du numérique , La décennie numérique 2024-2030.*

Concernant le renforcement de l'autonomie stratégique européenne, quatre axes stratégiques ont été retenus :

1. La poursuite de la stratégie nationale quantique lancée en 2021 avec l'objectif clé de fournir à l'UE son premier ordinateur quantique tolérant aux erreurs doté de 128 qubits logiques[53] en 2025 et le lancement d'un grand défi pour intégrer des accélérateurs quantiques de première génération (NISQ)[54] dans des supercalculateurs existants ;
2. Le développement des capacités de production et d'industrialisation des technologies électroniques pour augmenter nos capacités de production de 90 % d'ici 2027 et avoir une production européenne de semi-conducteurs équivalente à 20 % de la production mondiale ;
3. Le déploiement de 1 000 nœuds périphériques en France, sur les 10 000 nœuds prévus en Europe[55], pour accompagner le *edge computing* dans le cadre d'un Projet Important d'Intérêt Européen Commun (PIIEC)[56] avec l'Allemagne, la Hongrie, l'Italie, les Pays-Bas, la Pologne et l'Espagne, en plus des PIIEC existants sur le *cloud* ;

[53] *Un qubit physique est la plus petite unité d'information d'un système quantique, capable d'être dans une superposition, c'est-à-dire simultanément dans les états |0⟩ et |1⟩ jusqu'à ce qu'une mesure soit effectuée. Un qubit logique regroupe plusieurs qubits physiques pour corriger les erreurs et garantir la fiabilité du calcul quantique. Le programme « Proqcima » de la direction générale de l'Armement (DGA) a pour objectif de faire émerger des ordinateurs quantiques constitués de 128 qubis logiques d'ici à 2032.*

[54] *Dispositifs hybrides exploitant les algorithmes quantiques pour compléter les calculs classiques. Conçus pour fonctionner en tandem avec des supercalculateurs, ils traitent des sous-problèmes spécifiques comme l'optimisation, la simulation moléculaire ou le calcul du wake effect – énergie perdue à cause des turbulences générées par une éolienne sur les autres dans un parc, en lien avec sa position et l'impact sur le flux global – où les algorithmes quantiques offrent un avantage potentiel.*

[55] *Commission européenne, 5 décembre 2023, « Commission approves up to €1.2 billion of State aid by seven Member States for an Important Project of Common European Interest in cloud and edge computing technologies ».*

[56] *Un PIIEC est un cadre défini par l'Union européenne permettant aux États membres de soutenir financièrement des projets innovants d'envergure, jugés stratégiques pour la compétitivité et la souveraineté économique de l'Europe. Ces projets, souvent dans des secteurs clés comme les batteries, la microélectronique ou l'hydrogène, impliquent des collaborations transfrontalières entre entreprises et organismes de recherche, tout en respectant des règles spécifiques d'aide d'État pour éviter les distorsions de concurrence.*

4. La mise en place d'une nouvelle phase de la stratégie de cybersécurité annoncée en 2021, qui serait cette fois centrée sur les compétences nécessaires pour affronter les nouvelles menaces et le développement de technologies évolutives et de coopérations renforcées en conséquence.

En parallèle, la stratégie nationale sur l'intelligence artificielle s'est renforcée avec son volet 3, prévoyant un financement de 109 Mds € pour le développement de centres de données dédiés à l'IA en France. Ces investissements, majoritairement portés par des acteurs privés, incluent aussi des partenariats public-privé, comme le campus IA financé à hauteur de 50 Mds € par les Émirats Arabes Unis. Désormais, le déploiement de réseaux mondiaux *via* ces partenariats devient un levier stratégique pour rester compétitif dans la course à l'IA. **À l'échelle européenne, 200 Mds € supplémentaires viennent compléter cet effort, marquant une rupture avec l'attentisme technologique habituel et dotant pour la première fois l'Europe des moyens de ses ambitions**. L'IA semble enfin être perçue comme un moteur structurant des usages futurs, nécessitant un redimensionnement anticipé des infrastructures numériques. Son développement impose une révision en profondeur des capacités en réseau, stockage et traitement pour répondre aux exigences des technologies émergentes et de leurs cas d'usage. **En adoptant une approche proactive sur le dimensionnement des infrastructures, les pouvoirs publics ne se contentent plus d'ajustements réactifs face aux évolutions technologiques. Ils intègrent l'infrastructure comme un élément central de la transformation numérique, abandonnant la logique d'adaptation à la marge pour une stratégie capable d'anticiper et d'orienter les ruptures technologiques**.

a. Proposer des solutions locales de type IaaS : un impératif pour réduire les dépendance aux grands acteurs du *cloud* sur les usages les plus sensibles

Les annonces récentes visent à réduire la dépendance aux solutions PaaS des géants américains pour les usages sensibles de l'IA. **Aujourd'hui, l'Europe ne détient que 18 % des centres de données mondiaux, et moins de 5 % appartiennent à des entreprises européennes**. Ce déficit d'infrastructures freine déjà l'écosystème IA, comme l'a souligné Mistral. Les PaaS, tels que Google Vertex AI ou AWS SageMaker, offrent une infrastructure clé-en-main pour entraîner et déployer des modèles d'IA, mais au prix d'une perte de contrôle. En effet, en sous-traitant l'hébergement et l'optimisation des modèles aux *hyperscalers*, les entreprises européennes ont progressivement renoncé à une part essentielle de leur autonomie technologique et financière. **Si des acteurs comme Mistral ou Hugging Face ont su développer des modèles performants en s'appuyant sur ces services, cela pose un problème majeur pour les secteurs sensibles, comme la défense, où des entreprises comme Helsing privilégient des infrastructures internes**.

Les nouveaux centres de données annoncés par l'UE et la France au Sommet de l'IA pourraient bien combler cette lacune à condition d'être orientés vers les usages les plus sensibles que nous devons collectivement maîtriser, et donc de s'inscrire dans une stratégie cohérente. En proposant des solutions locales d'*Infrastructure-as-a-Service* (IaaS), ces infrastructures offriront une alternative aux plateformes *cloud* étrangères, renforçant la souveraineté numérique et permettant aux entreprises d'optimiser leurs coûts en adaptant leurs infrastructures à leurs besoins réels.

Ces solutions de type IaaS doivent prendre en compte le phénomène de « convergence » entre les réseaux, le *cloud*, l'*edge computing* et l'IoT. Cette interconnexion permet aux technologies de fonctionner ensemble de manière complémentaire. Les technologies

IoT collectent les données en temps réel à l'aide de capteurs. L'*edge computing* traite immédiatement les données urgentes ou nécessitant peu de puissance de calcul, comme la détection d'une anomalie sur une machine. Le *cloud*, quant à lui, prend en charge l'analyse et le stockage des données moins prioritaires ou nécessitant une forte capacité de calcul, par exemple pour prévoir des pannes ou optimiser l'efficacité globale. **Ce modèle redéfinit non seulement les architectures techniques, mais aussi les services proposés et les rôles des différents acteurs du numérique**.

Le *cloud computing* s'est imposé grâce à des technologies performantes adaptées au passage à l'échelle, notamment la virtualisation et la softwarisation des données[57], comme l'illustrait déjà un livre blanc publié par Syntec en 2012[58]. Historiquement centré sur de grands *data centers* abritant des millions de serveurs, le *cloud* a répondu à des besoins massifs de stockage et de calcul en offrant une infrastructure de traitement « élastique », capable d'évoluer en fonction de la charge de travail. Ce modèle, dans lequel le client paie à l'usage en substituant des OPEX à des CAPEX, a transformé la gestion des ressources informatiques. Cependant, l'augmentation des objets connectés et l'évolution des usages ont révélé ses limites. L'émergence de l'*edge computing*, au début des années 2010, a complété le *cloud computing* en déplaçant le traitement des données à la périphérie des réseaux. Cette approche répond à des besoins de latence réduite, d'optimisation de la bande passante, de résilience accrue et de maîtrise des coûts. **La convergence entre le cloud et les réseaux repose désormais sur des fonctions déployées virtuellement sur des équipements numériques polyvalents. Un serveur peut ainsi héberger simultanément des fonctions cloud classiques et des services réseau, comme la sécurité**

[57] *Ces modèles consistent à utiliser des logiciels pour rendre les ressources physiques (comme les serveurs ou les réseaux) accessibles et gérables de manière flexible, en les transformant en services numériques indépendants de leur infrastructure matérielle.*

[58] *Syntec Informatique, 2012, Le livre blanc du cloud computing : tout ce que vous devez savoir sur l'informatique dans le nuage.*

ou la gestion des flux, redéfinissant les rôles des infrastructures à la périphérie. Cette hybridation a également transformé les modèles de sécurité, les réseaux devenant des plateformes capables d'intégrer directement des services de protection dans les systèmes eux-mêmes.

Historiquement, Internet reposait sur une architecture en couches distinctes – par exemple, des réseaux dédiés à la transmission des données, d'autres au calcul ou au stockage. **Aujourd'hui, cette organisation hiérarchique est remplacée par des systèmes où les fonctions sont intégrées et les capacités redistribuées grâce à la « virtualisation », qui désigne le processus consistant à abstraire les ressources physiques, comme les serveurs, le stockage ou les réseaux, pour les transformer en ressources logicielles accessibles à la demande, indépendamment de leur emplacement ou de leur matériel sous-jacent**.

Ce que l'on appelle la « fin des couches » désigne cette évolution où un même acteur peut, s'il le souhaite, désormais gérer plusieurs fonctions critiques simultanément : calcul, intelligence artificielle (IA), *cloud* et connectivité. Cette convergence se reflète également dans l'émergence de services multisectoriels. Les infrastructures numériques ne se contentent plus de connecter des points : elles peuvent désormais intègrer des fonctions supplémentaires, telles que la capacité à observer et analyser l'environnement. Par exemple, des antennes qui, en plus de transmettre des données, agissent comme des radars pour détecter leur environnement immédiat, ce qui ouvre la voie à des innovations radicales. Une voiture intelligente, par exemple, pourrait non seulement optimiser son trajet en temps réel, mais aussi interagir avec une ville connectée, collaborer avec des fournisseurs d'énergie pour accéder à des sources peu coûteuses, et se recharger grâce à de l'énergie solaire. **Ces avancées bouleversent la structuration classique des infrastructures, créant de nouveaux défis en matière d'interopérabilité, de gestion des données et de gouvernance**. La 5G illustre déjà cette tendance : bien plus qu'une simple évolution de

la 4G, elle introduit des innovations comme le *slicing* réseau[59] et l'intégration native de services *cloud*[60].

Ainsi, dans une logique « d'infrastructure *as a service* », les utilisateurs peuvent choisir de payer à la consommation plutôt que d'acquérir des systèmes propriétaires. **Cette évolution impose une forte présomption de valeur sur l'infrastructure elle-même, au-delà des technologies sous-jacentes comme la virtualisation ou la *softwarisation***.

b. La nécessaire définition des besoins souverains à soutenir et leur articulation avec les hubs industriels existants

Les offres d'infrastructures numériques de type IaaS devront être conçues de concert avec les métiers qu'elles desservent, qu'il s'agisse de la santé, de l'énergie ou de l'industrie, et en cohérence avec les usages souverains qu'elles entendent permettre de maîtriser. Le succès de leur mise en œuvre dépendra en grande partie de la collaboration avec les acteurs locaux pour intégrer les besoins spécifiques des différents secteurs, dans un contexte où les réseaux sont appelés à devenir plus ouverts. Grâce aux interfaces mises à disposition, des tiers pourront contrôler et exploiter une partie de l'infrastructure. En effet, les infrastructures numériques sont désormais des plateformes de services intégrés, avec un fort potentiel de création de valeur en permettant à divers acteurs d'ajouter leurs propres fonctionnalités directement dans les infrastructures partagées. Il sera aussi conditionné à leur articulation avec les hubs industriels existants, et notamment les câbles sous-marins qui sont en capacité d'alimenter les clusters régionaux en

[59] *Le slicing réseau consiste à diviser un réseau physique en plusieurs segments virtuels indépendants, optimisés pour des usages spécifiques (comme l'IoT, la 5G industrielle ou le gaming), chacun disposant de ses propres ressources et niveaux de service.*

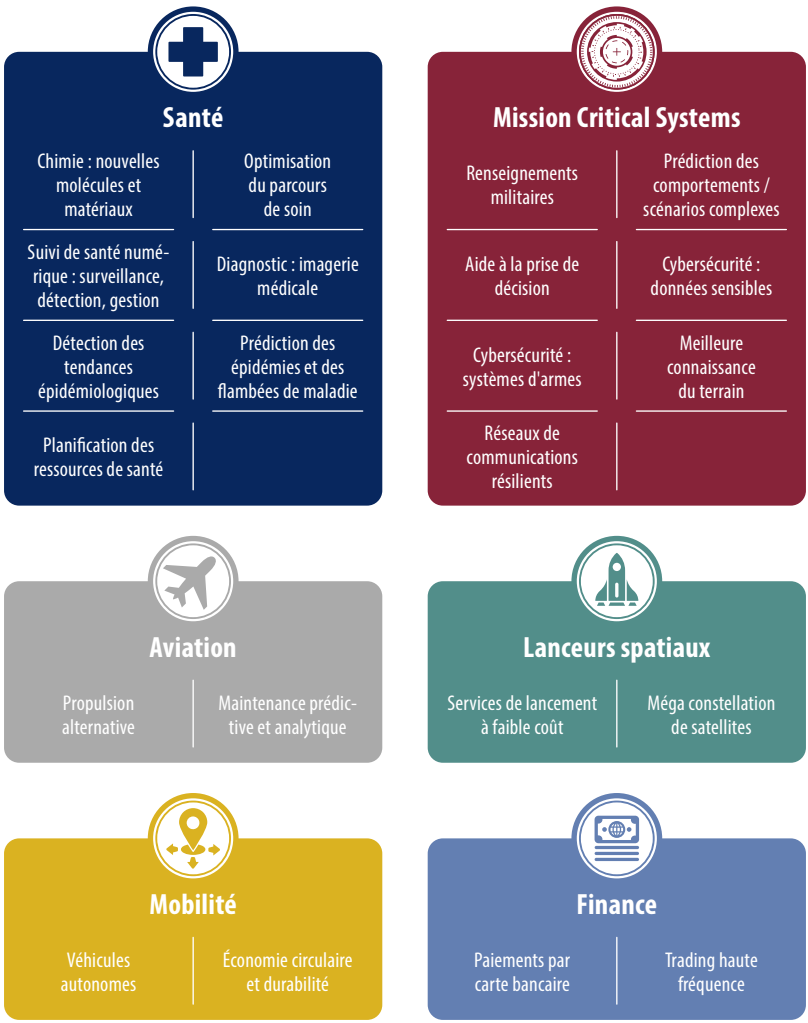
[60] *L'intégration native de services cloud désigne l'intégration directe des fonctionnalités cloud (stockage, calcul, sécurité) dans les infrastructures réseau, permettant aux services de fonctionner de manière fluide et optimisée sans intermédiaires supplémentaires.*

connectivité haute performance et en intégrant des solutions de *slicing* réseau pour attribuer des capacités dédiées aux secteurs verticaux tels que la santé, la logistique ou la finance.

L'autonomie technologique, bien que coûteuse, constitue un impératif stratégique pour la France. Investir dans une infrastructure européenne capable de justifier les coûts associés à cette autonomie technologique et stratégique est essentiel. **Cette démarche doit partir des usages prioritaires que l'Europe souhaite maîtriser pleinement**. Elle nécessite également de favoriser l'émergence d'initiatives locales et régionales pour développer des alternatives crédibles aux technologies étrangères, tout en positionnant l'Europe comme un acteur majeur de la souveraineté numérique à l'échelle régionale.

L'Institut Montaigne propose une catégorisation d'usages prioritaires pour lesquels déployer cette infrastructure en trois axes : (i) les domaines dans lesquels l'Europe dispose d'un savoir-faire reconnu à l'échelle mondiale ; (ii) les secteurs nécessitant une autonomie stratégique, que ce soit au niveau national ou européen ; et (iii) les champs où une accélération est impérative pour consolider ou renforcer un avantage compétitif.

Graphique n° 9 • Cas d'usage critiques



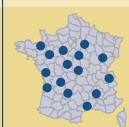
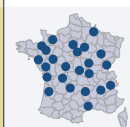
Les *data centers edge* complètent les *data centers cloud* et l'IoT en offrant des capacités accrues de traitement et de stockage plus proches des utilisateurs. Deux principales configurations existent : les *Metro Edge*, situés dans des zones métropolitaines pour desservir un grand nombre d'utilisateurs dans une région, et les *Last Mile Edge*, encore plus proches des utilisateurs finaux, souvent dans des environnements semi-urbains ou ruraux. Contrairement aux dispositifs embarqués, ces *data centers* offrent une redondance et des capacités d'interopérabilité qui permettent de gérer des applications critiques à plus grande échelle.

Graphique n° 10 • Technologies et infrastructures du *edge computing*

Le *edge computing* doit se comprendre comme une architecture en plusieurs niveaux et complémentaire au *cloud computing*. Au plus proche de l'utilisateur, les capteurs embarqués traitent rapidement la donnée des tâches simples (exemple : caméra locale). Ensuite, les armoires à rue prennent en charge des besoins spécifiques à petite échelle avec une faible latence (exemple : des feux de circulation intelligents). En parallèle, les *data centers edge*, qu'ils soient de type *Metro Edge* ou *Last Mile Edge*, assurent la prise en charge de tâches plus complexes (exemple : véhicules autonomes). Au dernier niveau, les *data centers cloud* gèrent les données consolidées pour le stockage à long terme et l'analyse approfondie (exemple : VOD). **Ces différents niveaux ont vocation à s'interconnecter plus étroitement et à fonctionner de concert**. Malgré leur décentralisation, les infrastructures de type « *edge* » requièrent une gestion et une administration rigoureuses, identiques à celles des serveurs traditionnels. Administrer un « parc *edge* » implique la mise en place de systèmes de gestion avancés pour assurer la maintenance, la sécurité et la coordination des équipements, tout comme dans un environnement centralisé.

Tableau n° 1 • Infrastructure du réseau de *datacenter*

Au plus loin de l'utilisateur → Au plus proche de l'utilisateur

	Datacenter Hyperscale (« Pôle Cœur »)	***Datacenter Metro Scale*** (« Opérateur Hôte »)	***Datacenter Metro Edge*** (« Sous marché »)	***Last Mile Edge*** (« Nœuds locaux »)	***On premise* / embarqué**
Type d'applications	• Stockage des données non sensibles à la latence. • Contenu statique. • Analyse de données et *machine learning*.	• Interconnexion et appairage.	• Diffusion de contenu et mise en cache. • Réseaux sociaux. • Jeux vidéos, réalités augmentée et virtuelle.	• IaaS et Bare Metals.[xx]	• Traitement sur place (applications sensibles à la latence et sécurisées). • Avant-postes AWS, GCP ou Azure.
Exemples d'applications	• Recherche en IA et *deep learning*.	• *Salesforce*. • API (ex. Twillo).	• Twitch (Amazon). • Oculus (Meta).	• StackEdge (Azure). • Publicité à faible latence et en temps réel.	• *Computing* embarqué dans les voitures autonomes et les drônes.
Capacités de puissance	• > 10 MW	• 3 MW - 10 MW	• 200 KW - 3 MW	• < 200 KW	• Embarqué
Marchés desservis (illustratif)	 Couverture nationale *via* les principaux nœuds régionaux.	 Déploiement unique pour desservir les principaux marchés.	 Sites multiples autour d'un marché principal pour répondre à la demande des sous-marchés.	 Point d'agrégation ultra local (ex. nœud câble, pôle d'agrégation mobile à mi-chemin).	 Capacité et *computing* sur site ou embarqué.

Source : EY, « Quelles sont les tendances autour des datacenters EDGE en France ? », mai 2023.

Le *edge computing* désigne ainsi le traitement des données au plus près de leur source, permettant un accès rapide et localisé directement « à la périphérie » du réseau. Cette approche, apparue dans les années 2010 avec l'essor des technologies IoT, répond à la nécessité de gérer efficacement les volumes croissants de données générées par ces dispositifs.

En 2031, plus de la moitié des données seront traitées directement en périphérie du réseau, soit en mode « *edge* ». Selon la Fondation Linux[61], d'ici 2025, le *edge computing* devrait surpasser le *cloud computing*, représentant quatre fois plus d'activités de traitement et générant 75 % des données mondiales. Aujourd'hui, les points de calcul et de stockage se multiplient pour rapprocher les données de l'utilisateur, en phase avec une explosion des volumes : 181 zettaoctets (Zo) attendus en 2025, contre 64 Zo en 2020 et seulement 2 Zo en 2010[62]. Parallèlement, nos capacités de stockage progressent, passant de 6,7 Zo en 2020 à 17 Zo en 2025. À l'échelle des organisations, le volume moyen de données a doublé tous les deux ans, atteignant 3 pétaoctets (Po)[63] par entité en moyenne, soulignant l'importance d'infrastructures capables de traiter ces volumes localement, avec rapidité et fiabilité.

Le développement d'infrastructures de traitement en mode *edge* constitue une opportunité stratégique majeure pour assurer la souveraineté des données européennes les plus sensibles. **Bien que la performance du *edge computing* repose sur une intégration fluide avec des solutions de *cloud computing*, c'est un segment de marché sur lequel la domination américaine ou chinoise n'est pas encore actée, offrant à l'Europe une réelle chance de développer des solutions locales compétitives. Cette technologie repose sur des infrastructures de proximité, permettant un traitement décentralisé des données. Elle réduit ainsi la dépendance aux services *cloud* étrangers pour les tâches critiques et favorise une gestion stratégique des données, notamment celles qui transitent au-delà des frontières européennes**. Le *edge computing* s'étend à des sous-segments clés où l'Europe excelle déjà, comme les plateformes *middleware* à faible

[61] *A. Joshipura, directeur général des réseaux au sein de la Fondation Linux (2019). Discours lors de l'Open Networking Summit.*

[62] *IDC, Seagate et Statista, « Le Big Bang du Big Data », mars 2021.*

[63] *Étude MEGA International et Entreprise Strategy Group, « Le rôle stratégique de la data Gouvernance et son évolution », octobre 2022.*

consommation énergétique ou les solutions de type *software-as-a-service* axées sur les données. Ces capacités, combinées à une expertise dans des secteurs stratégiques tels que les *smart grids*, la mobilité intelligente, et l'agriculture de précision, renforcent la compétitivité des entreprises européennes.

En s'appuyant sur l'infrastructure 5G portée par des leaders européens comme Nokia et Ericsson, l'Europe a l'opportunité de développer des solutions sur mesure adaptées aux besoins locaux et des infrastructures robustes.

Le principal frein pour faire émerger cette offre réside néanmoins dans la disponibilité limitée de données de qualité au plus proche de l'utilisateur.

Les données concernées sont de différentes natures :

- **Des données en temps réel issues de capteurs comme la température, l'humidité, les mouvements, les signaux acoustiques, etc., utilisées pour des applications comme l'agriculture de précision, la gestion des bâtiments intelligents ou les chaînes logistiques**. Dans des environnements industriels, cela se réfère surtout à des informations collectées dans les usines connectées pour surveiller les performances des machines ou anticiper les pannes (maintenance prédictive).
- **Des données sensibles ou critiques comme les données de santé** (images médicales de type IRM, scanners ou données de surveillance en temps réel, comme les moniteurs cardiaques, utilisés dans des hôpitaux ou cliniques connectés) ou de sécurité (vidéos de surveillance ou flux provenant de dispositifs de sécurité, comme les caméras intelligentes, pour détecter les anomalies ou renforcer la protection des lieux).
- **Des données liées à la mobilité des utilisateurs, comme les données des véhicules connectés** (informations sur la position, la

vitesse ou les performances des systèmes de conduite autonome. Ces données doivent être traitées localement pour réduire les délais critiques) **ou les capteurs dans les « villes intelligentes » utilisés pour gérer les feux de circulation et optimiser le trafic en temps réel**.

- **Des données météorologiques** collectées par des capteurs pour des applications comme la gestion des catastrophes naturelles ou la surveillance de la qualité de l'air.
- **Des données utilisateurs générées localement pour des biens et services de consommation**, comme par exemple les données nécessaires pour des expériences de réalité augmentée ou virtuelle, qui exigent une latence quasi nulle.

Le *edge computing* nécessite par conséquent une gestion répartie sur de nombreux points de traitement locaux. Chaque point de calcul doit être configuré et maintenu de manière indépendante, ce qui complique la gestion globale et la coordination des ressources. En effet, cette architecture décentralisée exige une surveillance constante des performances, la mise à jour régulière des logiciels et la gestion des pannes ou interruptions, souvent dans des environnements hétérogènes. Par exemple, dans une ville intelligente, des capteurs déployés aux intersections pour optimiser le trafic doivent fonctionner de manière autonome tout en étant synchronisés avec les autres systèmes urbains. Si l'un des capteurs tombe en panne ou devient obsolète, cela peut perturber l'ensemble du réseau, générant des ralentissements et des congestions imprévues. Cette complexité logistique, amplifiée par la nécessité de sécuriser chaque point contre les cyberattaques ou les intrusions physiques, peut augmenter les coûts opérationnels et ralentir le déploiement à grande échelle, particulièrement dans des secteurs critiques comme la santé ou la mobilité.

Plusieurs obstacles doivent être surmontés pour déployer efficacement des infrastructures de type Cloud-Edge-Iot sur le territoire européen :

- **Des verrous techniques :** les données à traiter localement doivent être adaptées à une gestion décentralisée. Cela implique qu'elles ne nécessitent ni synchronisation ni traitement préalable au niveau central, tout en étant stockées dans des formats standardisés afin de garantir leur accessibilité par les utilisateurs.
- **Des défis organisationnels :** une gestion proactive des données locales et des équipements physiques géographiquement répartis est indispensable pour éviter la surcharge des systèmes, ce qui repose sur des plateformes dédiées capables de prioriser les données les plus pertinentes pour le niveau local, qui, pour le moment, n'existent pas.
- **Des contraintes réglementaires :** certaines procédures d'autorisation peuvent empêcher un accès en temps réel aux données, ce qui réduit la performance des activités industrielles pour lesquelles ces dernières sont utilisées.
- **Problèmes liés aux usages :** des données locales mal structurées (mauvaise indexation ou formats inadéquats) ou non mises à jour en temps réel limitent la fiabilité des solutions et a fortiori l'adoption du *edge*.

Encadré n° 3 • Focus sur le *edge mesh*, une approche en devenir

Le *edge mesh*, un réseau décentralisé d'appareils interconnectés en périphérie, est un levier de compétitivité pour les ETI, PME et TPE en répondant aux enjeux de sobriété et de sécurité. Chaque nœud (appareil, serveur ou logiciel) peut produire, consommer et relayer des données, ce qui permet :

- Réduction des coûts : Le traitement local diminue les transferts et l'usage du *cloud*.

- Latence minimale : Les données sont traitées près de l'utilisateur, crucial pour des applications en temps réel (industrie 4.0, domotique, véhicules autonomes).
- Résilience accrue : L'absence de point de défaillance unique renforce la sécurité contre les cyberattaques.
- Flexibilité : Les nœuds peuvent être ajoutés ou supprimés selon les besoins.

La France accuse déjà un retard significatif dans le déploiement des infrastructures de traitement de type « *edge* ». Des acteurs tels qu'Amazon Web Services (AWS), Microsoft Azure et Google Cloud ont déjà implanté des centaines de nœuds périphériques à travers le monde, étendant continuellement leurs réseaux. Par exemple, AWS propose les Local Zones, des infrastructures de calcul décentralisées déployées dans plusieurs villes américaines, rapprochant les services des utilisateurs finaux et réduisant ainsi la latence pour des applications critiques comme les jeux en ligne ou les simulations industrielles. En Asie, des pays comme la Chine et le Japon ont massivement investi dans l'infrastructure *edge*. En Chine, des entreprises telles qu'Alibaba et Huawei ont déployé des milliers de nœuds périphériques pour répondre aux besoins de l'industrie locale et soutenir des projets nationaux tels que les villes intelligentes ou les réseaux IoT. Ces nœuds sont intégrés dans des initiatives stratégiques, contribuant à une infrastructure numérique robuste et adaptée aux besoins futurs.

En comparaison, les efforts européens restent modestes. Certaines initiatives, comme la plateforme BoostAeroSpace de Dassault, créée en 2011 en collaboration avec Airbus, Safran et Thales, visent à optimiser la compétitivité de l'industrie aérospatiale européenne en fournissant des services standardisés et interopérables à tous les acteurs de la chaîne logistique du secteur. De même, Docaposte, filiale numérique du

groupe La Poste, a développé des solutions de confiance numérique et a récemment lancé une offre souveraine d'intelligence artificielle générative en partenariat avec des acteurs français. **Cependant, ces initiatives demeurent isolées et manquent d'une coordination à l'échelle européenne pour le déploiement généralisé**.

Encadré n° 4 • L'exemple de la *Digital Public Infrastructure* (DPI), indienne, un choix de souveraineté technologique

La DPI indienne repose sur une plateforme inclusive et sécurisée, facilitant l'accès aux services publics et privés. Elle intègre des systèmes tels qu'Aadhaar pour l'identité numérique, UPI pour les paiements, et DigiLocker pour le stockage sécurisé de documents. Ces éléments offrent une infrastructure interopérable et évolutive, conçue pour réduire la fracture numérique et soutenir la croissance économique, notamment pour les populations rurales et marginalisées. La DPI accorde aussi une grande importance à la cybersécurité et à la protection des données, en s'appuyant sur des réglementations strictes et des technologies avancées pour garantir la confidentialité des informations citoyennes.

L'essentiel de la DPI est *open source*, avec un contrôle étroit de l'État indien sur des services intégrés ainsi qu'une plateforme de commerce numérique, *Open Network for Digital Commerce*, offrant une alternative à Amazon. La *Data Empowerment* and *Protection Architecture* (DEPA) structure la protection des données personnelles au cœur de l'infrastructure numérique, grâce au *consent manager*, qui centralise la gestion des autorisations entre utilisateurs et fournisseurs.

La DPI s'est également affirmée comme un levier diplomatique pour l'Inde, s'étendant à plus de 10 pays pour renforcer ses liens avec la diaspora, proposer une alternative aux réseaux de paiement comme Visa et Mastercard, et approfondir les relations avec des voisins tels que le Népal et Singapour.

c. La question des usages souligne aussi la nécessité d'anticiper des investissements structurants pour l'avenir

L'effort collectif pour généraliser la fibre en France met en lumière un paradoxe : pourquoi cet objectif politique fort ne s'applique-t-il pas à une autre priorité, celle de dimensionner nos infrastructures numériques de traitement de données pour répondre aux usages de demain ? L'IA, l'*edge computing* ou l'IoT, par exemple, imposent des infrastructures capables de gérer des volumes de données massifs, de répondre à des exigences de latence et de sécurité accrues, et de s'adapter aux évolutions rapides des technologies. **Ces besoins ne semblent pas mobiliser la même volonté politique que le raccordement universel à la fibre, alors même qu'ils conditionnent notre compétitivité économique et notre capacité à soutenir les transformations numériques à venir, et à garantir l'inclusion numérique de la population française par la même occasion, dans un contexte où 80 % des données des Français sont stockées et hébergées aux États-Unis**.

Le Plan Très Haut Débit de 2013 et le *New Deal Mobile* de 2018, avec un budget global de 36 Mds €[64] (soit la moitié de l'enveloppe allouée à l'innovation de rupture *via* France 2030[65]), ont produit des résultats mitigés

[64] *Banque des Territoires, juin 2024, « Rencontres en territoires connectés : Retour sur le Plan France Très Haut Débit ».*

et difficiles à mesurer. **Il reste impossible d'évaluer précisément le nombre de foyers raccordés à l'internet haut débit. Cela s'explique par le manque de données fiables sur les locaux équipés d'un PTO (Point Terminal Optique, qui connecte un logement à la fibre optique) mais sans accès fibre activé. De plus, le comptage des PTO installés peut entraîner des doublons**. Selon l'ARCEP (fin T3 2024[66]), la France compte, 44,5 millions de locaux, 39,9 millions de locaux raccordables à la fibre, et 23,6 millions d'accès fibre activés (locaux où la fibre est en service). **Ainsi, environ 53 % des locaux peuvent être considérés comme « raccordés » à la fibre (23,6 M activés sur 44,5 M)**.

Encadré n° 5 • État des lieux de la couverture fibre par zone du territoire français

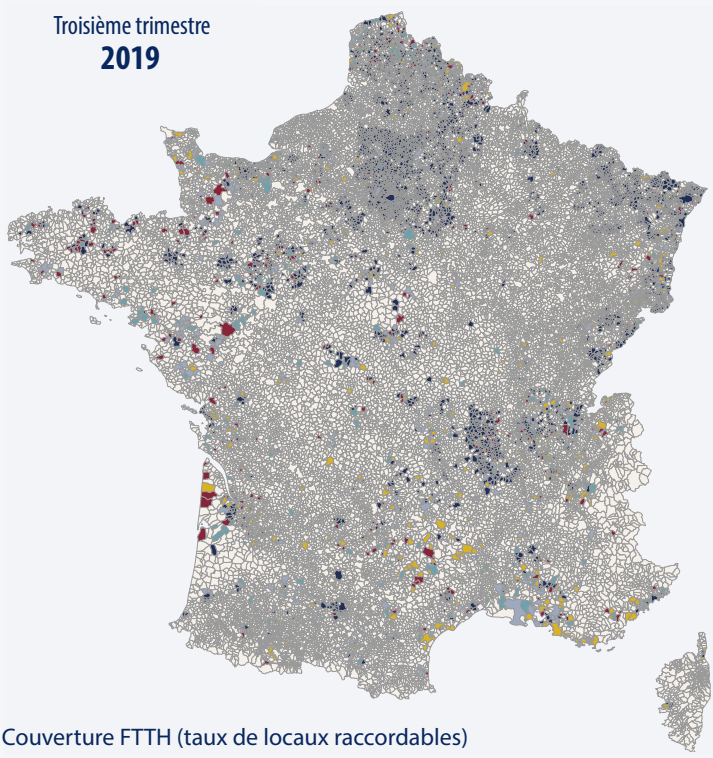
1. Zones très denses (ZTD) : 97 % de foyers couverts.
2. Zones moyennement denses (AMII) : 89 % de foyers couverts.
3. Zones d'initiative publique (RIP) : 77 % de foyers couverts.

Note pour le lecteur :
Les zones très denses (ZTD), zones moyennement denses (AMII) et zones d'initiative publique (RIP) correspondent à des classifications du territoire en fonction de leur densité de population et de leur attractivité économique, influençant le modèle de déploiement des réseaux : les ZTD sont déployées par des opérateurs privés, les zones AMII résultent d'un co-investissement public-privé, tandis que les RIP concernent les zones les moins rentables, où les collectivités locales pilotent le déploiement avec le soutien de subventions publiques.

[65] *Source : RAP, 2022, Investir pour la France de 2030. Les 54 Mds € correspondent à 34 Mds € votés en LFi 2022 auxquels s'ajoutent 20 Mds € du PIA 4.*

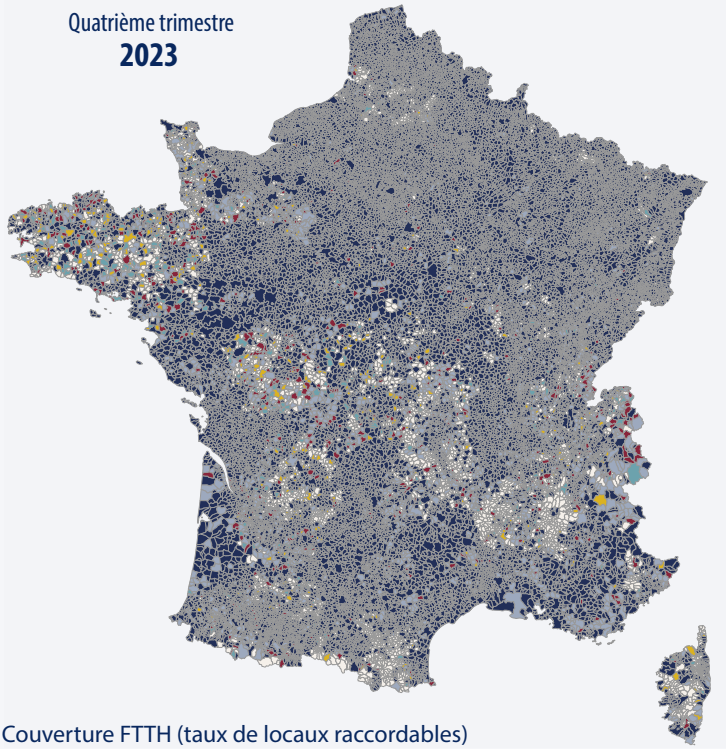
[66] *ARCEP, décembre 2024, « Marché du haut et du très haut débit fixe ».*

Graphique n° 11 • État des lieux de la couverture fibre par zone du territoire français



Source : https://cartefibre.arcep.fr/.

Graphique n° 12 • État des lieux de la couverture fibre par zone du territoire français

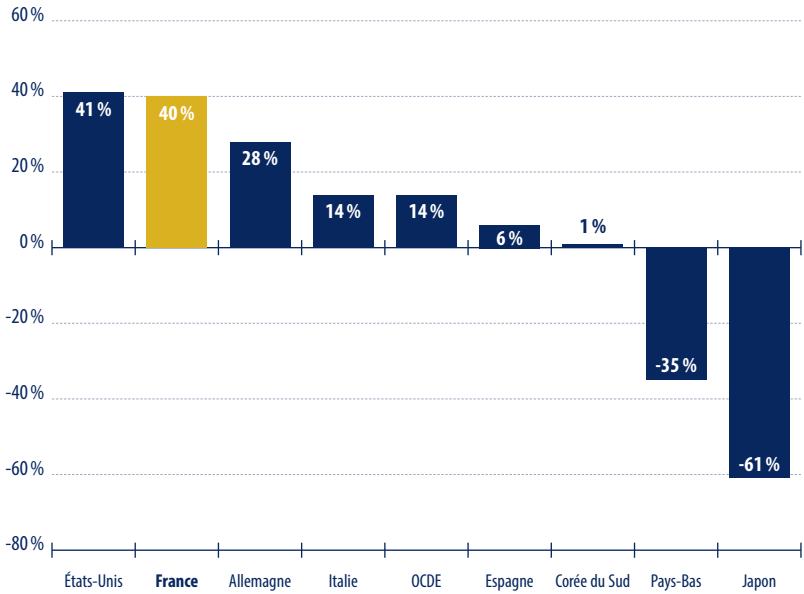


Source : https://cartefibre.arcep.fr/.

Les opérateurs télécoms ont investi massivement en ce sens, avec 14,7 Mds € [67] dépensés en France en 2022 dans l'ensemble des réseaux de connectivité (fixe et mobile), soit deux fois le montant investi dans le transport, le réseau ferré et la distribution d'électricité [68]. La France figure parmi les pays ayant enregistré les plus fortes hausses d'investissements dans les télécoms entre 2009 et 2018 [69].

Graphique n° 13 • Évolution des investissements entre 2009 et 2018 dans une sélection de pays

(évolution en euros courant)



Source : OCDE, Telecommunications database, Calculs DG Trésor.

[67] *ARCEP, « Les services de communications électroniques en France », résultats définitifs, année 2022.*

[68] *Étude Arthur D.Little, « L'économie du secteur des télécoms en France », Édition 2023.*

[69] *OCDE Telecommunications Database, Calculs DG Trésor.*

Cependant, ce choix soulève une question fondamentale : pourquoi avoir choisi de raccorder tous les foyers français sans distinction, indépendamment des besoins locaux ou des contextes géographiques ? Alors que cette démarche est justifiable économiquement dans les zones denses où les coûts sont maîtrisables et les retours sur investissement rapides, elle devient problématique dans les zones rurales ou isolées. Dans ces territoires, le raccordement des 20 % de foyers restants nécessite un investissement moyen de 300 à 400 euros par raccordement[70].

Pour ces zones, des alternatives comme la 5G fixe / Fixed Wireless Access (FWA)[71], ou des solutions satellitaires pourraient offrir une connectivité haut débit à moindre coût, tout en répondant aux besoins essentiels des habitants. C'est d'autant plus nécessaire que le modèle de financement des zones RIP nécessite des recettes complémentaires pour rester soutenable car il a été aligné avec celui des zones plus denses.

Encadré 6 • Deux technologies complémentaires à la fibre (FTTH) pour connecter à internet les derniers foyers français

1. **La 5G fixe :** la 5G fixe, une forme de FWA, utilise le réseau mobile pour relier des antennes fixes aux foyers, offrant des débits jusqu'à 1 Gbps, comparables à la fibre, dans les zones rurales où son déploiement est trop coûteux. Les quatre opérateurs français proposent déjà des offres basées sur des CPE internes (*customer premises equipment*, soit l'équipement qui permet de

[70] *Étude ARCEP 2024 sur les montants de la sous-traitance à l'opérateur commercial (STOC).*

[71] *Solution qui permet de fournir une connexion Internet à haut débit aux foyers et entreprises via le réseau mobile 5G, sans passer par une infrastructure filaire classique comme la fibre optique ou l'ADSL.*

capter le signal 5G), mais la couverture et l'éligibilité pourraient s'améliorer avec des modules externes pour les zones reculées. Des performances optimales nécessitent également le déploiement de la 5G sur la bande 3,5 GHz dans ces zones.

2. **Les technologies satellitaires :** les satellites en orbite basse (LEO), comme Amazon Kuiper ou Starlink, permettent d'offrir une connexion Internet jusqu'à 200 Mbps dans les zones où les infrastructures terrestres sont absentes. Cette solution est déjà utilisée en France dans des régions rurales ou montagneuses, avec l'avantage de pouvoir aussi servir de secours en cas de perturbations affectant les réseaux terrestres.

Dans un tel contexte, il est naturel de s'interroger sur le bon usage de nos moyens limités. 36 Mds € dédiés aux infrastructures de réseaux représentent une somme conséquente qui n'englobe pas les capacités de traitement, laissées principalement aux *hyperscalers*. **Le choix a donc été fait d'aller vers de la commodité au détriment d'investissements plus stratégiques servant un projet économique de long terme**.

Pour réussir, il est indispensable de clarifier les finalités des investissements d'avenir dans les infrastructures numériques. S'agit-il d'une dépense sociale destinée à réduire les fractures numériques, ou d'un investissement stratégique pour renforcer la compétitivité de nos infrastructures ? Si la première option peut se justifier, elle nécessite d'être explicitement assumée auprès des concitoyens, tandis que la seconde implique une réflexion approfondie sur les besoins à long terme et une priorisation des ressources en fonction des usages à fort potentiel économique. Ce n'est qu'en fixant un cap clair, centré sur les besoins des usages de demain, que la France pourra tirer pleinement parti de ses infrastructures numériques et soutenir efficacement la transition numérique.

Recommandation 1

Construire une offre souveraine cloud-réseau-edge-IoT « bout-en-bout » au niveau français et européen pour des usages aux dépendances maîtrisées.

Que souhaitons-nous réellement maîtriser et pourquoi ? Les réseaux, les infrastructures, les applications, les données ou les usages ? Cette réflexion implique de déterminer le périmètre géographique pertinent, et en particulier la taille du marché adressable, et les éléments de la « stack » technologique pour lesquels il faut faire émerger des acteurs souverains. La réponse réside sans doute dans une combinaison de ces dimensions. Dans un contexte où le paysage des infrastructures numériques évolue rapidement, nos opérateurs télécoms risquent d'être relégués aux segments à plus faible valeur ajoutée de la chaîne de valeur alors qu'ils disposent d'avantages comparatifs technologiques majeurs. Cela exige donc une anticipation des technologies et des usages souverains afin de planifier les options de mitigation des risques.

Pour relever ces défis, notre réponse repose sur trois axes essentiels : lutter contre l'ingérence technologique, renforcer la sécurité des systèmes numériques et mener la bataille du B2B industriel grâce à une offre compétitive et souveraine. Ces priorités sont indispensables pour préserver notre autonomie technologique tout en restant compétitifs à l'échelle mondiale.

Au vu des capitaux nécessaires pour développer en propre des plateformes numériques et compte tenu de nos dépendances aux chaînes d'approvisionnement de composants, matériels et intergiciels reliant les matériels aux applications, **il est impératif de comprendre les possibilités technologiques qui nous sont encore accessibles pour poser nos arbitrages. Les infrastructures de traitement de données**

de proximité, dites « edge », sont un élément de réponse adapté à des usages spécifiques car cette technologie est particulièrement appropriée à la cartographie et à la nature des données industrielles de nos ETI et PME.

Recommandation 1.1 : mettre en place une stratégie d'intégration verticale du continuum cloud-réseau-edge-IoT dans les secteurs suivants : la santé, les systèmes critiques (mission critical systems), l'aviation, la mobilité, les lanceurs spatiaux et la finance (cf. graphique n° 9).

Favoriser la création d'un acteur d'envergure couvrant l'ensemble de la chaîne de valeur, exclusivement dédié à ces secteurs stratégiques. Cette initiative doit s'appuyer sur les industriels français proposant déjà des services multiplateformes (Atos, Orange, Dassault Systèmes, Bouygues, Illiad, Docaposte, Eclairion, Sesterce, etc.), tout en évitant de disperser les efforts dans des secteurs où la loi du marché peut prévaloir.

Développer des passerelles IoT, des serveurs *edge*, des micro *data centers* et des solutions de 5G privée pour garantir la disponibilité et la sécurisation des données sur les territoires concernés. **Ces infrastructures devraient être déployées en priorité sur des sites sensibles, permettant d'expérimenter l'automatisation et la sécurisation des processus industriels critiques. À terme, elles serviraient de socle à un projet ambitieux de « territoires d'avenir »**, où l'Europe aurait un contrôle total sur les données sensibles, qu'il s'agisse de la gestion des infrastructures (routes intelligentes, gestion énergétique) ou de l'offre de services aux populations et aux entreprises.

Promouvoir une architecture cloud-edge-IoT combinant des connexions fibre haut débit reliant les bassins de population, économiques et industriels, et des solutions sans fil pour la périphérie. Les boxes 5G proposées

par Orange, Bouygues, Free ou SFR, offrant des débits proches de la fibre à un coût d'infrastructure réduit (une antenne étant bien moins onéreuse que le câblage intégral d'un quartier), illustrent le potentiel de ce modèle.

Explorer une approche complémentaire reposant sur des solutions satellitaires, telles que celles proposées par OneWeb, adaptées à des besoins spécifiques comme le backup dans les territoires ultra-marins, la couverture des zones blanches ou les régions difficiles d'accès. Ces infrastructures, encore coûteuses, nécessitent un financement à l'échelle européenne et restent pour l'instant majoritairement orientées vers des usages de niche en expansion.

Recommandation 1.2 : permettre une meilleure accessibilité de la donnée au plus proche de l'équipement de l'utilisateur pour permettre une offre souveraine fonctionnelle par le biais d'incitations à l'égard des créateurs et détenteurs de données.

Ces incitations peuvent être imposées aux collectivités locales et aux acteurs publics à des fins de participation à l'innovation nationale et d'exemplarité.

Elles peuvent être financières à l'égard d'acteurs privés (en santé, en formation, en mobilité…) afin de favoriser un partage optimal permettant ainsi la standardisation des formats de données industrielles, et leur structuration en temps réel.

Recommandation 1.3 : sécuriser l'offre souveraine qui en découlerait en constituant des *data spaces* (infrastructures de partage sécurisé de données) français dédiés.

Ce dispositif devra faire l'objet d'une redondance systématique sur les couches logicielles et s'assurer de la solidité de son approvisionnement en composants critiques. Il sera sans doute nécessaire de renforcer les prérogatives (autorités et moyens) du Délégué interministériel aux approvisionnements en minerais et métaux stratégiques et des instances impliquées (DGE et SISSE).

La question de l'investissement et de la participation européenne à un projet souverain reste aujourd'hui sans réponse véritable.

1.3. AU NIVEAU EUROPÉEN L'URGENCE D'UNE VISION COMMUNE SUR LES OPPORTUNITÉS TECHNOLOGIQUES À SAISIR ET LES DÉPENDANCES STRATÉGIQUES À MAÎTRISER

Pour construire un marché unique européen des infrastructures numériques, il est crucial de définir le projet politique numérique qui doit en être le socle. **Le problème, aujourd'hui, réside moins dans un défaut européen en matière d'investissements de rupture qu'en l'absence d'une vision commune claire et cohérente sur un tel projet. Nul ne s'inquiète également du coût précis de l'inaction ou de la non-convergence des réseaux et des logiciels dans les années à venir**. Or, la capacité à intégrer le nouveau continuum cloud-edge-IoT-réseaux de manière cohérente dans les offres des entreprises conditionnera non seulement la compétitivité technologique des acteurs, mais aussi leur aptitude à capter la valeur générée par les nouveaux usages. C'est pourquoi la Commission européenne a invité les pays membres à changer de vision sur les infrastructures numériques pour les penser de manière intégrée avec des réseaux de type 3C – connectés, collaboratifs, calcul –, soit en « écosystème »[72]. À la manière des corridors 5G, qui sont des zones géographiques spécifiquement aménagées pour permettre un déploiement continu et performant des réseaux 5G le long

[72] *Livre blanc de la Commission européenne, février 2024, « Comment maitriser les besoins en infrastructures numériques de l'Europe ».*

d'axes de transport existants, des European Digital Hubs ont ainsi été mis en place pour expérimenter les nouveaux modèles associés.

Si les 200 Mds € annoncés pour financer des infrastructures d'IA vont dans le bon sens, le projet européen ne peut se limiter à des ambitions générales. Il doit proposer une vision structurée et concrète, capable de susciter l'adhésion des États membres et des acteurs économiques, articulée autour de deux piliers essentiels :

1. Une méthode rigoureuse pour orienter les investissements vers des initiatives de rupture porteuses de valeur réelle, en réponse aux besoins concrets des secteurs stratégiques tels que la santé, l'énergie ou l'industrie.
2. Une définition explicite des dépendances stratégiques, identifiant celles que l'Europe peut tolérer et celles qu'elle doit impérativement maîtriser pour préserver son autonomie stratégique.

a. Une méthode incomplète pour investir dans des initiatives de rupture

La première question centrale concerne la méthode : comment les grandes entreprises, tous secteurs confondus, peuvent-elles contribuer au développement d'initiatives de rupture dans le domaine numérique à l'échelle européenne ? Plus encore, comment permettre aux entreprises qui ne relèvent pas directement du secteur numérique d'agir et d'influencer les orientations budgétaires européennes, afin de s'assurer que ces investissements répondent également à leurs besoins stratégiques ?

L'Europe fait face à un écart d'innovation qu'il devient urgent de combler pour réduire le fossé croissant avec les États-Unis, la Chine et les économies d'Asie du Sud-Est. Cet écart est particulièrement marqué dans des secteurs critiques comme les semi-conducteurs, l'IA et les technologies de rupture, où les investissements européens demeurent bien en deçà

de ceux de ses principaux concurrents. En 2021, l'Union européenne a investi environ 2,2 % de son PIB en R&D, contre 3,1 % pour les États-Unis et 2,4 % pour la Chine[73]. **Pour relancer la compétitivité européenne, le rapport Draghi[74] recommande un « choc d'investissement » de 800 Mds € par an sur la période 2025-2030, soit l'équivalent de 4,7 % du PIB de l'UE. Cependant, l'absence d'une méthode claire pour canaliser ces fonds vers des projets stratégiques limite l'impact de ces ambitions et ralentit les progrès attendus**.

La méthode actuelle pour mobiliser les investissements européens repose largement sur le secteur privé, mais les outils utilisés montrent leurs limites. Les incitations fiscales, comme les crédits d'impôt pour la R&D ou les déductions pour investissements, se révèlent insuffisantes pour inciter les entreprises à s'engager dans des projets risqués ou de rupture technologique. Par ailleurs, les mécanismes de mutualisation, inspirés de la réponse à la pandémie de Covid-19 avec les emprunts communs, sont difficiles à reproduire à l'échelle de l'Union européenne. Les divergences politiques entre États membres et la réticence de certains pays à accroître leur dette commune freinent leur mise en œuvre. **En réalité, les capacités techniques ne sont pas homogènes en Europe, comme le révèle l'écart dans les certifications de cybersécurité dans le cadre de la mise en œuvre de la directive NIS 2**. Certains États membres peinent à aligner leurs standards avec les exigences européennes, ce qui crée des disparités dans la sécurisation des infrastructures critiques. De plus, les instruments actuels sont souvent inadaptés aux enjeux de rapidité et d'innovation. Les Projets Importants d'Intérêt Européen Commun (PIIEC) manquent d'agilité, comme l'a démontré leur mise en œuvre dans l'industrie des semi-conducteurs, des batteries ou de l'hydrogène, où des retards dans les subventions et des processus complexes ont limité l'impact sur la

[73] *Contrepoints, J-B .Noé, juin 2024, « Selon Michel Cicurel (La Maison), pour rattraper les Etats-Unis et la Chine, l'Europe doit flécher l'épargne des Européens vers des investissements prioritaires ».*

[74] *M.Draghi, septembre 2024, Rapport: « The future of European competitiveness », https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en.*

compétitivité européenne. Des modèles inspirés de la DARPA *(Defense Advanced Research Projects Agency)* fonctionnent aux États-Unis, grâce à une centralisation des décisions et à une prise de risque assumée dans des projets innovants. En Europe, ces modèles peinent à s'appliquer en raison de la fragmentation institutionnelle, de la lenteur des processus décisionnels et de l'absence d'une doctrine commune sur le financement du risque technologique. **Il n'existe pas en Europe d'appels à projets de type *high risk high gain*, avec une méthode adaptée à une configuration dans laquelle un faible pourcentage, généralement autour de 10 %, des projets financés fonctionne réellement**. La bureaucratie européenne excessive, incarnée par le fonctionnement en silo de nombreuses agences, limite la mise sur le marché des avancées permises par une recherche fondamentale de qualité.

Un problème majeur réside dans l'absence de doctrine claire en matière de gestion du risque technologique en Europe. Contrairement à des pays comme la Chine ou les États-Unis, qui assument pleinement des stratégies de patriotisme économique ou de soutien massif à l'innovation locale, voire de « gâchis » de l'argent public pour faire émerger des vainqueurs, l'Europe adopte une posture excessivement prudente. Les entreprises européennes, souvent préoccupées par la minimisation de leurs responsabilités en cas d'échec, privilégient des solutions étrangères « sur étagère » – principalement américaines ou chinoises – perçues comme fiables et sans risques immédiats. Cette approche, axée sur l'évitement du risque, freine l'émergence de projets de rupture locaux et maintient l'Europe dans un rôle de suiveur technologique. Ce manque de prise de risque est particulièrement problématique dans des secteurs critiques comme la 5G ou la 6G, où la Chine avance rapidement grâce à des investissements massifs dans le développement de brevets et d'innovations nationales[75]. **Bien que l'Europe dispose de fondements solides grâce à ses infrastructures de base – câbles sous-marins, réseaux de télécommunications –, elle échoue**

[75] *Business France, 17 janvier 2023, « Le développement de la 5G en Chine en 2022 ».*

à investir dans les couches technologiques supérieures, comme les systèmes logiciels avancés ou les télécommunications de nouvelle génération, qui sont les véritables moteurs de compétitivité pour le futur. L'urgence est donc de définir une doctrine européenne du risque technologique, non pas pour éviter l'échec à tout prix, mais pour encourager activement l'audace et valoriser les innovations de rupture qui assureront la souveraineté technologique de demain.

Les semi-conducteurs illustrent le retard technologique de l'Europe sur le plan mondial. **Malgré une mobilisation américaine de 55 Mds $ pour la production de semi-conducteurs pour une part de marché de 13 %, contre 43 Mds €[76] pour l'Europe pour une part de marché de 11 %, l'essentiel de la production mondiale reste concentré en Asie qui détient près de 80 % des capacités globales de production**. TSMC domine la fabrication, captant 62 % des revenus mondiaux en 2024 (contre 59 % en 2023)[77], tandis que Samsung, détient 10 %[78] du marché. **Toutefois, les États-Unis dominent le marché de la conception et de la vente de produits finis avec 48 % de parts de marché en 2022[79] et des acteurs comme Qualcomm, Nvidia, Intel, AMD et Apple, ce qui n'est pas le cas de l'Europe.**

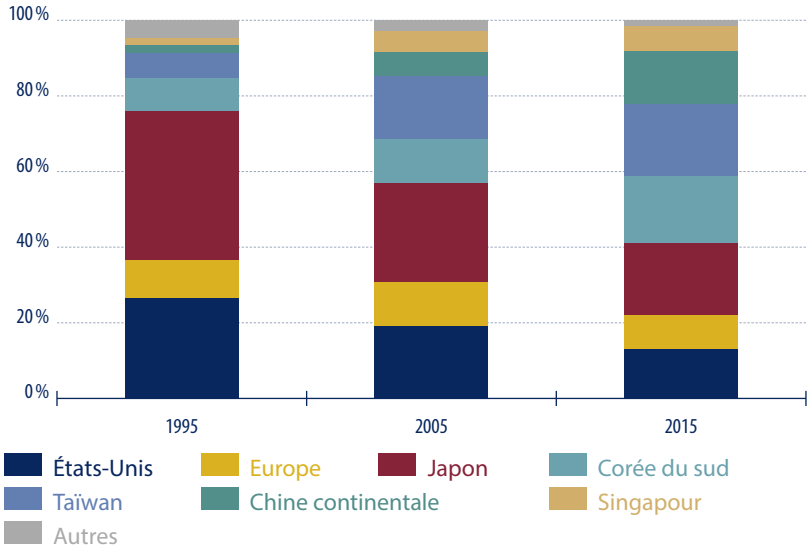
[76] *Commission européenne, « European Chips Act » de 2022 avec l'objectif d'atteindre 20 % des parts de marché en 2030, ce qui représente une multiplication par 4 des capacités de production européennes puisque le marché mondial sera amené à doubler d'ici là.*

[77] *Trendforce, « Global Foundry Revenu Share », Février 2024.*

[78] *Ibid.*

[79] *DGE, M. Ericher, J. Seux, S. Toe et L. Tournier, janvier 2025, « Les semi-conducteurs : un marché mondialisé et une dépendance européenne ».*

Graphique n° 14 • Répartition géographique des capacités globales de production de semi-conducteurs.



Note de lecture : en 2015, le Japon représentait 18 % des capacités de production mondiales de semi-conducteurs alors qu'il en représentait 40 % en 1995.

Source : Thema DGE, janvier 2025.

Les États-Unis ont non seulement investi dans la relocalisation de la production *via* TSMC en Arizona, action symbolique, Intel et Micron, mais ils ont également encouragé des collaborations stratégiques comme l'initiative Ultra Accelerator Link[80]. Ce projet, porté par Intel, Meta, Microsoft, et AMD, vise à standardiser les systèmes d'interconnexion pour accélérateurs d'IA[81] et à réduire la domination de Nvidia

[80] *HPC wire, D. Eadline, 30 mai 2024, « Everyone Except Nvidia Forms Ultra Accelerator Link (UALink) Consortium ».*

[81] *Le projet prévoit la création d'un système permettant de connecter plus de 1 000 accélérateurs d'IA en un seul module de calcul, pour ouvrir un système aujourd'hui contrôlé par Nvidia, avec les premiers composants attendus dans les prochaines années.*

avec NVLink. Ces démarches ne se limitent pas à combler des lacunes, mais instaurent des écosystèmes flexibles et innovants, témoignant d'une approche tournée vers le leadership technologique. **En parallèle, la Chine investit massivement pour réduire sa dépendance vis-à-vis des machines de lithographie[82] européennes d'ASML[83]**. Les récents progrès de Naura Technology Group et de la Semiconductor Manufacturing International Corporation chinoise (SMIC), capables de produire des puces en 5 nm, illustrent une stratégie proactive pour maîtriser les technologies critiques. De même, le Japon a lancé le projet Rapidus[84] un effort national soutenu par le Gouvernement et des industriels tels que Toyota et Sony, pour développer localement des puces à 2 nm d'ici 2027. Avec un investissement de 550 M $, il a permis la création d'un consortium avec 8 entreprises[85], pour construire une base de production de puces de 2 nm et moins à Hokkaido.

L'Europe, pour sa part, manque d'infrastructures et de compétences pour produire des technologies avancées, notamment les puces de moins de 7 nm, indispensables pour l'intelligence artificielle et les équipements numériques avancés. **Cette dépendance l'expose à des risques multiples : restrictions à l'exportation, comme celles récemment imposées par les États-Unis à la Chine, ou perturbations liées aux tensions géopolitiques**. Sa posture est principalement défensive avec une approche dite « systématique »[86] de la sécurité économique, désormais considérée comme une priorité stratégique[87]. La création

[82] *Technique de gravure utilisant un faisceau d'électrons pour créer un motif en creux sur une surface. Cette technologie de pointe est particulièrement utilisée dans la production de nanotechnologies (semiconducteurs).*

[83] *Vietnam.vn, « Huawei a trouvé un moyen de produire des puces de 5 nm malgré les efforts américains pour l'empêcher », 2024.*

[84] *Siècle Digitale, Z. Tazrout, 3 avril 2024, « Pour que Rapidus passe à la vitesse supérieure, le Japon met les bouchées doubles ».*

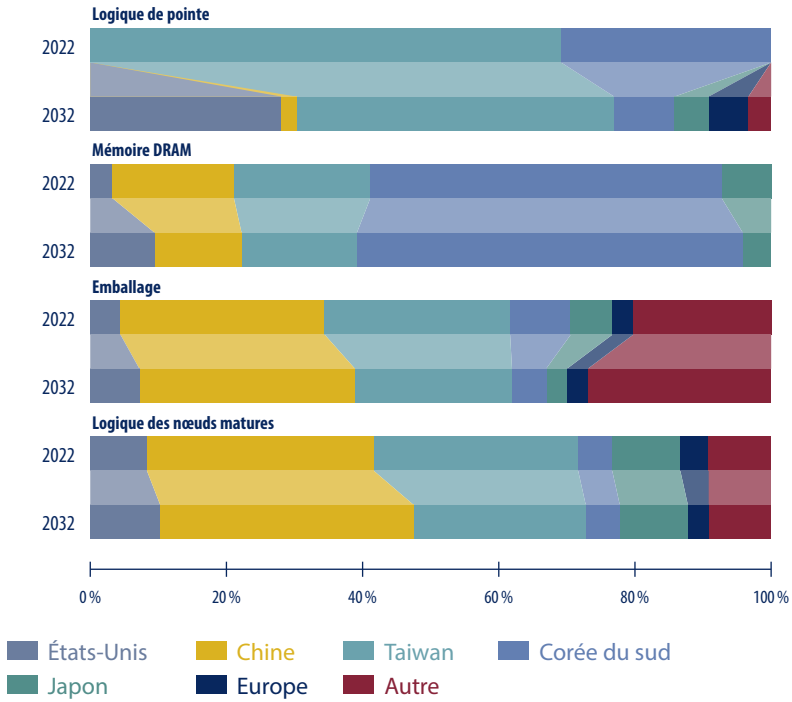
[85] *Denso, Kioxia, NEC, NTT, Softbank, Sony, Toyota et MUFG Bank.*

[86] *Commission européenne, 24 janvier 2024, « La Commission propose de nouvelles initiatives pour renforcer la sécurité économique ».*

[87] *Commission européenne, juin 2023, « Une approche de l'UE pour renforcer la sécurité économique ».*

d'un poste de commissaire au commerce et à la sécurité économique, ainsi que des instruments comme le règlement européen sur les biens à double usage ou la directive sur la résilience des entités critiques, montrent toutefois une prise de conscience réelle des menaces pesant sur ses infrastructures numériques.

Graphique n° 15 • États des lieux de l'avance des acteurs européens sur la production mondiale de semi-conducteurs



Note : les données pour 2032 représentent les prévisions de la SIA. La logique de pointe est définie comme étant inférieure à 10 nanomètres, la logique de nœud mature est définie comme étant supérieure ou égale à 28 nanomètres. Le plan de production américain comprend d'autres technologies qui ne sont pas représentées ici.

Source : Semiconductors Industry Association, Boston Consulting Group.

Graphique n° 16 • Les étapes de fabrication d'une puce

Dépot
Une couche de film est déposée sur une plaquette de silicium.

Revêtement avec photorésine
Un revêtement photosensible, la photorésine ou résine photorésistante, est appliqué sur le dessus.

Lithographie
Un motif est créé sur la photorésine à l'aide d'ultraviolets.

Cuisson et développement
La plaquette est cuite et un développement est effectué pour intégrer le motif.

Gravure
L'excès de matière est éliminé à l'aide de gaz (processus sec) ou de bains chimiques (processus humide).

Implantation ionique
La plaquette est bombardée d'ions, ce qui créé des transistors et facilite le flux de l'électricité.

Ce processus doit être répété pour chaque couche supplémentaire

Retrait de la photorésine
La photorésine restante est enlevée.

Conditionnement
Les puces individuelles sont obtenues en découpant la plquette à l'aide d'une scie diamant. Elles sont placées sur un « substrat » qui envoie des signaux à d'autres puces et surmontées d'un couvercle « calo-porteur » pour éviter qu'elles ne chauffent.

Source : ASML, Mega.

Encadré n° 7 • Différence entre concepteurs et fabricants de puces

- Les concepteurs de puces, souvent appelés *fabless companies*, se spécialisent dans la conception et le développement de circuits intégrés sans disposer des infrastructures nécessaires pour les fabriquer. Des entreprises telles qu'Apple, AMD, Nvidia et Qualcomm conçoivent des processeurs et des puces graphiques, mais dépendent de fonderies tierces pour la production.

- En revanche, les fabricants de puces, comme TSMC (Taiwan Semiconductor Manufacturing Company), Intel, Apple et GlobalFoundries, possèdent et exploitent des installations de production sophistiquées utilisant des technologies avancées de lithographie, de gravure et de dépôt pour produire des puces à grande échelle. Intel est un cas particulier, car il conçoit et fabrique ses propres puces, ce qui en fait un IDM (Integrated Device Manufacturer).

Ce constat met en lumière un problème stratégique : ce ne sont pas uniquement les montants alloués qui posent question, mais surtout la cohérence entre ces montants et les objectifs fixés. **Sans une redéfinition réaliste des priorités et une meilleure articulation entre investissements et ambitions, la France (et l'Europe avec elle) risque de diluer ses efforts et de ne pas atteindre ses propres cibles**. C'est l'objet de ce rapport : appeler à une prise de décision lucide et pragmatique, où les moyens alloués, la structuration des initiatives et la définition des objectifs s'alignent pour garantir la réussite des ambitions françaises et européennes.

b. Un manque de consensus sur les dépendances à maîtriser

Les infrastructures numériques françaises reposent sur des chaînes de valeur mondiales fragmentées et dépendantes de technologies étrangères. Les acteurs français et européens sont en effet fortement dépendants de technologies numériques produites par des acteurs étrangers, ce qui expose l'Europe à des vulnérabilités stratégiques. Plus de 70 % des données françaises sont hébergées sur des *clouds* américains, tandis que la production de semi-conducteurs repose largement sur des fournisseurs asiatiques[88].

Le champ de contraintes européen est triple en matière de dépendances stratégiques, et si le pragmatisme fait désormais consensus, aucun choix clair n'a été fait par l'ensemble des pays européens.

1. **L'Europe fait face à des vulnérabilités stratégiques liées aux pratiques extraterritoriales des États-Unis**, notamment le *Cloud Act* de 2018, qui élargit le cadre du *US Stored Communications Act* de 1986 pour permettre, dans le cadre d'une enquête fédérale, aux autorités américaines d'accéder aux données stockées à l'étranger par des entreprises américaines *via* des accords bilatéraux d'accès aux données. Ces risques ont été exacerbés par l'extension en avril dernier de la section 702 du *Foreign Intelligence Surveillance Act* (loi FISA[89]) de 1978. Cette mesure élargit les capacités des agences de renseignement américaines à accéder aux données de non-ressortissants, même lorsqu'elles sont hébergées hors des États-Unis, dès lors qu'elles sont détenues par des fournisseurs de services électroniques américains. **Elle cible également les *meet-me rooms*[90] des *data centers*, ces espaces où transitent des données**

[88] *Oliver Wyman, 2020, European Digital Sovereignty. NB que 92 % des données du « monde de l'ouest » sont stockées sur des serveurs aux États-Unis.*

[89] *Champ d'application élargi aux entités « ayant accès à des équipements qui sont ou peuvent être utilisés pour transmettre ou stocker des communications filaires ou électroniques ».*

stratégiques, telles que des échanges sensibles entre multinationales et leurs filiales ou des données critiques appartenant à des gouvernements. Ces infrastructures sont particulièrement vulnérables, malgré les mécanismes de chiffrement, face à des interceptions de clés ou aux obligations légales de divulgation imposées par cette législation.

- Pour renforcer la sécurité des infrastructures numériques, la récente loi de programmation militaire française (LPM) autorise l'ANSSI à déployer des outils techniques, comme des sondes qualifiées, directement au sein des *data centers*, et ce sans intervention judiciaire.

- Toutefois, la priorité pour les États membres européens reste l'accès au marché américain et la préservation des relations économiques transatlantiques.

- Bien que les lois américaines prévoient généralement une procédure judiciaire ou administrative pour justifier l'accès aux données sensibles, ces garanties restent insuffisantes pour protéger les données critiques européennes, notamment celles liées aux infrastructures essentielles et aux projets de défense. **En réponse, l'Europe a tenté d'élaborer un référentiel commun avec Gaia-X, destiné à établir des normes de confiance pour structurer des écosystèmes de données et d'infrastructures**. Cependant, ce projet a fait l'objet de nombreuses critiques, notamment en raison de son ouverture aux acteurs non européens, ce qui a affaibli son impact en matière de souveraineté numérique.

2. Deux vulnérabilités en matière d'approvisionnement stratégique : les matériaux critiques et les ressources énergétiques.

[90] *Espace sécurisé où différents opérateurs ou clients peuvent interconnecter leurs équipements via des points d'accès communs, facilitant ainsi l'échange de données et la mutualisation des infrastructures.*

- Pour ce qui concerne les matériaux critiques, les fibres optiques, les GPU ou les équipements 5G dépendent de matériaux critiques extraits ou produits sur divers continents. Cette dispersion amplifie les risques de ruptures, comme l'ont montré les récentes restrictions chinoises[91] sur des métaux essentiels tels que le gallium, le germanium et l'antimoine, en réponse aux sanctions américaines. Ces mesures perturbent les chaînes logistiques, allongent les délais de production et aggravent les goulots d'étranglement, particulièrement dans les secteurs stratégiques des infrastructures numériques.

- Pour ce qui concerne l'approvisionnement énergétique, les quantités d'uranium, de gaz, et de pétrole sont limitées, alors que ces ressources sont indispensables au fonctionnement des infrastructures numériques. La croissance de la demande liée à la construction de réacteurs ou à l'expansion des infrastructures numériques pourrait exacerber cette dépendance, augmenter les coûts et limiter la disponibilité des ressources critiques. Cette combinaison de vulnérabilités, technologiques et énergétiques, souligne la fragilité structurelle des chaînes d'approvisionnement européennes.

La dépendance aux intrants chinois dans le secteur des batteries électriques et aux fournisseurs étrangers dans plusieurs segments de l'industrie des semi-conducteurs illustre la vulnérabilité des infrastructures numériques européennes. Ces situations ne sont pas hypothétiques : elles traduisent un nouvel ordre économique où les chaînes d'approvisionnement peuvent devenir des armes stratégiques. **Pour anticiper ces défis, l'agenda de sécurité économique européen adopte désormais une approche pragmatique, en diversifiant ses fournisseurs tout en investissant dans des capacités locales de production stratégique, mais le « comment » reste à déterminer**.

[91] *Le Figaro avec l'AFP, 3 décembre 2024, « La Chine restreint les exportations de composants essentiels à la fabrication de puces vers les États-Unis », https://www.lefigaro.fr/flash-eco/la-chine-restreint-les-exportations-de-composants-essentiels-a-la-fabrication-de-puces-vers-les-etats-unis-20241203.*

3. Les semi-conducteurs : une absence de consensus sur le niveau de maîtrise de la chaîne de valeur en Europe.

Bien que le Chips Act ait fixé des principes directeurs, la concurrence fait rage en Europe pour attirer des projets d'usines étrangères de semi-conducteurs, au lieu de se coordonner pour construire des usines européennes. Si l'usine de STMicroelectronics et GlobalFoundries à Crolles répond aux besoins critiques de secteurs comme l'automobile et l'IoT, aucune usine européenne ne peut concevoir des puces <11 nm. Ce déficit, aggravé par une approche fragmentée des investissements, exacerbe la concurrence entre États membres pour attirer des usines capables de produire ces puces avancées. Ainsi, l'Allemagne a alloué 10 Mds € de subventions publiques pour attirer Intel et TSMC, bien que le projet d'Intel ait finalement été reporté. **Cela empêche paradoxalement les acteurs européens de tirer parti des équipements d'ASML et développer des capacités locales sur des segments à haute valeur ajoutée**. En effet, bien que ASML soit le leader mondial de la lithographie avancée (EUV), 46 %[92] de ses ventes en 2023 ont été réalisées en Chine, dépassant Taïwan et la Corée du Sud.

Construire des usines ne suffit pas car rentabiliser ces installations nécessite des volumes de production élevés et un marché suffisamment large pour absorber les stocks. À ce jour, l'Europe peine à instaurer cette dynamique, malgré une concentration de 40 % des investissements mondiaux en semi-conducteurs[93]. Par ailleurs, les composants à haute densité, largement dominés par des acteurs asiatiques comme TSMC et Samsung, restent un verrou stratégique pour l'Europe. Selon une étude publiée en octobre 2024 par l'Université d'Oxford[94], les États-Unis ont pu concentrer la majorité des puces Nvidia H100, grâce à

[92] *ASML, leader européen pour la fabrication de puces, restreint ses exportations vers la Chine. (2024, January 3). Journal De L'Économie, https://www.journaldeleconomie.fr/asml-leader-europeen-pour-la-fabrication-de-puces-restreint-ses-exportations-vers-la-chine/.*

[93] *École de Guerre Économique, (date), La difficulté pour l'Europe de penser la réduction de ses dépendances dans le numérique et dans l'énergie.*

des infrastructures déjà adaptées à leurs exigences techniques dès leur commercialisation en 2022.

Pour combler ces lacunes, l'Europe a lancé la *Chips Joint Undertaking*, en étroite collaboration avec le secteur privé, pour transformer la chaîne de valeur des semi-conducteurs et garantir une indépendance technologique. Une telle démarche vise à inscrire l'Europe dans une dynamique proactive et à sécuriser sa position sur les segments stratégiques des infrastructures numériques. Elle est toutefois conditionnée à une collaboration plus étroite entre pays membres de l'Union européenne, et à un consensus clair sur le niveau de maîtrise à atteindre le long de la chaîne de valeur (fabrication et conception).

2 Le développement sécurisé de nos infrastructures numériques de traitement de données doit être une priorité stratégique de l'État

Le retard dans l'adoption des nouvelles technologies d'infrastructure, telles que la 5G, les *data centers* et les solutions de calcul avancés, s'explique essentiellement par l'absence de vision claire sur leurs cas d'usage concrets. Ce manque freine la perception de leur valeur ajoutée et limite l'engagement des acteurs privés et publics. Sans une feuille de route structurée, définissant les usages stratégiques, l'adoption de ces technologies continue d'être perçue comme un risque plutôt qu'une opportunité pour la compétitivité et l'économie.

[94] *Université d'Oxford, V. Lehdonvirta, B. Wu, and Z. Hawkin, 16 octobre 2024, "Compute North vs. Compute South: The Uneven Possibilities of Compute-based AI Governance Around the Globe", https://osf.io/preprints/socarxiv/discover.*

2.1. LA NÉCESSITÉ D'UN DÉVELOPPEMENT MASSIF DE NOS CAPACITÉS DE CALCUL INTENSIF

Le calcul intensif est devenu un outil stratégique essentiel pour le traitement et l'analyse de données massives, car il joue un rôle clé dans des domaines tels que la recherche scientifique, l'intelligence artificielle et les simulations complexes. Le marché mondial des supercalculateurs, estimé à 32,4 Mds $[95] en 2023, devrait atteindre près de 50 Mds $[96] en 2026, reflétant une croissance rapide et soutenue. Cette progression s'accompagne d'une augmentation exponentielle de la puissance de calcul disponible, qui croît d'environ 40 % par an. Par exemple, la capacité mondiale est passée de 2,2 exaflops en juin 2020 à 8,2 exaflops en juin 2024 pour le TOP 500, soit les 500 supercalculateurs les plus puissants du monde, témoignant de l'intensification des besoins et des investissements dans ce secteur stratégique. Des pays comme le Royaume-Uni se sont déjà engagés à multiplier par 20 la capacité de calcul d'ici 2030, en créant notamment un nouveau « *superordinateur* »[97] .

a. La demande en capacités de calcul intensif est amenée à fortement augmenter dans les 5-10 prochaines années

Le développement de nouveaux usages, notamment les contenus récréatifs, et la facilité d'accès aux infrastructures numériques ont entraîné une hausse très importante du trafic global, et du besoin en

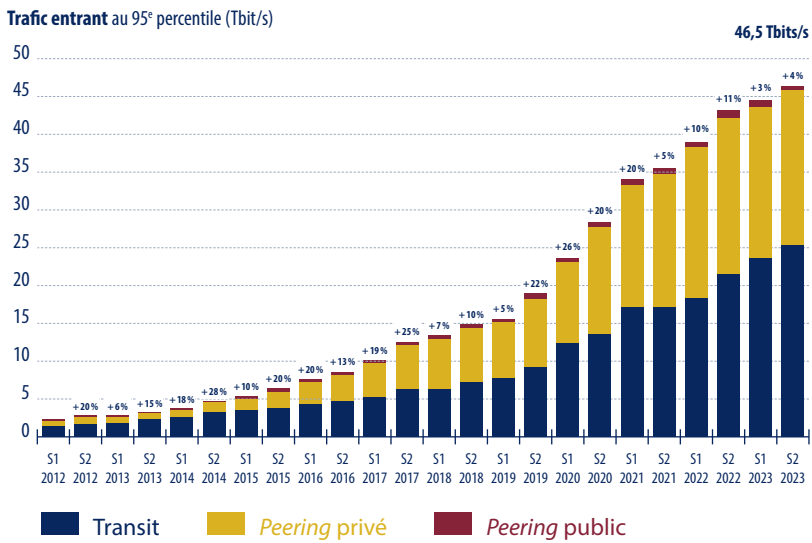
[95] *Le Monde, C. De Laubier, 11 septembre 2023, « Le retour de la course aux supercalculateurs entre Etats-Unis, Chine et Europe », https://www.lemonde.fr/economie/article/2023/09/11/le-retour-de-la-course-aux-supercalculateurs-entre-etats-unis-chine-et-europe_6188810_3234.html.*

[96] *Ibid.*

[97] *Le Figaro, A. Alexandre, 13 janvier 2025, « L'IA peut transformer la vie des travailleurs : l'Angleterre dévoile un grand plan pour « libérer » le potentiel de l'intelligence artificielle », https://www.lefigaro.fr/secteur/high-tech/le-royaume-uni-lance-un-plan-pour-liberer-l-intelligence-artificielle-et-dynamiser-l-economie-20250113.*

bande passante[98]. De fait, le trafic global entrant a été multiplié par plus de deux entre S1 2020 et S2 2023 sur les principaux FAI, pour atteindre 46,5 Tbits/s, soit une hausse de 7,5 % par rapport à fin 2022[99]. Par conséquent, les besoins en débit, latence et en résilience ont augmenté pour garantir la qualité du réseau sur de plus grandes distances.

Graphique n° 17 • Évolution du trafic entrant à l'interconnexion vers les principaux FAI en France entre S1-2012 et S2-2023

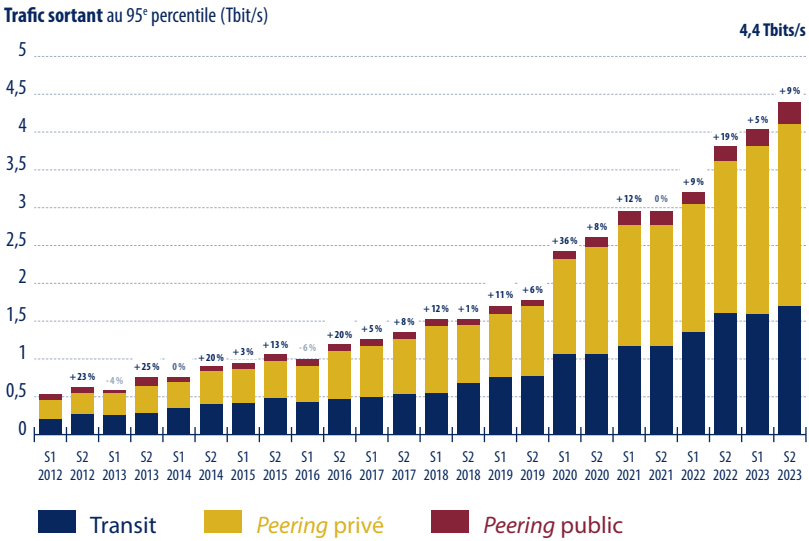


Source : Arcep.

[98] *Au début d'internet, il était calibré pour transmettre du texte (soit un ensemble restreint de 128 caractères) à une vitesse de 50 kilobits par seconde. Aujourd'hui, la vidéo représente près de 65 % du trafic, ce qui nécessite une vitesse de transfert de +100 mégabits par seconde.*

[99] *ARCEP, « État de l'Internet en France », juillet 2024.*

Graphique n° 18 • Évolution du trafic sortant à l'interconnexion vers les principaux FAI en France entre S1-2012 et S2-2023



Source : Arcep.

L'accès aux infrastructures numériques se fait de manière toujours plus proche de l'utilisateur final, avec des interfaces simplifiées et parfois intégrées directement dans son environnement, ce qui démultiplie les besoins futurs en matière de connectivité . En France, un citoyen possède en moyenne 15 équipements connectés en 2022, ce qui est supérieur à la moyenne mondiale de 8[100].

L'entraînement des modèles d'IA va également fortement augmenter les besoins en bande passante ces prochaines années. Le rythme de développement des capacités de calcul mobilisées pour l'entraînement

[100] *ADEME, « Évaluation de l'impact environnemental du numérique en France et analyse prospective », janvier 2022.*

des modèles de fondation est multiplié par près de 4 tous les ans. **Si les investissements en IA se maintiennent à rythme constant, un saut comparable à celui que nous avons observé entre les capacités de calcul mobilisées pour GPT-4 et GPT-2 pourrait se produire d'ici à 2030**[101]. Autrement dit, les cycles d'entraînement des modèles pourraient être 10 000 fois plus importants que ceux observés actuellement sur les modèles de pointe de type GPT-4, Grok, Llama, Gemini, Claude ou Yi Lightning. Ces besoins ont un impact majeur sur les coûts d'entraînement des modèles de fondation qui sont amenés à augmenter avec des montants compris entre 78 M$ pour GPT-4 et 191 M $ pour Gemini Ultra[102]. Il est important de noter que ce rythme concerne uniquement les modèles de fondation, qui sous-tendent les autres types de modèles plus étroits et spécialisés le long de la chaîne de valeur de l'IA, qui sont beaucoup moins coûteux à entraîner. **Le principal défi réside toutefois dans la capacité de calcul nécessaire pour l'inférence, compte tenu de la rapide diffusion des usages et de l'évolution des *scaling laws*[103] vers un modèle centré sur le *test-time compute*[104]**. Cela signifie que les performances accrues dépendent d'une augmentation des ressources de calcul au moment de l'inférence, les FLOPs utilisés après l'entraînement (post-training) surpassant désormais ceux mobilisés durant l'entraînement (pre-training) pour les modèles à l'état de l'art.

[101] *Epoch AI, August 20, 2024, Can AI Scaling Continue Through 2030?*

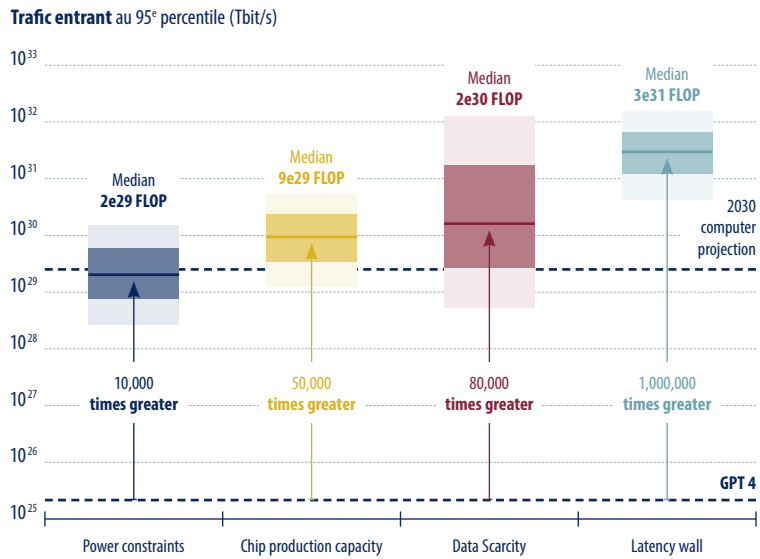
[102] *Stanford Institute for Human-Centered Artificial Intelligence, avril 2024, AI Index, https://spectrum.ieee.org/ai-index-2024.*

[103] *Relations mathématiques ou empiriques qui décrivent comment les performances des modèles d'IA évoluent en fonction de la taille des données d'entraînement, de la taille du modèle (nombre de paramètres) et des ressources de calcul utilisées. Ces lois montrent que, généralement, plus le modèle est grand et plus les données et les ressources de calcul sont importantes, meilleures seront ses performances.*

[104] *Quantité de calcul utilisée lors de la phase d'inférence, c'est-à-dire au moment où un modèle pré-entraîné est utilisé pour produire des prédictions ou des réponses. Contrairement à la phase d'entraînement, où les ressources sont principalement utilisées pour ajuster les paramètres du modèle, il met l'accent sur l'efficacité et la rapidité nécessaires pour exécuter les calculs sur des données en temps réel ou après le déploiement.*

Pour l'Union européenne, la priorité ne réside pas dans une course aux modèles de fondation contre les États-Unis, mais dans le développement de quelques modèles multi-modaux spécialisés, idéalement *open source*, qui soient maintenus sur le long terme. Ces modèles doivent être conçus pour répondre aux besoins spécifiques et accélérer leur adoption au sein de la société, offrant ainsi un levier stratégique pour l'innovation européenne.

Graphique n° 19 • Estimation par scénarios des besoins futurs en puissance de calcul pour entraîner les modèles d'IA d'ici à 2030



Source : Epoch AI.

Les algorithmes d'IA nécessitent des capacités de calcul substantielles pour entraîner des modèles sur de grands ensembles de données, en particulier pour le *machine learning* et le *deep learning*, qui demandent des infrastructures robustes afin de traiter les données et d'entraîner les modèles avec performance et évolutivité. Les réseaux neuronaux profonds, par exemple, exigent des calculs intensifs pour ajuster les paramètres des modèles. Selon une étude d'Intel de juin 2024[105], l'utilisation des accélérateurs d'IA (neural processing units – NPU) devrait croître fortement, les PC dotés de capacités IA représentant 80 % du marché d'ici 2028.

Les NPU sont une des promesses pour déployer à l'échelle des *small language models* (SLM), des modèles de langage optimisés pour fonctionner sur des infrastructures aux ressources limitées, offrant des capacités de traitement local. Les SLM sont plus légers que les grands modèles et consomment moins d'énergie, ce qui les rend adaptés à des environnements contraints, comme les appareils embarqués ou les systèmes décentralisés.

Encadré n° 8 • Focus sur le More Moore vs More Than Moore

Le concept de Moore désigne la poursuite de la miniaturisation des transistors, conformément à la loi de Moore. Cette loi, formulée en 1965, prédit que le nombre de transistors sur une puce double tous les 18 à 24 mois, ce qui améliore la puissance de calcul tout en réduisant les coûts. Depuis l'invention du premier microprocesseur par Intel en 1971, l'augmentation exponentielle des performances, décrite par la loi de Moore, a

[105] *Étude Intel, juin 2024, « AI Everywhere, Redefines Power, Performance and Affordability ».*

permis une progression spectaculaire des capacités des ordinateurs. Cependant, depuis 2010, des ralentissements dans la production de puces par des leaders comme Intel, Samsung et TSMC remettent en question cette loi, certains suggérant de réviser son rythme à un doublement des performances tous les trois ans au lieu de deux.

Ce ralentissement s'explique notamment par les défis liés à l'augmentation de la fréquence des processeurs, devenue difficile à gérer en raison des problèmes de dissipation thermique. Alors qu'on envisageait à l'époque d'atteindre des fréquences de 10 GHz (représentant le nombre de cycles d'horloge par seconde, mesuré en milliards), la stratégie a évolué pour intégrer davantage d'unités de calcul ou de mémoire par puce (multi-cœur), tout en maintenant des fréquences stabilisées entre 2 et 3 GHz. Désormais, les processeurs incluent non seulement des cœurs CPU, mais aussi d'autres composants dans un même système intégré, appelé System on a Chip (SoC), qui combine des CPU, GPU, unités de réseau et même des accélérateurs spécialisés comme les TPU (Tensor Processing Units) et les NPU (Neural Processing Units). Cette évolution rend de plus en plus complexe la mesure des performances d'un processeur. Les architectures SoC, avec leurs multiples unités optimisées pour des tâches spécifiques, ne se prêtent plus à une évaluation unique basée sur la fréquence ou le nombre de transistors, nécessitant des critères plus adaptés pour évaluer leur efficacité dans des usages variés.

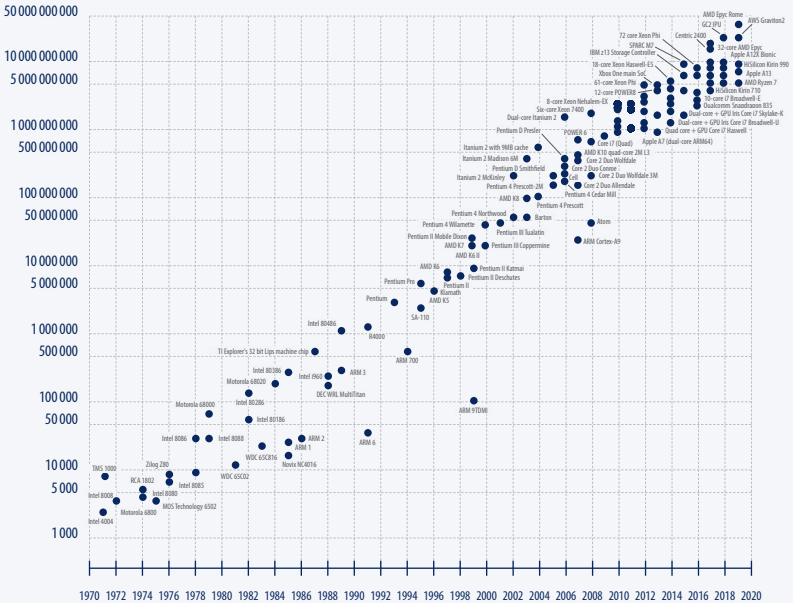
Le concept de More than Moore, en revanche, ne se concentre pas sur la taille des transistors, mais sur l'ajout de nouvelles fonctionnalités aux semi-conducteurs. Il s'agit d'intégrer des capacités telles que la détection, la communication ou la gestion d'énergie, pour répondre à des besoins technologiques

diversifiés. Tandis que More Moore vise à maximiser les performances des technologies CMOS (transistors traditionnels), More than Moore cherche à enrichir les puces pour des usages plus variés, comme l'IoT ou les dispositifs médicaux connectés.

Cependant, le développement de l'IA dépasse largement le rythme de la loi de Moore. Depuis 2013, la demande en puissance de calcul pour l'IA double tous les quatre mois[106], soit six fois plus vite que prévu par la loi de Moore. Par ailleurs, de nouvelles technologies émergent régulièrement, comme celle d'IBM, qui utilise un laser infrarouge pour séparer une puce de son support et atteindre des performances inédites. Ces avancées montrent que, bien que la loi de Moore reste une référence, il devient essentiel d'adopter une vision plus large, en s'interrogeant sur les progrès globaux en informatique. Ce regard élargi permettra d'anticiper les défis liés aux capacités futures de stockage et de traitement des données.

[106] *Banque Transatlantique, G.Ozon, novembre 2022, « La loi de Moore est morte, vive la loi de Huang ! ».*

Graphique n° 20 • Loi de Moore, le nombre de transistors sur les puces double tous le deux ans



La loi de Rock, quant à elle, met en lumière un défi majeur de la miniaturisation : le coût. Depuis 1971, le coût de production des puces les plus avancées a été multiplié par 18, selon le MIT. Plus les puces sont petites, plus les procédés de fabrication sont complexes et coûteux. Par exemple, seules les machines d'ASML peuvent produire des puces de moins de 5 nm à l'aide de rayons UV extrêmes, avec un prix unitaire de 180 M $. Malgré cela, Nvidia a lancé en 2023 une production de masse de puces de 3 nm grâce à des investissements massifs.

Concernant le calcul quantique, des incertitudes demeurent sur la performance des futurs calculateurs et sur la montée en puissance des technologies habilitantes associées[107]. Néanmoins, il est communément admis que ces machines transformeront durablement le marché du calcul haute performance (HPC), dans certains domaines adaptés pour le calcul quantique comme l'optimisation, la chimie ou encore la cryptographie. Dans un premier temps, elles coexisteront avec des systèmes de type NISQ[108] et des machines analogiques au sein des centres HPC. À plus long terme, l'arrivée de processeurs dits *« fault-tolerant quantum computing »* (FTQC) et le développement d'architectures modulaires capables d'interconnexion pourraient redéfinir certains processus industriels et élargir les capacités des centres de calcul.

La demande porte également sur la frugalité des méthodes de calcul des fournisseurs d'infrastructures numériques. À l'instar d'une voiture de sport, plus énergivore qu'une citadine, les infrastructures numériques consomment différemment selon leurs usages. Pour les usages peu intensifs, comme la visioconférence, 80 % des émissions proviennent du matériel, contre seulement 20 % liées à l'usage. En revanche, pour des usages intensifs comme l'IA générative, cette proportion s'inverse : 80 % des émissions proviennent de l'usage et seulement 22 % du matériel[109]. Ainsi, c'est l'utilisation des infrastructures numériques, plus que leur existence, qui constitue la principale menace environnementale.

La recherche et développement pour concevoir des infrastructures numériques à forte densité énergétique s'est fortement développée. En plus de l'évaluation en temps réel des sources

[107] *CEA, & Futuribles. (2024, Mars). Étude Quantum 2042 – Y.SPOT, https://yspot.fr/etude-quantum-2042/.*

[108] *Ordinateur quantique actuel, caractérisé par un nombre limité de qubits (généralement quelques dizaines à quelques centaines) et par une sensibilité élevée aux erreurs dues au bruit quantique.*

[109] *Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2023). Estimating the carbon footprint of bloom, a 176b parameter language model. Journal of Machine Learning Research, 24(253), 1–15.*

d'amélioration dans la fabrication et dans l'utilisation des infrastructures numériques, ces initiatives se concentrent sur la performance énergétique comme levier d'innovation. Sur les réseaux, des dispositifs comme les produits radios améliorés permettent de diminuer la consommation d'énergie, avec des entreprises comme Ericsson qui ont diminué de 30 % leur consommation d'énergie de stations basse radio de 2021 à 2023, soit une amélioration moyenne de 10 % par an. Pour compenser l'augmentation des émissions liées au traitement des données, les *hyperscalers* travaillent sur des innovations technologiques comme la fusion atomique (SMR) ou l'énergie géothermique, de nouveaux modes de refroidissement liquide par plaque froide au niveau composant en circuit fermé, de la réutilisation de chaleur fatale, de la pile à combustible à hydrogène, des innovations au niveau gestion de l'eau, l'utilisation de bois à la place du béton, etc. Des chercheurs de l'université de Santa Cruz travaillent actuellement sur une méthode scientifique pour faire fonctionner des modèles d'IA avec plusieurs milliards de paramètres sans énergie supplémentaire[110]. En outre, certains équipementiers utilisent le *machine learning* pour économiser l'énergie à l'aide de fonctionnalités optimisées. **Ces actions se sont toutefois davantage concentrées sur l'amélioration de la performance des infrastructures numériques et non sur leur usage**.

b. Une réponse a été apportée par la puissance publique mais elle demeure insuffisante au regard des besoins

Pour répondre aux besoins croissants en calcul intensif, la France a créé en 2007 le Grand équipement national de calcul intensif (GENCI), sous l'impulsion du ministère de la Recherche et de l'Enseignement supérieur. Le GENCI est une infrastructure stratégique axée sur la partie amont de la chaîne de valeur des infrastructures numériques,

[110] *Ils ont déjà réussi à faire fonctionner un modèle de type Llama-2 avec 13 watts d'énergie, soit l'équivalent de ce que consomme une ampoule LED.*

rassemblant les principaux acteurs de la recherche française, tels que le CEA, le CNRS et l'INRIA. Avant sa création, les chercheurs français souffraient d'un déficit en puissance de calcul, les obligeant souvent à se tourner vers des infrastructures dans des pays voisins comme l'Allemagne, l'Italie ou l'Espagne. Le GENCI met aujourd'hui à disposition gratuitement des moyens de calcul et d'accompagnement pour les organismes publics et privés, avec une puissance de calcul atteignant 240 pétaflops (PFlops) en 2024, qui est doublée chaque année. Le pétaflop est une unité de mesure qui correspond à un million de milliards d'opérations par seconde, ces opérations impliquant des nombres à virgule flottante[111], un quasi doublement par an depuis 2007, condition indispensable pour rester dans la course technologique. En 2023, plus de 3,5 milliards d'heures de calcul ont été mises à disposition par le GENCI, permettant de soutenir aussi bien la recherche fondamentale qu'appliquée. Ces ressources sont accordées sous condition de publication des résultats obtenus, sur le principe de la recherche ouverte.

En parallèle des besoins civils, le supercalculateur Exa, hébergé par la Direction des Applications Militaires (DAM) du CEA, répond aux exigences en puissance de calcul militaire de la France. Dans ce cadre, l'Agence des participations de l'État (APE) a racheté les supercalculateurs d'Atos, regroupés au sein de sa division Advanced Computing, pour un montant estimé entre 500 et 625 M€[112]. Par ailleurs, le ministère des Armées a récemment confié la réalisation d'un nouveau supercalculateur dédié à l'IA à un tandem formé par Orange et HPE.

Les capacités de calcul fournies par le GENCI jouent un rôle clé dans de nombreux domaines, allant de la science fondamentale à l'innovation technologique et à l'aide à la décision. Ces simulateurs numériques,

[111] *La virgule flottante permet de représenter des nombres très grands ou très petits avec une précision adaptée, ce qui est crucial pour des calculs complexes en physique, en climatologie ou en intelligence artificielle.*

[112] *Les Echos, A. Drif, 25 novembre 2024, « Coup de théâtre dans la vente des actifs stratégiques d'Atos à l'État ».*

extrêmement consommateurs de ressources, sont utilisés pour des applications variées telles que la modélisation du climat, la recherche en énergie, le développement automobile ou pharmacologique, et même pour des situations d'urgence comme les tremblements de terre ou la gestion des priorités de recherche pendant la crise du Covid-19. **En 2023, le GENCI a soutenu 1 500 projets de recherche, tous domaines confondus, dont près de 1 000 étaient liés à l'IA. Avec 1 400 projets en IA attendus d'ici 2025, cette tendance reflète l'explosion de l'IA dans des secteurs toujours plus diversifiés**.

Encadré n° 9 • Rappel des procédures pour accéder à de la puissance de calcul en France

1. Pour des usages IA, possibilité d'avoir un accès dynamique avec un appel à projet ouvert toute l'année permettant d'accéder à une machine en une semaine.

2. Fourniture d'un accès régulier à de la puissance de calcul en 6 mois selon la chronologie actuelle des procédures en place pour postuler sur la plateforme EDARI.

Le GENCI met à disposition des machines spécifiques, optimisées en fonction de leur dotation en CPU ou en GPU[113], conçues pour répondre de manière flexible et ciblée à des besoins variés.

[113] *Le GPU (Graphics Processing Unit ou processeur graphique) est un circuit électronique spécialisé initialement conçu pour traiter et afficher des graphiques, notamment dans les jeux vidéo et les applications 3D. Contrairement aux CPU (Central Processing Units), qui sont polyvalents et gèrent un large éventail de tâches informatiques, les GPU sont optimisés pour exécuter des calculs parallèles massivement.*

- **Le supercalculateur Jean Zay, situé au centre de calcul de l'IDRIS du CNRS à Orsay, est une plateforme nationale de calcul intensif dédiée à la simulation numérique et aux usages de l'IA**. Doté d'environ 4 000 GPU, Jean Zay a été mis en place à la suite des annonces présidentielles de 2019 visant à offrir une machine spécialisée dans l'IA à la recherche française. Cette infrastructure stratégique permet de répondre aux besoins croissants en calcul intensif, avec une mise à niveau annuelle pour suivre les avancées rapides des générations de GPU, contrastant avec le rythme traditionnel de mise à jour tous les six ans. **Jean Zay est capable de réaliser des calculs massivement parallèles[114], indispensables pour des simulations complexes et des entraînements de modèles d'apprentissage profond. À titre de comparaison, cette puissance permettrait de faire fonctionner une usine 4.0 sophistiquée ou d'exécuter simultanément plusieurs milliards de calculs dans des systèmes connectés, comme les véhicules autonomes**. Jean Zay est utilisé aussi bien par la recherche publique que par l'industrie. Du côté académique, des modèles linguistiques faisant partie du projet Albert ainsi que Bloom ont été entraînés grâce à cette infrastructure. Dans le secteur privé, des entreprises comme Valeo exploitent cette puissance pour développer des technologies avancées, notamment dans le domaine des voitures autonomes. Cette polyvalence fait de Jean Zay un pilier essentiel pour le développement scientifique et technologique en France.

[114] *Dans le contexte des supercalculateurs, effectuer massivement des calculs en parallèle signifie diviser une tâche complexe en de nombreuses sous-tâches, exécutées simultanément sur des milliers, voire des millions, de cœurs de calcul, afin d'accélérer le traitement et de résoudre des problèmes d'une grande envergure.*

Graphique n° 21 • Panorama des acteurs privés entraînant leurs modèles sur Jean Zay, une diversité unique au monde et reconnue par les plus gros acteurs

- **Le supercalculateur Adastra, installé au Centre Informatique National de l'Enseignement Supérieur (CINES) à Montpellier, est équipé de processeurs CPU et GPU fournis par AMD**. Ce supercalculateur se distingue par son système de refroidissement à l'eau chaude, une technologie novatrice qui offre des performances énergétiques nettement supérieures aux systèmes de refroidissement à l'air traditionnels. Adastra est l'un des rares supercalculateurs au monde, avec LUMI en Finlande, à adopter ce type de solution. Cette approche technologique constitue un différenciateur stratégique pour Adastra, notamment vis-à-vis des *hyperscalers*, qui privilégient souvent des volumes massifs de cycles de calcul à faible coût.

- **TGCC Joliot-Curie, situé au sein du centre de calcul du CEA, est aussi une infrastructure clé pour le calcul intensif en France**. Ce centre évoluera prochainement avec l'installation du supercalculateur de classe exascale Alice Recoque (anciennement Jules Verne) d'ici deux ans. **Doté de 25 000 GPU, Alice Recoque apportera une capacité de calcul significative pour répondre aux besoins croissants en simulation et en apprentissage profond dans la recherche et l'industrie**. Joliot-Curie joue également un rôle important dans la mise en œuvre de la stratégie nationale quantique. **Il héberge déjà un processeur développé par Pasqal, spécialisé dans le calcul quantique analogique, et accueillera l'année prochaine un processeur de Quandela, processeur quantique basé sur de la photonique**. Les domaines prioritaires identifiés à ce stade se concentrent sur des applications dans des domaines tels que la chimie, les matériaux et l'optimisation, où le calcul quantique peut compléter les capacités des supercalculateurs traditionnels.

Encadré n° 10 • Quelles différences entre les ordinateurs dits « exaflopiques » et les ordinateurs quantiques ?

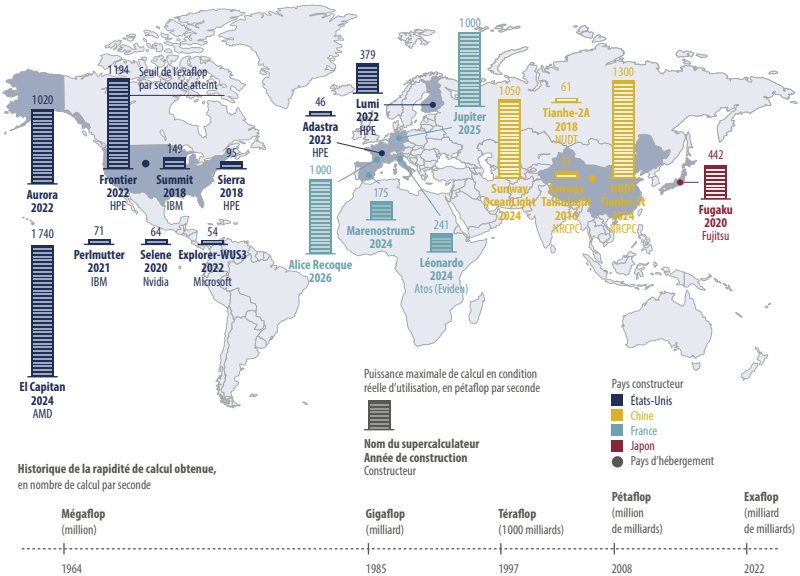
Dans le cas d'ordinateurs dits « exaflopiques », comme Alice Recoque d'EuroHPC et GENCI/CEA qui entrera en service en 2026 en France, ou le Frontier doté de 1,2 exaflops et situé au Oak Ridge National Laboratory au Tennessee, cela signifie que l'ordinateur en question est capable d'effectuer en une seconde autant de calculs que plusieurs millions d'ordinateurs personnels combinés pourraient en réaliser en une année. En termes scientifiques, cette capacité équivaut à un exaflop, soit un milliard de milliards (ou 10^18) d'opérations en virgule flottante par seconde.

Avec l'arrivée de l'ordinateur quantique, l'échelle est tout autre, car l'unité de mesure de puissance est le qubit, qui utilise le principe de superposition[115] pour traiter les calculs de manière exponentielle. Si un ordinateur quantique atteint le pallier de 128 qubits logiques fixé par Proqcima en 2032, cela signifie qu'il peut théoriquement représenter 2^128 états différents en simultané. **Autrement dit, pour certains calculs spécifiques, un ordinateur quantique de 128 qubits logiques pourrait résoudre en quelques secondes ce qui prendrait des milliers, voire des millions d'années à un supercalculateur classique exaflopique**.

Le site du TGCC/CEA joue un rôle central pour maintenir la France dans la compétition mondiale en matière de puissance de calcul. **Grâce aux investissements réalisés dans le cadre du programme EuroHPC, il accueillera prochainement des capacités de calcul de type exascale, offrant une puissance 50 fois supérieure à celle disponible actuellement**. EuroHPC, une co-entreprise basée au Luxembourg, regroupe la Commission européenne et les États membres de l'Union européenne pour développer des infrastructures de calcul intensif et quantique dotées de capacités de calcul exascale. Ce programme vise à renforcer l'autonomie technologique européenne tout en positionnant ses infrastructures parmi les plus performantes au monde. En outre, *l'European Processor Initiative* (EPI) vise à doter les supercalculateurs européens de processeurs fabriqués en Europe, fournis par l'entreprise SiPearl qui développe des CPU basés sur la technologie Arm.

[115] *Le principe de superposition décrit la capacité d'une particule quantique (comme un électron ou un photon) à exister simultanément dans plusieurs états possibles tant qu'on ne l'a pas mesurée. Ce principe permet aux ordinateurs quantiques de traiter de nombreuses possibilités en parallèle, rendant certains calculs infiniment plus rapides que dans un système classique. C'est ce qui leur donne leur puissance, notamment pour les applications qui nécessitent l'exploration de vastes ensembles de solutions, comme dans la cryptographie et l'optimisation.*

Graphique n° 22 • Positionnement de la France dans le marché des supercalculateurs en 2023, avant la construction de Alice Recoque



Pour rappel, 1 Pflop = 10^15 opérations vs 1 EFlop = 10^18 soit un milliard de milliards.

Source : Le Monde actualisée avec les données 2025 GENCI.

Les projets européens de supercalculateurs, comme Jupiter en Allemagne et Alice Recoque en France, visent à positionner l'Europe sur un pied d'égalité avec les États-Unis en termes de puissance publique de calcul d'ici 2026[116]. **Avec trois supercalculateurs exascale prévus en Europe et trois déjà opérationnels aux États-Unis, l'écart semblait se réduire, avant les annonces relatives au projet Stargate de**

[116] *En effet, avec 2 supercalculateurs exascale prévus pour 2026 (Jupiter et Alice Recoque) et 3 supercalculateurs à un stade pré-exascale existants, cela permet d'envisager une puissance équivalente à 3 supercalculateurs exascale pour l'Europe en 2026, aux côtés des 3 supercalculateurs exascale américains existants.*

l'ordre de 500 Mds $. Toutefois, cette vision est limitée et ne tient pas compte de l'augmentation rapide de la puissance de calcul mondiale, où le secteur privé joue un rôle déterminant.

c. La réponse étatique doit désormais se compléter d'une réponse privée pour suivre la concurrence internationale

La véritable fracture entre l'Europe et les États-Unis réside dans la puissance de calcul rendue disponible par des acteurs privés, qui constitue une part essentielle des capacités mondiales : en 2026, les États-Unis devraient cumuler environ 450 000 GPU, contre seulement 50 000 en Europe, soit un rapport de 1 à 9[117]. L'Union euroéenne accuse un lourd retard par rapport aux États-Unis et à la Chine en matière de puissance de calcul[118]. En effet, alors que la puissance mondiale cumulée des supercalculateurs augmente de près de 40 % par an, sans compter l'arrivée potentielle d'un ordinateur quantique, l'Europe n'héberge que 24 % de la puissance de calcul mondiale, contre 53 % pour les États-Unis, et dispose de seulement 22,3 % des centres de données, contre 37,8 % pour les États-Unis.

Cette disparité s'explique par l'absence d'acteurs privés européens capables de fournir une infrastructure compétitive, à l'exception d'un ou deux acteurs tels qu'Iliad, pour un continuum fort public/privé de la recherche fondamentale à la commercialisation de solutions. Alors que des entreprises américaines comme NVIDIA et Tesla déploient des dizaines de milliers de GPU pour des projets stratégiques, les initiatives européennes restent trop modestes. Par exemple, Scaleway, bien qu'ayant augmenté sa capacité de 2 000 à 5 000 GPU, est encore loin des besoins pour compléter un Alice Recoque à 25 000 GPU.

[117] *Source : données GENCI.*

[118] *Commission Nationale de l'IA, mars 2024, IA : notre ambition pour la France.*

Une véritable transition vers le calcul intensif privé nécessiterait des investissements de l'ordre de 2,5-3 Mds €, soutenus par des initiatives publiques comme le projet Cluster de Paris-Saclay, souvent cité par la DG CONNECT, mais qui manque d'un relais privé suffisant. Cette faiblesse du secteur privé européen a des conséquences directes sur les usages stratégiques. Alors que la demande explose dans des domaines comme l'IA générative, la recherche scientifique et les industries de pointe, l'Europe risque de rester dépendante des infrastructures non européennes pour répondre à ces besoins critiques.

Cette situation appelle à des approches différenciées pour les acteurs publics et privés. Du côté des acteurs publics, il est essentiel de préparer l'après Jupiter et Alice Recoque en adoptant une vision à long terme sur les investissements nécessaires, tout en tenant compte des perspectives offertes par le calcul quantique et des usages qu'il pourra desservir. **À ce titre, le projet CLUSSTER (Cloud Unifié Souverain de Services, de TEchnologies et d'infrastructuRes) porté par ATOS/Eviden associant pour la première fois acteurs publics (dont GENCI, CEA, CNRS, Inria) et privés (ATOS, OVH Cloud, Qarnot, CS), labellisé avec Bpifrance dans le cadre de l'appel à projet Cloud Souverain, joue un rôle clé en fédérant des projets de recherche ouverte sur les infrastructures numériques publiques et en soutenant l'offre commerciale grâce à des infrastructures privées**.

Pour les acteurs privés, l'enjeu réside dans le développement d'une offre européenne intégrée, allant de la production des infrastructures numériques jusqu'à leur exploitation. Cela passe par des investissements significatifs dans les GPU, afin de répondre à la demande croissante en puissance de calcul dans le marché de l'exaflop. Cela passe aussi par de la mise à disposition de puissance de calcul modulaire, soit des ressources de calcul évolutives, et agrégée aux côtés de dispositifs de type Adastra dans le marché du pétaflop. **En effet, la bataille pour les entreprises privées ne concerne pas uniquement l'exaflop mais aussi les usages à forte valeur ajoutée. Ces usages nécessitent des**

machines pétaflopiques capables de traiter des données de type « synthétique », soit les données d'entraînement générées par des supercalculateurs, et les biais associés à ces données. Il convient de rappeler qu'il y a 497 supercalculateurs de capacité pétaflopique et inférieure dans le TOP 500 en novembre 2024. En effet, de nombreuses entreprises, membres du CAC 40, ETI ou PME n'ont pas besoin de capacités de type exascale pour réaliser leurs calculs intensifs.

Encadré n° 11 • Focus sur le supercalculateur Gefion

Le supercalculateur Gefion est une initiative danoise visant à renforcer les capacités américaines en intelligence artificielle (IA). Financé par la Fondation Novo Nordisk, enrichi par le succès des médicaments amaigrissants Ozempic et Wegovy, ce projet représente un investissement de 100 M $.

Gefion est équipé de 1 528 unités de traitement graphique (GPU) fournies par Nvidia, ce qui équivaut à la dernière partition de calcul installée par GENCI sur Jean Zay en 2024. Il est destiné à soutenir les entreprises et les chercheurs danois dans des domaines tels que la santé, la biotechnologie et l'informatique quantique, en leur offrant une puissance de calcul avancée pour surmonter les obstacles liés à la capacité de traitement.

Le lancement de Gefion a été marqué par la présence de Jensen Huang, PDG de Nvidia, et de membres de la famille royale danoise, soulignant l'importance nationale de ce projet. Ce supercalculateur est conçu pour accélérer les progrès technologiques et scientifiques en fournissant des ressources computationnelles sans précédent aux acteurs locaux.

Pour les acteurs du *cloud* privés cherchant à se différencier sur l'IA, l'un des axes les plus prometteurs est le développement de modèles spécialisés dits *distillés*. Contrairement aux grands modèles de langage généralistes, ces modèles sont conçus comme des versions optimisées, adaptées à des tâches spécifiques. Grâce au processus de *knowledge distillation*, ils conservent une partie des performances des modèles initiaux tout en étant allégés, ce qui réduit leur consommation énergétique et améliore la latence. Cette approche est essentielle dans un contexte où l'efficacité de l'inférence ne repose plus sur la seule puissance brute, mais sur une optimisation fine de toute la *stack* technologique.

L'intérêt des modèles distillés est particulièrement évident pour des applications nécessitant un passage à l'échelle sans explosion des coûts. Par exemple, les derniers modèles d'IBM à 2 milliards de paramètres, appartenant à la catégorie des *Small Language Models* (SLM), ont été conçus pour minimiser les risques d'hallucination tout en maintenant un coût d'inférence soutenable. **De même, *DeepSeek* a misé sur une approche alternative en limitant la redondance des données d'entraînement, privilégiant des jeux de données plus ciblés plutôt qu'un volume massif nécessitant des infrastructures lourdes comme les 500 000 processeurs H100 de Nvidia**. Ce travail d'optimisation passe notamment par une méthode dite de « quantization », une technique qui réduit l'empreinte mémoire en adaptant le nombre de bits nécessaires au codage des informations, notamment dans les modèles impliquant des éléments physiques. Cette réduction de précision, ajustée selon les besoins réels, est optimisée par des accélérateurs capables d'ajuster dynamiquement le nombre d'opérations. **Ces optimisations ont eu un impact tangible comme l'illustre le fait que *Mistral AI* soit passé d'un besoin de 8 GPU à 4 GPU après quantization, réduisant drastiquement les coûts d'exécution et les besoins en puissance de calcul**.

Un modèle conçu pour un cluster universitaire ne peut toutefois pas être transposé tel quel dans un environnement industriel ou commercial sans repenser l'architecture sous-jacente. L'exemple du projet de l'Université d'Alabama à Huntsville avec IBM illustre bien cette contrainte. **Dans un cadre académique, le développement d'IA repose souvent sur des clusters HPC traditionnels, où l'objectif est d'optimiser la précision des modèles sans contrainte immédiate de consommation énergétique ou d'intégration dans une infrastructure cloud existante. Or, dans un environnement industriel, ces modèles doivent être déployés sur des architectures optimisées en termes de consommation d'énergie, de latence et de compatibilité matérielle**. L'initiative conjointe entre IBM et UAH s'attaque précisément à ce problème avec l'utilisation des puces *SyNAPSE*, qui imitent le fonctionnement du cerveau humain pour exécuter des tâches d'IA avec une consommation énergétique jusqu'à 100 fois inférieure à celle des architectures classiques basées sur GPU ou CPU.

Les acteurs du *cloud* privés doivent relever plusieurs défis techniques pour exploiter pleinement le potentiel des modèles distillés. Le premier concerne la sélection et la préparation des données d'entraînement, qui doivent être choisies avec soin pour limiter les biais tout en garantissant une acceptabilité large des résultats produits. Ensuite, le *reinforcement learning* joue un rôle clé pour affiner les modèles en fonction des cas d'usage spécifiques. Enfin, l'infrastructure doit être optimisée pour éviter les Goulets d'étranglement, en tenant compte des différences fondamentales entre HPC et IA : les jeux de données, les schémas d'accès mémoire et les architectures matérielles diffèrent profondément entre ces deux domaines.

d. *A minima*, un doublement de nos ambitions en puissance de calcul est à sécuriser d'ici à 2030

Comme au début des réseaux télécoms, ne pas rater le virage de l'intelligence artificielle implique de se doter, à l'échelle européenne, des moyens en puissance de calcul à la hauteur de nos ambitions, avec une participation impérative et complémentaire du secteur privé. Selon le TOP 500, la puissance de calcul mondiale, publique et privée, a progressé à un rythme annuel de 39 % entre juin 2020 et juin 2024. Si cette tendance se maintenait – une hypothèse conservatrice – la puissance de calcul mondiale passerait de 8,2 exaflops aujourd'hui à 15,8 exaflops en 2026, et atteindrait 59 exaflops en 2030.

Avec l'entrée en service des supercalculateurs Jupiter et Alice Recoque en 2026, la part de l'Europe atteindrait 19 % de la puissance mondiale à cette date. Cependant, sans une accélération significative des investissements en calcul intensif, cette part chuterait à seulement 5 % en 2030. **Pour éviter un tel recul, il est indispensable de lancer dès à présent la construction d'au moins six nouveaux supercalculateurs exaflopiques ou post exaflopiques, portant ainsi la capacité européenne à 9 exaflops, soit environ 15 % des 59 exaflops prévus en 2030**.

Il convient de noter que c'est le strict minimum pour rester compétitif sur le marché du calcul intensif. Être plus ambitieux impliquerait de lancer dès aujourd'hui la construction de 12 nouveaux supercalculateurs exaflopiques – ou l'équivalent en puissance avec des supercalculateurs pétaflopiques – ce qui permettrait de hisser l'Europe à 25 % de la part mondiale de puissance de calcul en 2030. Les secteurs qui en bénéficieraient le plus seraient l'IA,le *big data* et la recherche scientifique.

Tableau n° 2 • Estimation des besoins futurs en puissance de calcul en Europe (2024 – 2030)

	2024	2026	2030	# nouveaux supercalculateurs requis
Puissance de calcul du TOP 500 en exaflops	8,2	15,8	59,0	
Rythme constant	**12 %**	**19 %**	**5 %**	**0**
# exaflops	1	3	3	
Rythme *a minima*	**12 %**	**25 %**	**15 %**	**6**
# exaflops	1	4	9	
Rythme ambitieux	**12 %**	**43 %**	**25 %**	**12**
# exaflops	1	7	15	

Note pour le lecteur : le scénario *a minima* s'appuie sur le taux de croissance annuel de la puissance de calcul disponible mondialement de juin 2020 à juin 2024 (source : Top500).

Encadré n° 12 • Focus sur le post exascale, une nouvelle frontière pour la puissance de calcul

Le post-exascale marque une avancée décisive dans l'évolution des supercalculateurs, visant à dépasser une puissance de calcul de 10^{21} FLOPS (flops par seconde). Cette progression repose sur la convergence de trois axes technologiques : les architectures classiques à haute performance (HPC), les systèmes d'intelligence artificielle (IA) intégrés, et des approches innovantes comme les processeurs neuromorphiques et quantiques. Par exemple, la puce Loihi 2 d'Intel illustre les avancées en architectures neuromorphiques, imitant le fonctionnement des réseaux

neuronaux biologiques pour traiter efficacement des tâches d'IA complexes avec une consommation énergétique minimale.

Dans cette perspective, des projets comme Frontier, au Oak Ridge National Laboratory (ORNL), montrent comment les processeurs quantiques sont ajoutés à des infrastructures HPC existantes pour résoudre des problèmes jusqu'alors inaccessibles avec des architectures classiques. De même, IBM a intégré son Q System One, un ordinateur quantique fonctionnel, dans des projets visant à évaluer la complémentarité entre calcul quantique et classique, ouvrant la voie à des applications hybrides particulièrement prometteuses dans les domaines de la simulation et de l'optimisation.

Par ailleurs, des initiatives comme le supercalculateur Fugaku au Japon intègrent nativement des modèles d'IA au sein des machines HPC pour des applications de grande envergure, comme la simulation climatique. Cette combinaison IA-HPC permet d'accélérer les calculs tout en améliorant la précision des prévisions, soulignant le rôle central de la convergence entre ces technologies pour relever les défis scientifiques, industriels et sociétaux du post-exascale. Ces efforts internationaux illustrent l'importance pour l'Europe de s'inscrire dans cette dynamique afin de ne pas dépendre uniquement des avancées américaines et asiatiques.

Recommandation 2

Entamer, dès aujourd'hui, *a minima*, en France, la construction de 6 supercalculateurs exaflopiques additionnels afin de proposer à l'Europe une capacité de calcul de 9 exaflop.

Le besoin en puissance de calcul, recherche et industries incluses est estimé à 15 exaflops d'ici à 2030, soit une part de marché mondiale de 25 % pour l'Europe. Cela correspond à une base installée de 15 supercalculateurs exaflopiques nécessitant donc la construction de 12 supercalculateurs additionnels européens.

Nous préconisons le strict minimum pour rester dans la course mondiale, compte tenu de nos marges de manœuvre limitées, soit 6 supercalculateurs exaflopiques additionnels. Cela permettrait à l'Europe de sécuriser 15 % de la puissance de calcul mondiale.

Les avis divergent quant à la nécessité de confier cette ambition à l'Europe ou à la France en particulier.

L'estimation budgétaire nécessaire pour construire ces infrastructures repose sur une analyse comparative des projets existants. **Le supercalculateur Alice Recoque, installé en 2026 sur le site TGCC du CEA, a été cofinancé par EuroHPC à hauteur de 540 M € sur cinq ans dans le cadre d'un partenariat avec la France (GENCI) et les Pays-Bas**. Sur cette base, et en prenant en compte l'évolution des coûts liés à l'approvisionnement en GPU et à l'échelle des infrastructures, le financement nécessaire est estimé comme suit : **3 Mds € correspondant à 6 supercalculateurs exaflopiques et 6 Mds € correspondant à 12 supercalculateurs exaflopiques**. Ces estimations reposent sur une répartition équilibrée des financements : 50 % issus des crédits EuroHPC et 50 % provenant de fonds publics et privés, notamment des entreprises européennes du *cloud* prêtes à investir dans des infrastructures stratégiques.

Recommandation 2.1 : confier à la France la construction de 6 supercalculateurs sur le territoire national, le reste pouvant être réparti sur le territoire européen, en capitalisant sur nos capacités existantes (composants, compétences et énergie).

La répartition concertée des responsabilités entre le public et le privé sera une condition de succès.

Les infrastructures publiques doivent conserver leur rôle central dans l'entraînement de modèles de grande échelle, en s'appuyant sur des initiatives européennes.

Les acteurs du *cloud* privé ne doivent pas limiter leurs investissements à l'entraînement de modèles d'IA, une tâche déjà prise en charge par des infrastructures publiques comme celles d'Euro HPC ou du GENCI en France. **Ils doivent se concentrer sur l'industrialisation et l'inférence à grande échelle des modèles open-source ou propriétaires, en les adaptant à des usages spécifiques**. Leur rôle est également crucial pour fournir des infrastructures d'inférence dédiées aux besoins métiers de communautés variées, tout en répondant à des exigences de performance et de confidentialité. **Une attention particulière pourrait être portée aux petits modèles de langages distillés dits « SLM », qui sont plus légers et optimisés que les grands modèles, et permettent de gagner en efficacité sur l'inférence. Sur ce type de modèles, ce sont les modalités de sélection des données de base et leur préparation qui feront toute la différence**.

Recommandation 2.2 : mettre en place une Gouvernance efficace qui incite des acteurs comme Outscale (Dassault Systèmes), Scaleway (Iliad) ou OVH Cloud à investir dans des machines exascale dédiées aux usages commerciaux, prenant le relais des infrastructures publiques axées sur la recherche. L'objectif central est de faire émerger des acteurs français capables d'exporter une capacité de calcul suffisante pour être compétitif. Les comités stratégiques de filière (CSF) sont des outils intéressants et ayant une vocation éminemment stratégique, mais ils peinent à se déployer

et souffrent d'une lenteur d'exécution pénalisante faute de bénéficier de l'intérêt politique nécessaire.

2.2. DES BESOINS FORTS DE DÉVELOPPEMENT DE *DATA CENTERS* QUI SONT PARTIELLEMENT ADRESSÉS FAUTE D'UNE PLANIFICATION STRATÉGIQUE

Le pendant de cette capacité de calcul massive réside dans les *data centers*, infrastructures essentielles pour soutenir les stratégies disparates menées sur l'IA, le quantique et la cybersécurité au plus haut niveau de l'État, tout en répondant aux besoins qu'elles génèrent.

Les *data centers*, conçus dès les années 1960-1970 par des pionniers comme IBM, centralisent le stockage et le traitement des données en regroupant serveurs, systèmes de stockage et équipements réseau. **La France compte 250 *data centers* et leur nombre croît de 13 % à 14 %[119] par an, avec une croissance de 18 % par an prévue à partir de 2030, selon France Data Center. En termes de puissance, 600-700 MW de data centers sont opérationnels aujourd'hui, ce qui représentera entre 1,5 et 1,9 GW en 2030[120]**.

Le marché français des *data centers* se divise en deux catégories : les grands *data centers* internationaux et les *data centers* de proximité gérés par des PME, ETI ou opérateurs spécialisés. Les premiers, exploités par des acteurs comme Equinix, Digital Realty ou Data4, hébergent

[119] *EY, T. Solelhac, A. Gonnet, 14 septembre 2023, Baromètre : « Les datacenters, un marché en forte croissance, qui investit fortement dans l'économie française », https://www.ey.com/fr_fr/insights/tmt/premier-barometre-de-la-filiere-des-datacenters.*

[120] *NB que la puissance se compte en W et la consommation en Wh, et que de nombreuses entreprises utilisent des salles informatiques assimilables à des data centers, ce qui complique la réalisation d'une cartographie exhaustive.*

les données des *hyperscalers* (Google, Amazon, Microsoft, etc.) et de grandes entreprises et traitent d'importants volumes de données à l'échelle mondiale, avec des capacités énergétiques de 10 à 130 MW[121], connectées au réseau haute tension transporté par RTE. Les seconds, opérés par des entreprises comme Telehouse ou Thésée Datacenter, répondent à des besoins locaux avec une puissance moindre (environ 10 MW) et un raccordement au réseau de distribution d'Enedis. Adaptés à des secteurs comme la santé ou la recherche, ils permettent, par exemple, à un CHU d'héberger localement des données critiques telles que des dossiers médicaux ou des images médicales.

Encadré n° 13 • Différents paliers de disponibilité et de sécurité des *data centers*

La classification en tiers permet de certifier la sécurité, les capacités de maintenance, de disponibilité et atteste de la qualité des choix technologiques et énergétiques.

Tableau n° 3 • Estimation des besoins futurs en puissance de calcul

(2024 - 2030)

Classification	Redondance[122]	Indisponibilité annuelle	Usage
Tier 1	Aucune	28 heures	Acteur pouvant se permettre une défaillance temporaire, rapide à déployer

[121] *Le plus grand data center en France, situé à La Courneuve, affiche une puissance de 130 MW.*

[122] *La redondance appliquée aux centres de données signifie que les services de base et les systèmes auront des doublons (équipement, liaisons, alimentation et chemins, données, logiciels…) afin de garantir les fonctionnalités dans l'éventualité où l'un de ces composants s'avérerait défaillant.*

Classification	Redondance	Indisponibilité annuelle	Usage
Tier 2	Partielle (certains composants électroniques)	22 heures	Acteurs de taille intermédiaires
Tier 3	Redondance pour chaque équipement d'origine (N+1)	1,6 heure	Données confidentielles à forte exigence de disponibilité
Tier 4	Double redondance pour chaque équipement d'origine (2N+1)	48 minutes	Données d'importance vitale

À ce jour, la majorité des DC sont Tiers 3.

Source : Data4.

Aujourd'hui, les *data centers* sont de plus en plus refroidis à l'eau pour améliorer leur efficacité énergétique et réduire leur empreinte carbone. Contrairement au refroidissement à l'air, le refroidissement liquide permet une dissipation thermique plus efficace, essentielle face à la densité croissante des serveurs et à la montée en puissance de l'IA. Cette approche limite aussi la consommation d'énergie des systèmes de climatisation traditionnels, rendant les infrastructures plus durables.

Encadré n° 14 • Précisions sur la puissance et la consommation électrique des *data centers*

La plupart des *data centers* de grande capacité qui se développent depuis quelques années sont conçus pour pouvoir être partagés entre différents acteurs et pour accompagner la croissance du marché. Ils ont vocation à héberger des capacités de calcul au sein de bâtiments offrant des garanties de sécurité, de continuité d'alimentation en électricité, d'efficacité énergétique… Les autorisations administratives et les raccordements au réseau électrique à très haute tension sont demandés sur la base de la puissance électrique correspondant à leur taux maximal de remplissage, qui ne pourra être atteint qu'au bout de 7 à 10 ans d'exploitation. Au départ, ce sont donc des coquilles vides (« shell ») qui se remplissent peu à peu de serveurs et dont la consommation électrique croît proportionnellement à leur remplissage.

La « puissance électrique » d'un *data center*, exprimée en MW, doit être considérée comme le « débit maximal » d'électricité que le datacenter pourra utiliser, une fois pleinement occupé et en période de forte activité. Ainsi, un *data center* dimensionné pour 100 MW, peut ne « souscrire » qu'une puissance de 30 MW, s'il est certain de ne pas dépasser cette puissance dans les prochains mois. Pour autant, en l'état actuel des règles, RTE doit être en mesure de lui allouer les 100 MW initiaux s'il les demande. La consommation d'électricité est pour sa part le produit de la puissance par le nombre d'heures d'utilisation. Sur 24h, un *data center* fonctionnant à 30 MW consommera donc 720 MWh, et près de 260 000 MWh sur une année.

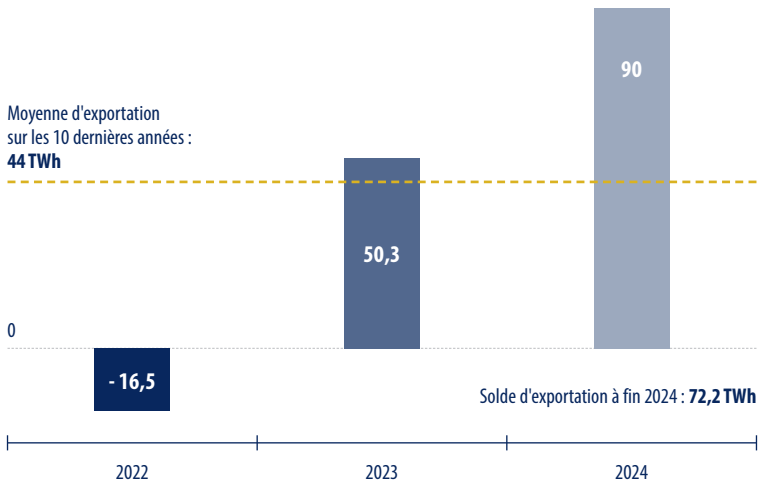
À fin 2024, sur une dizaine de *data centers* raccordés au réseau à très haute tension et totalisant une puissance d'environ 1 000 MW, RTE estimait la puissance réellement utilisée à environ 100 MW, soit 10 %. Sur une année entière, cela représente une consommation de l'ordre de 0,85 TWh – la consommation française s'élevant à environ 450 TWh.

a. Dans les 5 prochaines années, la possibilité inexploitée de mobiliser plus stratégiquement nos ressources énergétiques

La production d'électricité est excédentaire en France quel que soit le scénario de consommation identifié par les acteurs. En 2024, la France a exporté près de 90 TWh d'électricité, un record historique, avec une moyenne de 14 GWh exportés quotidiennement vers ses voisins. **En effet, la France est redevenue exportatrice nette depuis 2023 grâce à un très net redressement de sa production d'énergie nucléaire et à un développement rapide de ses énergies renouvelables**. C'est un facteur d'attractivité notable pour la France, qui lui permet d'être une « terre d'accueil » de *data centers*, comme l'ont bien illustré les annonces pendant le Sommet de l'Action de l'IA qui s'est tenu à Paris en février.

Graphique n° 23 • Évolution des exportations nettes de la France en électricité depuis 2022

(exportations nettes de la France en TWh)



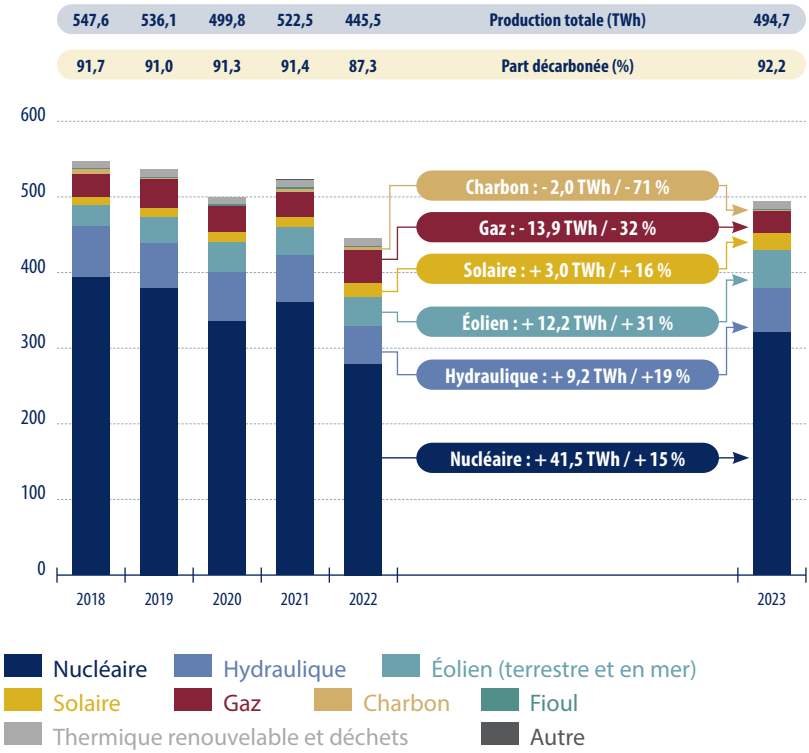
Source : RTE.

En 2023, 494,7 TWh d'électricité ont été produites, alors que 446 TWh ont été consommées. **À plus long-terme, dans chaque scénario, la production électrique est capable de répondre à la surconsommation potentielle dans de bonnes conditions technico-économiques**. RTE a retenu les variables suivantes pour construire ses scénarios : l'efficacité énergétique, la sobriété, le développement des énergies renouvelables et la maximisation de la production nucléaire. Ces leviers permettent d'évaluer différents chemins vers la décarbonation et la réindustrialisation, en intégrant des marges de sécurité pour éviter les risques de sous-dimensionnement du système.

Éléments sur la disponibilité de l'énergie électrique produite en France

Graphique n° 24 • Évolution de la production totale d'électricité par filière et part de production décarbonnée

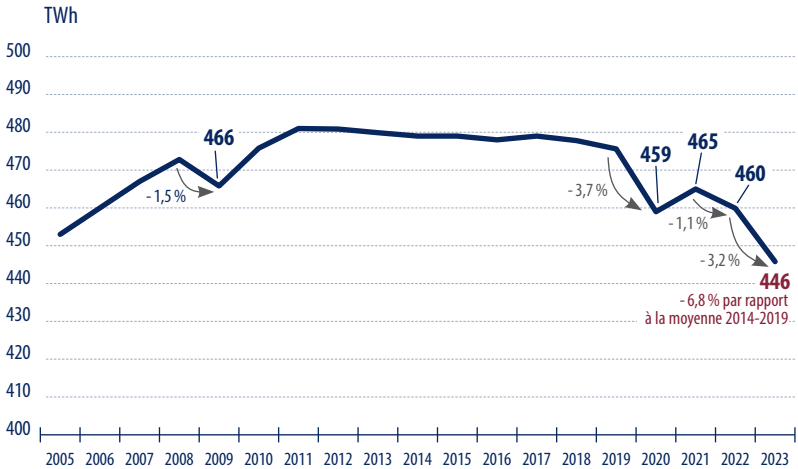
(en France entre 2018 et 2023)



Note : la production à partir de déchets ménagers est considérée renouvelable à 50 %. La production hydraulique est retranchée de 70 % de la consommation de pompage des STEP selon la Directive européenne 2009/28/CE.

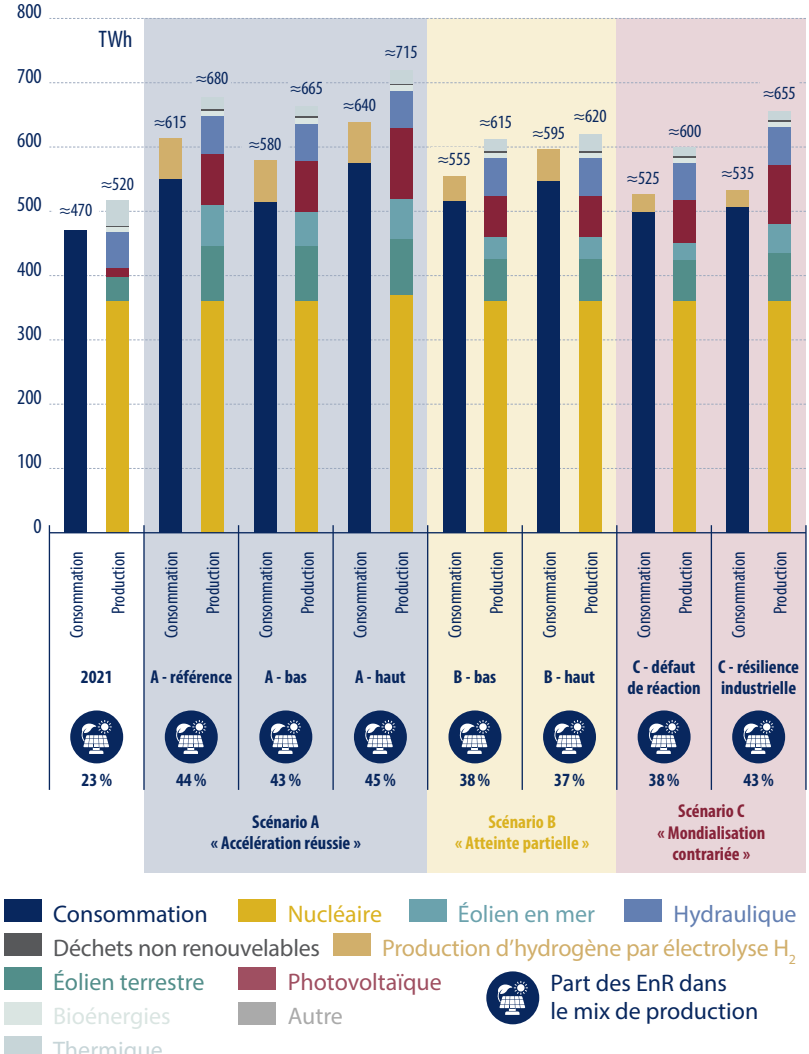
Source : bilan électrique 2023 de RTE.

Graphique n° 25 • Évolution entre 2005 et 2023 de la consommation corrigée des effets météorologiques et calendaires



Source : bilan électrique 2023 de RTE.

Graphique n° 26 • Bouclage et mix à 2035 des différents scénarios



Source : scénarios du bilan prévisionnel 2023 de RTE.

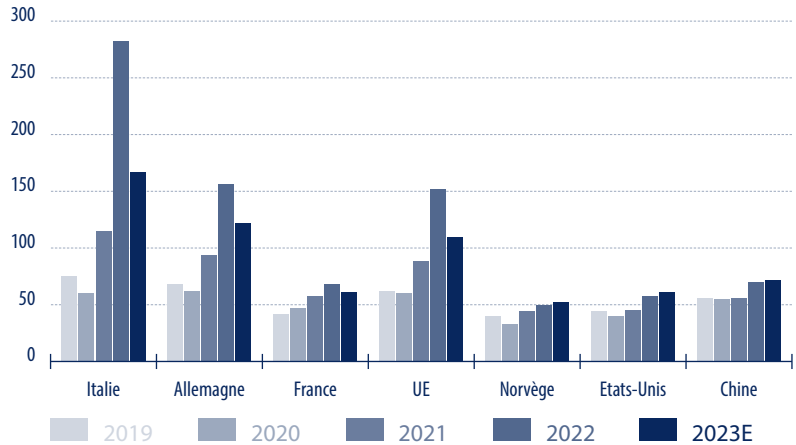
Bien que le Bilan Prévisionnel 2023-2035 de RTE[123] n'ait pas explicitement intégré la croissance de l'IA générative, il prévoit une augmentation de la consommation électrique en France, estimée entre 580 et 640 TWh/an d'ici 2035, contre 460 TWh actuellement. Cette hausse est notamment attribuée à l'électrification accrue de secteurs tels que la mobilité, l'industrie et les *data centers*. Cependant, le coût de l'électricité en France demeure élevé par rapport à d'autres pays. **Au premier trimestre 2024, le prix moyen pour les ménages français était de 0,28 €/kWh, tandis qu'il était de 0,16 €/kWh aux États-Unis, 0,197 €/kWh au Japon et 0,07 €/kWh en Chine[124]. Cette différence de coût pose la question de la compétitivité énergétique de la France, malgré sa production nucléaire**. Bien que le parc nucléaire français assure une production d'électricité bas carbone et stable, les coûts associés à la maintenance, au renouvellement des infrastructures et à la gestion des déchets nucléaires influent sur le prix final de l'électricité. Ainsi, la capacité du nucléaire à garantir des tarifs compétitifs face à des pays aux sources d'énergie diversifiées reste un défi majeur pour la France.

[123] RTE, 28 juillet 2024, Bilan Prévisionnel 2023 -2035 – La Consommation.

[124] Selon une étude de Verivox utilisant les données de GlobalPetrolPrices de 2021.

Graphique n° 27 • Estimation du prix final de l'électricité pour les grands clients industriels dans les secteurs à forte consommation d'énergie

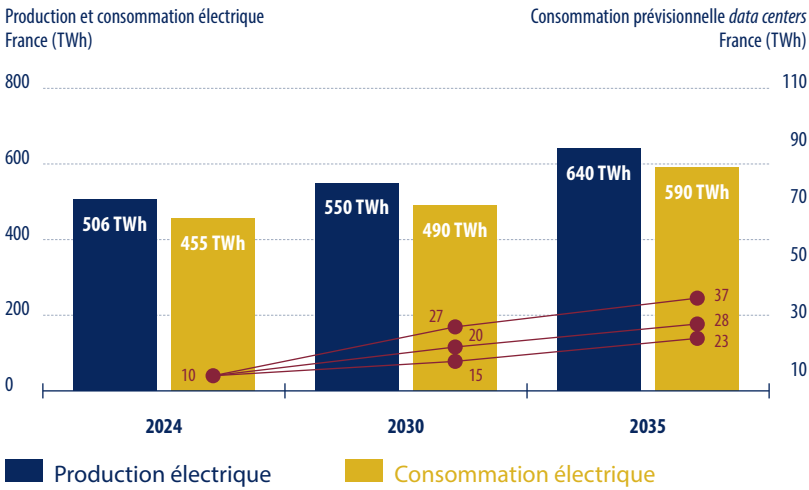
(2019-2023)



Note de lecture : l'analyse porte sur les coûts de l'électricité dans les industries dont la consommation annuelle d'électricité est supérieure à 150 GWh pour les pays européens, sur la base des données d'Eurostat. La compensation des prix de l'électricité est incluse pour les pays qui participent au système européen d'échange de quotas d'émission. Pour le calcul de l'aide d'État maximale possible pour la compensation des prix de l'électricité dans les pays européens, l'analyse suppose que le produit spécifique a un repère de consommation d'électricité de 0,8 et que l'entreprise en question reçoit l'aide d'État maximale possible une fois que ce repère est incorporé dans le calcul de l'aide maximale. Les prix pour les États-Unis et la Chine sont indicatifs du prix moyen rapporté. Les industries individuelles, en fonction de leurs niveaux de consommation d'énergie et de leur localisation, peuvent être confrontées à des prix différents.

Source : analyse de l'AIE basée sur la base de données des prix des utilisations finales de l'AIE, données d'Eurostat (2023), Journal officiel de l'Union européenne (2021), Southeastern US Industrial Rate Survey, Brubaker & associates (2023), Shanghai Metal Market (2023), Intercontinental Exchange (2023), EUA Futures (consulté en novembre 2023).

Graphique n° 28 • Capacité actuelle pour répondre aux besoins futurs de *data centers*



Guide de lecture : l'estimation de RTE selon laquelle les data centers devraient consommer 10 TWh au début de la décennie 2020 est reprise comme hypothèse de départ. Selon RTE la consommation des data centers pourrait atteindre 15 à 20 TWh en 2030, puis 23 à 28 TWh en 2035. Ces deux hypothèses constituent respectivement le scénario conservateur et le scénario central. Une « hypothèse haute » incluant la montée en charge de l'IA générative, qui représenterait 25 %[125] des data centers d'ici à 2030, a également été ajoutée, ce qui représente une augmentation de 33 % par rapport à l'hypothèse haute.

[125] *CSIS, Gregory C. Allen et I. Goldston, 25 octobre 2024, « The Biden Administration's National Security Memorandum on AI Explained ».*

Un enjeu de planification du raccordement électrique

La très grande majorité des projets de *data center* de forte puissance ayant déjà contractualisé leur raccordement au réseau à très haute tension sont des *data centers* en colocation, plutôt destinés à des usages à latence faible. Ils sont concentrés dans deux grandes zones : l'Île-de-France et la région marseillaise.

Sur la base des dates contractualisées pour la mise en service des raccordements, la croissance devrait être la suivante :

- fin 2024, RTE aura raccordé 10 grands *data centers*, pour une puissance cumulée de 1 000 MW ;
- d'ici fin 2028, 27 nouveaux *data centers* devraient être raccordés au réseau à très haute tension, pour une puissance cumulée de 2 800 MW supplémentaires ;
- enfin, en raison des besoins de renforcement du réseau liés à cet afflux concentré de demandes, une dizaine de projets ne seront raccordés qu'entre 2029 et 2031, représentant environ 1 600 MW supplémentaires[126].

Le raccordement électrique des *data centers* de grande capacité est différent de celui des infrastructures électro-intensives classiques en raison des incertitudes qui entourent l'utilisation réelle de la puissance demandée. En effet, ce type de *data center* fonctionne souvent selon des modèles de colocation, où plusieurs clients partagent la même infrastructure, ce qui impose une montée en charge progressive, et a fortiori une plus grande flexibilité dans l'allocation de puissance. Le taux de remplissage des *data centers* en colocation augmente généralement de 10 à 15 % par an sur une période de 7 à 10 ans. En outre, la consommation énergétique des *data centers* varie fortement en fonction de la saisonnalité et des événements commerciaux, comme

[126] *Source : données internes, RTE.*

le Black Friday ou les fêtes de fin d'année. Ces variations impliquent des marges de puissance conséquentes, généralement comprises entre 25 % et 50 %, pour absorber les pics de consommation tout en évitant de sous-dimensionner le réseau.

L'enjeu de l'accueil des data centers réside donc dans la capacité du réseau électrique à fournir en temps et en heure la puissance demandée, sans mettre en danger l'alimentation des utilisateurs déjà raccordés. En effet, sur un réseau électrique, il ne peut pas y avoir « d'embouteillage » ou de ralentissement du débit : en cas de surcharge d'une ligne électrique, des dispositifs de sécurité coupent le courant et les conséquences peuvent affecter l'alimentation en électricité de territoires étendus, bien au-delà des seuls utilisateurs responsables de la surcharge.

Pour répondre à ces défis, RTE a déployé des systèmes de phasage qui, bien qu'encore contestés par certains acteurs, permettent une montée en puissance progressive des *data centers*, tout en préservant la stabilité et la résilience du réseau électrique.

Encadré n° 15 • Mode de fonctionnement de RTE pour raccorder les projets au réseau électrique à haute et très haute tension

RTE (Réseau de Transport d'Électricité) est responsable de l'exploitation et du développement du réseau électrique français. Encadré par des textes européens, RTE doit garantir un accès au réseau pour tous les utilisateurs, en respectant le principe de non-discrimination. **En application de ces principes, RTE ne priorise pas les projets en fonction de leur nature ou de leur importance, mais traite les demandes et leur alloue les capacités disponibles en fonction de leur ordre d'arrivée**.

Le mode de fonctionnement historique consiste à gérer chaque demande individuellement, et à déterminer les besoins d'extension et, le cas échéant, de renforcement du réseau existant. Des règles permettent ensuite de répartir le coût entre RTE et le demandeur. **Cette approche reste l'approche dominante pour des projets géographiquement isolés**.

Cependant, elle trouve ses limites dans des zones où se concentrent de nombreuses demandes, comme à Marseille ou en Île-de-France. Ainsi, en Ile-de-France, la somme des demandes des projets de *data centers* représente la puissance de 2 EPR, ou de 3 millions d'habitants supplémentaires. Entre Marseille et Aix-en-Provence, les besoins sont estimés entre 300 et 500 MW, soit l'équivalent de la consommation de 300 000 à 500 000 habitants. Au-delà d'un certain seuil critique de demandes, RTE doit renforcer le réseau pour répondre à ces besoins sans mettre en risque les utilisateurs déjà raccordés.

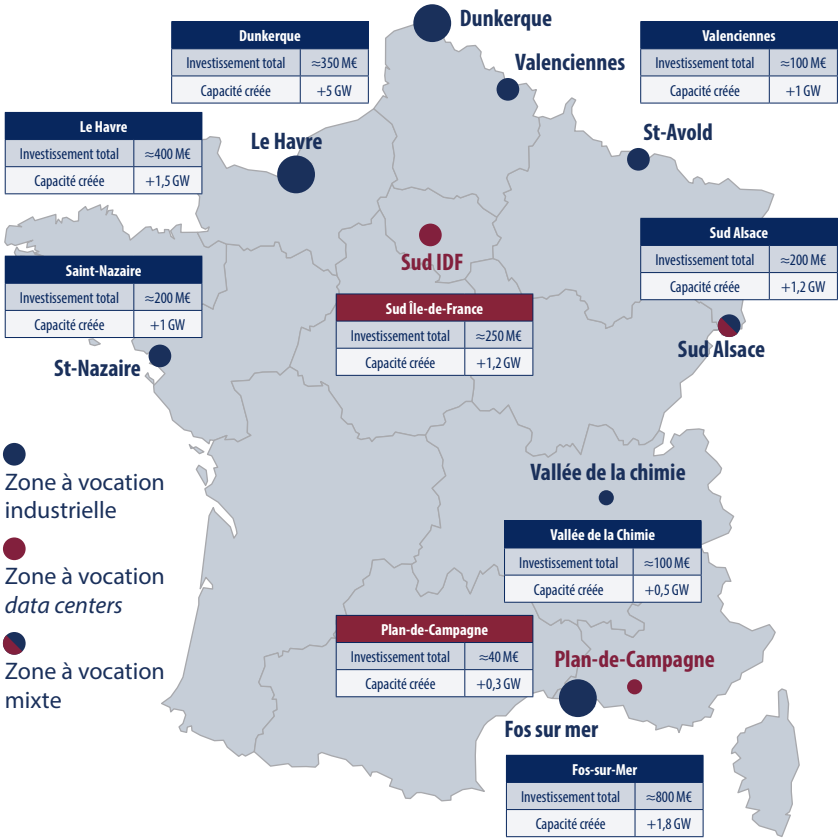
Ce processus peut rallonger les délais de raccordement, parfois jusqu'à 5 à 7 ans, en raison de la complexité des travaux nécessaires (par exemple, construction de nouvelles lignes aériennes ou renforcement des infrastructures existantes). Ces délais s'expliquent par la nécessité de convaincre les habitants et les collectivités locales du bien-fondé de ces aménagements, ainsi que par les contraintes techniques inhérentes à l'ajout de capacités.

Pour raccourcir les délais tout en assurant un développement rationnel du réseau, RTE a développé une pratique d'anticipation des investissements dans des « zones d'accueil mutualisées ». Ces solutions ont été en premier lieu déployées dans les grandes zones industrielles et industrialo-portuaires, comme Dunkerque, Le Havre

ou Fos, au service de secteurs stratégiques, tels que la décarbonation industrielle, l'hydrogène ou encore les carburants de synthèse. **Le raccordement mutualisé est mis en œuvre dès lors que RTE anticipe un besoin de capacité dans une zone donnée, quand bien même les demandes contractualisées n'ont pas encore atteint le seuil de saturation**. Concrètement, il s'agit de périmètres géographiques où les infrastructures de raccordement électrique sont conçues pour être partagées entre plusieurs utilisateurs, tels que des consommateurs industriels, des producteurs d'énergie renouvelable ou des gestionnaires de réseaux de distribution.

Cette démarche a été étendue aux besoins des infrastructures numériques et a donné naissance à des projets spécifiques pour les besoins en data centers, notamment dans le sud de l'Île-de-France, autour de Marseille (Plan de Campagne) et dans le sud de l'Alsace. Ces infrastructures sont déployées par étapes, avec un horizon temporel adapté à la montée en charge progressive des besoins. Ces sites sont situés à proximité de nœuds IXP *(Internet Exchange Points)*, qui sont des hubs où les réseaux d'opérateurs et de services numériques échangent directement des données *via* des accords de *peering*. Le *peering* permet à deux réseaux de transférer des données entre eux sans passer par un intermédiaire, réduisant les coûts de transit et la latence.

Graphique n° 29 • Répartition des zones d'accueil mutualisées sur le territoire



Source : données RTE.

Malgré la vocation affichée pour les différentes zones, RTE ne peut pas « réserver » telle ou telle zone à un usage particulier. La coexistence de projets critiques dans les zones d'accueil mutualisées

(*data centers*, *gigafactories*, décarbonation industrielle, etc.) peut donc créer des tensions sur l'accès à la capacité. Aujourd'hui, l'État intervient de manière limitée en jouant un rôle d'arbitre *via* les préfets de région, conformément à la loi n° 2023-175 du 10 mars 2023[127]. Ce dispositif transitoire, bien qu'utile pour prioriser les projets selon des critères d'intérêt général, reste temporaire et insuffisant. **Une planification nationale stratégique s'impose pour déterminer quels territoires sont les plus adaptés à ces infrastructures. Cette planification devrait prendre en compte des critères tels que la disponibilité des ressources (électricité, eau), l'accessibilité du foncier et le potentiel de développement territorial. L'État doit ainsi intervenir en bonne intelligence avec le secteur, en dépassant son rôle d'arbitre pour devenir plus prescriptif, afin que les bons projets soient raccordés au bon endroit**.

À l'international, plusieurs initiatives offrent des pistes intéressantes. Au Québec, les projets de plus de 5 MW nécessitent une accréditation ministérielle pour garantir leur alignement avec les objectifs stratégiques du développement économique. En Angleterre, un processus de rationalisation a permis de libérer des capacités inutilisées en éliminant des projets dormants, facilitant ainsi leur réaffectation vers des initiatives plus stratégiques. En France, la CRE a proposé en juillet dernier un mécanisme de régulation incitative basé sur un budget cible pour certains investissements, avec un système de bonus-malus en fonction des écarts par rapport aux prévisions budgétaires. **Dans cette logique, un dispositif de réservation assorti de clauses de révision périodiques pourrait être envisagé pour les *data centers* : les projets qui n'atteindraient pas leurs objectifs de consommation dans un délai donné pourraient voir leurs capacités réattribuées à d'autres acteurs**.

[127] *Plus précisément, c'est l'article 28 de cette loi qui confère aux préfets de région la responsabilité de fixer un ordre de classement des demandes de raccordement aux réseaux électriques pour certains projets, tels que les data centers, les initiatives de décarbonation industrielle ou les gigafactories de batteries.*

Pour atteindre cet objectif, une meilleure coordination entre l'État, RTE et les collectivités locales est également essentielle. **L'État doit dépasser son rôle actuel d'arbitre et intervenir de manière prescriptive pour aligner les priorités nationales et locales sur les usages de l'électricité**. Cela nécessite une cartographie fine des réserves foncières et des capacités électriques afin d'optimiser les implantations et d'éviter les surcoûts. Cette démarche doit répondre à des questions clés : où se situent les réserves foncières alignées avec les capacités de soutirage du réseau électrique ? Quels sites permettent un raccordement court et efficace ? Comment coordonner les stratégies des collectivités territoriales avec les impératifs nationaux pour prévenir les blocages institutionnels ? **Des « zones stratégiques » pourraient émerger, adaptées à l'accueil de *data centers* à faible latence et inscrites dans une dynamique territoriale cohérente, en s'appuyant sur les travaux déjà réalisés par RTE**. Pour cela, l'État gagnerait à partir de ses infrastructures existantes pour identifier les plus aptes à répondre aux besoins et usages émergents. Cela permettrait de décentraliser l'approche pour optimiser les ressources nécessaires pour l'ensemble des acteurs désireux de construire des infrastructures de traitement de données en France.

Une tendance notable est la demande de raccordements entre 30 et 60 MW, à la frontière de la limite RTE – à 40 MW – et des plus gros raccordements Enedis pour des solutions d'IA générative spécialisées dans des domaines spécifiques, comme la détection de fraudes dans le secteur bancaire, l'analyse prédictive en santé ou la personnalisation de l'expérience client dans le commerce de détail. Pour répondre à ces nouveaux besoins, des entreprises comme Eclairion développent des *data centers* dits « modulaires » composés de *data centers* en conteneurs, capables d'héberger des solutions à très haute densité, allant de 30 kW à 200 kW par baie contre 5 à 20 kW dans les *data centers* traditionnels. **Souvent appelés « *data centers* sans murs », ces machines permettent de mettre un supercalculateur dans un *data center* traditionnel ou modulaire**. Conçus sous forme de conteneurs

préfabriqués, ils regroupent plusieurs équipements – alimentation, refroidissement, sécurité, et serveurs – dans un format compact et déployable rapidement.

Les « *data centers* sans murs » sont en outre bien adaptés aux solutions à très haute densité. La densité correspond à la quantité d'énergie consommée par les serveurs dans une armoire informatique standard. **La densité est critique pour les calculs qui se font de manière concentrée dans des espaces réduits, notamment l'entraînement de modèles d'IA de fondation**. Les infrastructures modulaires sont optimisées pour ces charges de travail grâce à des systèmes avancés de refroidissement, tels que le refroidissement liquide qui dissipent efficacement la chaleur même dans des configurations très compactes. En maximisant la puissance disponible par mètre carré, ces *data centers* permettent d'économiser de l'espace et de réduire les coûts liés à la construction d'infrastructures étendues.

Ces nouvelles infrastructures incarnent aussi une convergence croissante entre le HPC et l'IA, qui permet de mutualiser les ressources, optimisant à la fois la vitesse et la polyvalence des infrastructures pour des applications variées allant de la recherche scientifique à l'analyse en temps réel. Par exemple, une usine optimisant ses processus grâce à l'IA peut installer un *data center* modulaire sur site pour garantir une faible latence et une indépendance vis-à-vis des réseaux distants. Dans le domaine bancaire ou médical, ces infrastructures permettent de traiter des données critiques tout en respectant des contraintes strictes de souveraineté et de confidentialité. Enfin, dans la recherche scientifique, elles offrent une capacité de calcul supplémentaire pour des projets collaboratifs ou des besoins temporaires.

Encadré n° 16 • Focus sur l'entreprise Sesterce, une illustration de plusieurs dynamiques clés liées à la convergence IA et HPC

- L'entraînement des modèles avancés d'IA, au premier rang desquels les LLM – SLM – exige des infrastructures de calcul à très haute densité énergétique et de performance, caractéristiques des environnements HPC. Sesterce cible des racks de 150 kW, adaptés aux GPU nécessaires pour l'IA.

- Ces capacités dépassent largement les standards traditionnels des *data centers* classiques, illustrant comment l'IA pousse les exigences du HPC encore plus loin.

- En proposant un modèle de colocation, Sesterce permet à des entreprises tierces de louer des espaces et d'installer leurs propres équipements, ce qui réduit les barrières à l'entrée pour des acteurs spécialisés en IA sans qu'ils aient à construire leurs propres *data centers*. Ce modèle limite l'investissement initial pour Sesterce, qui ne porte pas le coût total des équipements et peut mutualiser les infrastructures critiques (énergie, refroidissement).

- Bien que la colocation offre une flexibilité, Sesterce anticipe des besoins spécifiques pour lesquels il pourrait être nécessaire de posséder et d'exploiter ses propres capacités, notamment pour des projets complexes ou des clients ayant des exigences particulières.

Une logique étatique de planification territoriale des usages de l'électricité doit tenir compte de ces nouvelles infrastructures de traitement des données. **En intégrant ces solutions dans les zones d'accueil mutualisées déjà identifiées par RTE, la France pourrait favoriser l'émergence de « zones stratégiques » offrant des capacités et des latences mieux adaptées aux différents besoins**. Cette approche permettrait d'optimiser l'implantation des projets tout en répondant aux exigences croissantes des industries, des collectivités et des acteurs de l'IA. En associant leur flexibilité à une planification étatique proactive, ces infrastructures modulaires ont le potentiel de devenir un atout majeur pour relever les défis technologiques et énergétiques de demain.

Enfin, une évolution du cadre réglementaire applicable aux infrastructures numériques de traitement doit également prendre en compte les contraintes financières. Si RTE devait anticiper les besoins de raccordement de futurs projets, les coûts pourraient devenir significatifs. **Avec 90 % de ses ressources financières provenant du tarif d'utilisation des réseaux publics d'électricité[128] (TURPE), RTE ne pourrait pas assumer ces investissements sans recapitalisation ou recours à des financements privés**. Cela met en lumière l'urgence de définir un cadre adapté pour coordonner les infrastructures électriques et les besoins croissants liés à l'implantation de *data centers* dans un maillage territorial optimisé.

L'impératif de valorisation des installations nucléaires françaises

Sous réserve de disponibilité du foncier, certains usages peuvent être raccordés à des zones de production nucléaires existantes. Dans ce contexte, l'énergie nucléaire est une option de plus en plus prometteuse pour les acteurs du *cloud*, car elle permet d'entraîner des

[128] *EDF, février 2024, Qu'est-ce que le TURPE (Tarif d'Utilisation des Réseaux Publics d'Électricité).*

modèles de fondation avec une énergie décarbonée et à bas coût. Lors du Sommet de l'IA à Paris en février 2025, EDF a annoncé la pré-identification de quatre sites industriels sur son foncier, offrant une puissance totale disponible de 2 GW, spécifiquement destinés à accueillir des *data centers*. Cela représente environ 10 % de la consommation électrique totale des centres de données en Europe en 2022. Ces sites, situés en Rhône-Alpes, Grand Est et Île-de-France, sont déjà raccordés au réseau électrique, réduisant ainsi les délais de mise en service.

Toutefois, le raccordement direct de *data centers* aux centrales nucléaires existantes soulève des défis spécifiques liés aux particularités du parc nucléaire français, qui diffère sensiblement du modèle américain. Aux États-Unis, les réseaux électriques sont organisés à l'échelle des États et restent faiblement interconnectés, ce qui pousse les *hyperscalers* à se brancher directement aux centrales pour sécuriser leur approvisionnement énergétique. En revanche, en France, les centrales nucléaires sont intégrées à un réseau national unifié, conçu pour maximiser la flexibilité et la résilience de l'approvisionnement. Contrairement aux infrastructures américaines, les architectures de postes des centrales françaises ne sont pas adaptées à un raccordement direct des *data centers*. Lorsqu'une tranche est arrêtée pour maintenance, l'accès au réseau peut être limité, ce qui complexifie l'exploitation continue des infrastructures numériques. Si un tel raccordement n'est pas une impossibilité technique, il nécessiterait une refonte de l'organisation actuelle du réseau et impliquerait d'importantes analyses d'impact en matière de sûreté nucléaire et de gestion des flux énergétiques. **L'enjeu clé réside dans la capacité à transférer efficacement la puissance électrique entre différentes zones, un domaine où la France bénéficie d'un avantage compétitif grâce à son réseau haute tension maillé et à l'expertise de RTE dans l'optimisation des flux énergétiques à grande échelle**.

Les États-Unis adoptent une stratégie offensive en intégrant l'énergie nucléaire au cœur de leurs infrastructures numériques, en s'appuyant à la fois sur des réacteurs nucléaires existants et sur le développement des petits réacteurs modulaires (SMR). **L'objectif des *hyperscalers* est double : sécuriser un approvisionnement énergétique stable et décarboné tout en s'affranchissant du réseau électrique traditionnel. Pour cela, ils explorent deux approches complémentaires : l'exploitation de capacités existantes dans des centrales nucléaires classiques et l'investissement dans les SMR pour une production plus flexible et décentralisée**. Microsoft illustre bien cette double logique. D'une part, l'entreprise a signé un contrat d'achat d'électricité (PPA) avec la centrale de Three Mile Island pour exploiter des capacités disponibles et garantir l'alimentation en énergie nucléaire de ses *data centers*. D'autre part, elle a noué un partenariat avec TerraPower[129] pour développer des réacteurs de quatrième génération, misant sur une indépendance énergétique totale à terme. Amazon suit une approche similaire en s'associant à plusieurs acteurs du nucléaire pour connecter ses infrastructures directement aux centrales, contrairement à OVH qui, bien que proche des sites de production, reste connecté au réseau *via* un poste source.

Parallèlement, les *hyperscalers* investissent massivement dans les SMR, perçus comme une solution énergétique standardisable et adaptable aux infrastructures numériques. Google a franchi un cap en signant le premier PPA basé sur des technologies SMR avec Kairos Power[130] avec un déploiement attendu dès 2027-2028. Oracle, Alphabet et OpenAI s'engagent également dans cette voie, séduits par la flexibilité et la capacité de ces réacteurs à alimenter directement leurs centres de traitement de données.

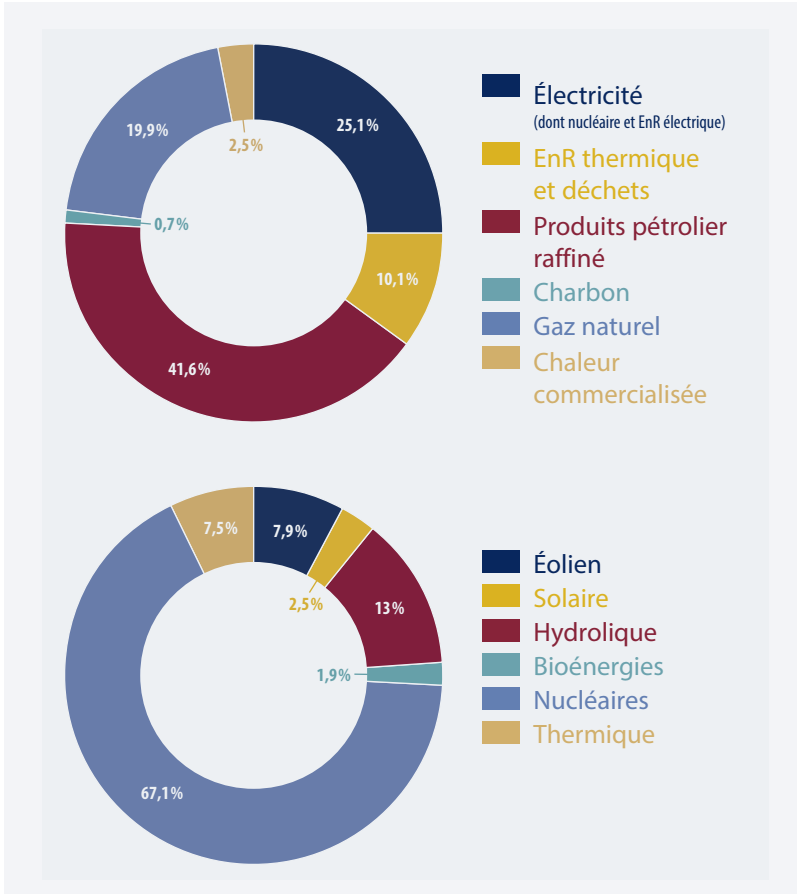
[129] *TerraPower est une entreprise américaine fondée en 2006 par Bill Gates, spécialisée dans le développement de réacteurs nucléaires avancés visant à fournir une énergie propre et durable.*

[130] *RGN, 15 octobre 2024, « Google et Kairos Power s'engagent pour déployer une flotte de SMR ».*

En France, la situation est différente car avec un surplus énergétique encore notable, l'enjeu ne porte pas sur l'indépendance énergétique des *data centers*, mais sur l'optimisation de l'électrification des usages. La priorité est de mieux allouer cette ressource aux opérateurs industriels existants, en tenant compte des enjeux de compétitivité et des différences structurelles avec nos voisins européens. **Si l'intégration des *data centers* aux infrastructures nucléaires est envisageable, elle soulève des défis techniques et réglementaires, notamment en raison de l'architecture du réseau français, qui n'a pas été conçu pour un tel modèle**. Contrairement aux États-Unis, où la transition est déjà engagée, l'Europe reste en retrait : alors que les premiers SMR américains pourraient être opérationnels d'ici 2027-2028, leur déploiement en Europe est attendu plutôt entre 2030 et 2035, creusant ainsi l'écart stratégique. **La France dispose pourtant de conditions particulièrement avantageuses pour implanter des *data centers* avec un mix énergétique bas carbone et peu coûteux , composé à 70 % d'énergie nucléaire, d'autant plus que cela permettrait d'augmenter la demande en électricité**.

Encadré n° 17 • Répartition des dépenses énergétiques dans les *data centers* français

L'énergie dans les *data centers* représente 54 % des dépenses totales, avec une consommation moyenne de 5,15 MWh/m^2/an (soit l'équivalent d'une ville de 50 000 habitants pour un datacenter de 10 000 m^2).



Dans ce contexte, tout projet de raccordement direct de *data centers* à des centrales nucléaires françaises devra prendre en compte les éléments suivants :

1. **L'impact sur les tarifs de l'électricité :** Le raccordement direct de *data centers* de grande capacité aux centrales nucléaires pourrait entraîner des distorsions tarifaires, notamment si ces centres

bénéficient de tarifs préférentiels au détriment d'autres utilisateurs. C'est pour cette raison qu'en novembre 2024[131], la Commission fédérale américaine de régulation de l'énergie a rejeté un accord modifié pour un *data center* d'Amazon connecté directement à une centrale nucléaire, estimant que cela pourrait augmenter les factures des consommateurs finaux et affecter la fiabilité du réseau électrique. En France, la réglementation reste très protectrice du principe de péréquation géographique des tarifs et ne prévoit pas de tarification spécifique pour des consommateurs situés à proximité des grandes centrales de production. Un raccordement *via* une « ligne directe » est juridiquement possible, mais il suppose une autorisation administrative spécifique.

2. **L'acceptabilité sociale :** l'implantation de *data centers* à proximité de centrales nucléaires suscite des préoccupations parmi les riverains, notamment concernant les nuisances potentielles (bruit, trafic, chaleur) et l'utilisation locale des ressources naturelles, telles que l'eau pour le refroidissement. Chaque projet est soumis à une évaluation rigoureuse, incluant des consultations publiques orchestrées par la Commission nationale du débat public (CNDP), tant pour la construction d'un *data center* que d'une centrale nucléaire. L'acceptabilité sociale est donc cruciale pour la réussite de ces projets.

3. **Les capacités des centrales nucléaires :** les centrales nucléaires à quatre tranches (= nombre de réacteurs) sont capables de fournir une alimentation stable et de gérer des charges importantes, ce qui les rend aptes à assurer un raccordement fiable pour des infrastructures critiques comme les *data centers*. En revanche, les centrales à deux tranches peuvent ne pas disposer des capacités nécessaires pour assurer une telle stabilité, limitant leur aptitude à soutenir seules ces projets stratégiques : à certaines périodes, un

[131] *The Hindu, 2 novembre 2024, "U.S. regulators reject amended interconnect agreement for Amazon data center".*

complément par le réseau pourra être nécessaire. Leur potentiel peut toutefois faire l'objet d'une étude dans les situations où une tranche serait à l'arrêt.

4. **L'impact sur les grands flux nationaux d'électricité :** il est faux de croire que le raccordement d'un *data center* en sortie directe d'une centrale nucléaire, elle-même nécessairement raccordée au réseau public de transport, n'a pas d'impact sur les flux circulant sur ce réseau, même à des distances éloignées de la centrale. Les modèles de RTE montrent un déséquilibre croissant entre l'Ouest de la France, plutôt excédentaire en production d'électricité, et l'Est de la France, plutôt déficitaire. L'installation d'un *data center* de très forte puissance à proximité d'une centrale nucléaire de la Vallée du Rhône, par exemple, conduirait à consommer localement une partie de l'électricité produite par cette centrale. Ce faisant, cette électricité ne sera plus disponible pour les besoins industriels et urbains de cette même Vallée du Rhône au sens large, de Fos-sur-Mer jusqu'à Genève. RTE devra l'acheminer depuis des zones excédentaires, notamment dans l'Ouest de la France. Il est donc préférable de privilégier une implantation des *data centers* dans la moitié Ouest de la France, afin de limiter les externalités négatives sur les grands flux nationaux.

Recommandation 3

Construire une réelle planification étatique en matière d'approvisionnement électrique pour mailler le territoire français en *data centers* de grande capacité en anticipant les usages futurs.

Aujourd'hui, les projets de construction de *data centers* relèvent d'une dynamique propre aux acteurs privés et de raccordements effectués dans l'ordre des demandes, sans hiérarchisation (premier arrivé, premier servi). **Une planification souple est nécessaire qui s'intéresse à la fois à la quantité de *data centers* nécessaires sur la base des besoins futurs et à l'ordonnancement des projets selon leur criticité et leur localisation**. Un des problèmes stratégiques majeurs est la mise en concurrence indistincte des projets hors de toute analyse des besoins territoriaux et des avantages concurrentiels. L'enjeu des raccordements au réseau électrique en est une manifestation criante. Il se double d'une crainte liée à la rareté de l'électricité disponible, aux conflits d'accès et à la concurrence indue entre réindustrialisation, décarbonation, aménagement du territoire et transition numérique.

Recommandation 3.1 : communiquer au plus haut niveau de l'État pour démontrer la compatibilité entre disponibilité énergétique et infrastructures numériques. L'État doit porter un discours clair et stratégique sur l'adéquation entre les besoins numériques, industriels et de décarbonation, faisant valoir l'excédent de production électrique français. **Une initiative conjointe entre acteurs publics (RTE, CNDP) et privés (opérateurs de *data centers*, industriels) pourrait inclure une campagne de communication mettant en avant les preuves terrain (exemples de sites opérationnels redynamisant l'économie dans le respect des engagements de trajectoire climat)**.

Recommandation 3.2 : planifier et piloter le raccordement électrique des *data centers* – ceux de grande capacité comme ceux de proximité – à l'échelle nationale. L'État doit engager une planification systémique, intégrant toutes les parties prenantes (RTE, collectivités, acteurs économiques), pour identifier les besoins réels de la population. Cela passe par une **cartographie nationale co-construite avec RTE, l'État et les collectivités, localisant des « zones stratégiques » pour des infrastructures adaptées à des usages à latence faible ou forte**. Cette cartographie devra inclure les infrastructures modulaires (« sans murs ») pour les usages IA-HPC et se fonder sur quatre critères prioritaires : le différentiel de coût avec de l'infrastructure fixe (fonction du niveau de flexibilité attendu), la pertinence énergétique (proximité des sources), le potentiel économique (emplois locaux), et l'impact social (développement des territoires).

À plus long terme, une fois l'infrastructure opérante, une réflexion devra être initiée :

- **Sur la nature des offres commerciales et la priorisation des projets et des demandes de raccordement afin d'éviter l'engorgement de projets « zombis » ou d'usages à très faible valeur ajoutée**. Cette réflexion doit aussi permettre de réduire les délais de fabrication des transformateurs de puissance qui s'étendent constamment dans le secteur des *utilities*.

- **Sur la manière de valoriser la capacité du parc nucléaire français à transférer efficacement la puissance électrique entre différentes zones**, un domaine où la France bénéficie d'un avantage compétitif grâce à son réseau haute tension maillé et à l'expertise de RTE dans l'optimisation des flux énergétiques à grande échelle.

Recommandation 3.3 : capitaliser sur l'avantage géographique de la France dans les négociations internationales et les partenariats stratégiques. La position géographique de la France, au carrefour des axes nord-sud en Europe et bénéficiant d'une connectivité privilégiée avec les États-Unis, constitue un atout majeur à valoriser. De plus, le territoire français offre des caractéristiques naturelles uniques pour le développement d'infrastructures numériques durables. Par exemple, les régions montagneuses, avec leur potentiel hydroélectrique, permettent d'alimenter des *data centers* en énergie renouvelable. Implantés dans des vallées où les températures plus basses contribuent au refroidissement, ces centres pourraient réduire significativement leur consommation énergétique liée au refroidissement, qui représente une part importante de leur empreinte carbone.

b. La possibilité de cartographier plus finement les usages sous-tendus par les *data centers* qui seront amenés à être construits sur le territoire français

Selon la latence requise, différents usages dans différents types de data centers

Selon le type d'usage et la capacité de traitement requise, un *data center* n'a pas les mêmes exigences en matière d'énergie, de connectivité à un point d'échange internet (IXP[132]) ou de foncier. **L'un des critères essentiels est la latence, c'est-à-dire le délai entre l'envoi d'une requête**

[132] *Infrastructure physique qui permet à plusieurs réseaux (comme ceux des fournisseurs d'accès à Internet, des entreprises, des fournisseurs de contenu, ou des réseaux universitaires) d'interconnecter directement leurs trafics. Cette interconnexion, appelée peering, permet d'échanger des données de manière efficace, rapide, et à moindre coût, sans passer par des intermédiaires comme les opérateurs de transit.*

et la réception d'une réponse. Certains usages, comme l'inférence des modèles d'IA dans des applications interactives, nécessitent une latence très faible, de l'ordre de quelques millisecondes, pour garantir une réactivité immédiate. C'est le cas des assistants vocaux, du trading algorithmique ou encore de la télémédecine, où le moindre décalage peut dégrader l'expérience utilisateur ou compromettre la performance du service. À l'inverse, d'autres traitements, comme l'entraînement des modèles d'IA, tolèrent une latence plus élevée. Ce type de calcul repose sur le *batch processing*, un mode de traitement en lots qui peut s'étaler sur plusieurs heures, voire jours, sans impact immédiat sur l'usage final.

L'entraînement des modèles illustre bien cette distinction. Contrairement à l'inférence, qui exige une interaction rapide avec l'utilisateur, il s'effectue principalement en interne, au sein du *data center*, sans nécessité de connexions à faible latence avec l'extérieur. La majorité des échanges de données se fait entre GPU et TPU *via* des interconnexions ultra-rapides conçues pour optimiser la communication locale, réduisant ainsi la dépendance aux infrastructures réseau externes. En centralisant ces opérations dans des sites dédiés, les opérateurs optimisent les coûts énergétiques et la gestion des ressources, en concentrant la puissance de calcul dans des environnements spécifiquement conçus pour les charges de travail intensives.

Graphique n° 31 • Exemples d'usages de *data centers* selon les besoins en latence

Une classification plus précise des usages pourrait se baser sur le niveau de symétrie des transferts de données impliqués. Par exemple, dans des cas où une grande quantité de données est reçue (analyse de capteurs, flux vidéo, etc.) mais où peu de données sont renvoyées en retour (commandes ou instructions simples), le trafic est qualifié d'asymétrique. Comprendre cette asymétrie nécessite toutefois une analyse approfondie des usages industriels sous-jacents, afin de déterminer les exigences spécifiques en matière de latence, de volume de données et de puissance de calcul.

Pour les data centers de grande capacité, poursuivre le mouvement engagé de simplification des procédures administratives

Les *data centers* de grande capacité, soit des projets de 1 GWh, sont indispensables pour soutenir l'adoption croissante de l'IA générative par les entreprises. **Cette technologie repose sur des modèles nécessitant une puissance de calcul massive et des infrastructures capables de gérer des volumes colossaux de données en temps réel**. À l'international, les États-Unis ont accéléré la construction de tels centres avec le *National Security Memorandum on AI*, adopté en 2024, et des projets comme Stargate, un *data center* géant avec 100 Md $ déjà engagés et un objectif de 500 Mds $ dans les quatre prochaines années, lancé par OpenAI en partenariat avec Microsoft et Vodafone. D'après le baromètre EY / France Data Center 2024, la demande énergétique liée à l'IA pourrait croître de 13 à 14 % par an, atteignant 1 GW d'ici 2033. Pourtant, seulement 22 % des entreprises dépassent le stade de la preuve de concept, et seuls 4 % de ces projets apportent une réelle valeur ajoutée, selon une étude récente du BCG[133]. Ce décalage s'explique par l'insuffisance d'architectures adaptées dans les outils de production existants, telles que les systèmes multi-agents, et par le déficit de puissance de calcul.

[133] *Étude BCG, octobre 2024, "Where's the Value in AI?".*

Ainsi, pour que les entreprises puissent pleinement tirer parti de l'IA générative, la France doit investir dans des *data centers* de grande capacité capables de répondre aux exigences croissantes en termes de connectivité, de rapidité d'exécution et d'efficacité énergétique.

Le Sommet de l'Action de l'IA a été l'occasion d'annoncer la mise à disposition de 35 sites « clés en main » pour accueillir des data centers de grande capacité en France. Ces terrains, dont les superficies varient de 18 à plus de 150 hectares, sont prévus pour des projets pouvant atteindre jusqu'à 1 GW. Le Gouvernement s'est engagé à assurer un raccordement au réseau électrique pour une capacité élevée à partir de 2027.

Cette initiative vise à renforcer l'attractivité du territoire français pour l'implantation de *data centers*, en réponse aux investissements des *hyperscalers* qui se sont jusqu'à présent davantage orientés vers d'autres pays européens, principalement en raison de la lenteur des procédures administratives françaises pour la réalisation de tels projets.

Pendant la phase préparatoire d'un projet de *data center*, le principal problème réside dans la lisibilité des procédures pour les porteurs de projet. Cette phase, qui comprend les validations administratives et les travaux de raccordement, dure en moyenne entre 3 et 6 ans en France, en raison de deux obstacles majeurs :

- **D'une part, la multiplicité des entités impliquées dans le processus décisionnel constitue un frein important**. En France, diverses administrations nationales et locales interviennent : les préfectures régionales et départementales, les Directions Départementales des Territoires (DDT), les Directions Régionales de l'Environnement, de l'Aménagement et du Logement (DREAL), et les missions régionales d'Autorité environnementale (MRAe). Chaque entité traite des

aspects spécifiques, tels que l'urbanisme, les permis de construire ou les évaluations environnementales, souvent sans coordination. Par exemple, la mission régionale d'autorité environnementale émet des avis indépendants dans une procédure distincte, ce qui rallonge les délais. De plus, bien que le maire ait le pouvoir final d'octroi du permis de construire, il peut être contraint par la communauté d'agglomération, notamment en cas de contraintes liées au raccordement électrique ou à la fibre. **Ce cadre institutionnel fragmenté peut ajouter plusieurs années à la phase préparatoire, portant les retards jusqu'à 5 ans**.

- **D'autre part, les *data centers* sont classés comme des entrepôts en droit de l'urbanisme, des industries selon le code des impôts, et des bâtiments tertiaires au regard du décret tertiaire**. Cette incohérence réglementaire crée un manque de lisibilité pour les investisseurs et interdit leur implantation dans de nombreuses zones où ils généreraient pourtant moins de nuisances qu'un entrepôt logistique, réduisant ainsi les terrains disponibles pour leur construction. La qualification industrielle repose en grande partie sur l'utilisation de groupes électrogènes[134], indispensables pour garantir la continuité de service en cas de coupure électrique. Ces groupes, alimentés par des énergies fossiles, interviennent dès la coupure d'énergie ce qui explique qu'ils soient classés comme des équipements à risque industriel en raison des cuves de carburant qu'ils utilisent. Même s'ils ne sont activés qu'en situation d'urgence, ils nécessitent des tests réguliers pour s'assurer de leur bon fonctionnement, ce qui explique leur classification industrielle. **Ces contraintes allongent la procédure d'un an supplémentaire en moyenne, rendant le processus encore plus complexe et coûteux pour les porteurs de projet**.

[134] *Les groupes électrogènes sont des dispositifs conçus pour fournir de l'électricité en cas de défaillance du réseau principal.*

Pendant la phase de construction, le principal problème réside dans la stabilité juridique pour les porteurs de projets, qui rencontrent trois principaux obstacles, prolongeant souvent les délais de 4 à 7 ans[135], voire jusqu'à 15 ans dans les cas les plus complexes.

- **D'abord, la construction est freinée par des exigences administratives lourdes, notamment les agréments de la mission régionale de l'autorité environnementale**. Ces agréments, requis pour les sites abritant plus de 22 espèces protégées, nécessitent des dossiers complexes (plus de 300 pages), ajoutant en moyenne 2,5 ans aux délais. En Île-de-France, une contrainte supplémentaire est l'obligation de récupérer la chaleur fatale pour des usages tels que le chauffage urbain. Bien que cette exigence ait des avantages écologiques, elle est souvent impraticable sans infrastructures adaptées à proximité. À l'international, des critères comme le *Power Usage Effectiveness* (PUE) et le *Water Usage Effectiveness* (WUE) avec des techniques innovantes de refroidissement, telles que le refroidissement à eau en circuit fermé[136] ou le *river cooling*[137], offrent des solutions plus flexibles et durables. **Le *Climate Neutral Data Center Pact*, qui réunit industriels et décideurs publics sur le sujet, préconise la mise en place d'un ratio de type PUE/WUE**. Ces approches, déjà en place dans les trois centres nationaux du GENCI, à Marseille ou en Allemagne, démontrent qu'il est possible d'intégrer des *data centers* dans des écosystèmes locaux tout en minimisant leur impact environnemental. **Surtout, privilégier le PUE comme critère de conception influence directement le développement des applications dans le data center. En optimisant le**

[135] *Business Immo, mai 2021, Implanter un data center en France : regards croisés en droit immobilier et de l'environnement.*

[136] *Ce type de système peut réduire l'impact environnemental en diminuant de 30 % la chaleur générée par rapport aux systèmes de ventilation traditionnels.*

[137] *Exploitation de méthodes adiabatiques pour retarder l'utilisation de la climatisation quand les conditions ambiantes le permettent. Les approches adiabatiques désignent des méthodes qui font évoluer un système lentement et sans échange de chaleur ou d'énergie avec son environnement, de manière à rester proche de son état initial ou de son état d'équilibre tout au long du processus.*

ratio puissance utilisée pour le fonctionnement des serveurs / énergie totale consommée (kWh), un PUE performant incite à concevoir des applications frugales en énergie. Cela signifie, par exemple, réduire les ressources nécessaires pour exécuter une tâche ou limiter les besoins en refroidissement. Ces applications, conçues pour être légères et efficaces, permettent de maximiser la densité énergétique des installations tout en réduisant leur empreinte carbone. **Ainsi, un faible PUE peut devenir un levier de compétitivité, en attirant des entreprises soucieuses de durabilité et en répondant aux exigences croissantes des régulateurs et des clients pour des infrastructures écologiquement responsables**.

- Ensuite, même après l'obtention des autorisations, les projets restent exposés à des recours pouvant s'étendre sur plusieurs années. **En l'absence de mécanismes de filtrage, les opposants peuvent soumettre des requêtes à trois niveaux de juridiction (tribunaux administratifs, cours administratives d'appel, Conseil d'État), retardant les projets de sept ans en moyenne**. Cette insécurité juridique décourage les investisseurs et alourdit considérablement les coûts. Pour l'éolien offshore[138], depuis la simplification des procédures d'autorisation de 2020, le Conseil d'État traite en premier et dernier ressort les litiges, accélérant ainsi les procédures et réduisant le délai de recours potentiel. **Une adaptation pour les *data centers* limiterait les recours abusifs et améliorerait la visibilité pour les investisseurs**.

- **Enfin, les *data centers* souffrent d'une mauvaise image auprès des populations locales, souvent perçus comme ayant une faible valeur ajoutée**. Pourtant, ils génèrent en moyenne 2,5 fois plus d'emplois directs[139] qu'un entrepôt logistique de taille équivalente et ont contribué à hauteur de 5 Mds € de valeur ajoutée

[138] *Rapport du Conseil Général de l'Environnement et du Développement Durable, « La simplification des procédures d'autorisation applicables aux éoliennes en mer », juin 2021.*

[139] *France datacenter, 31 janvier 2022, Guide pratique à destination des élus franciliens.*

directe et indirecte[140] en 2023. **Ce potentiel économique reste mal compris et manque d'incarnation locale. Des projets intégrant un « récit territorial », comme le parc technologique de Ledger à Vierzon ou la réutilisation de sites industriels désaffectés comme le fait Eclairion, pourraient renforcer leur acceptabilité**. Toutefois, le mécanisme de zéro artificialisation nette (ZAN), introduit par la loi Climat et Résilience de 2021 complique leur implantation, en attribuant des quotas d'artificialisation désavantageux aux communes peu urbanisées[141]. Ce déséquilibre freine le développement de ces territoires, accentuant les écarts avec les collectivités déjà urbanisées. Enfin, l'acceptabilité de la construction de nouveaux *data centers* à proximité de centrales nucléaires pourrait également être renforcée si le financement de ces infrastructures s'appuie sur de nouveaux consommateurs pour éviter d'augmenter la TURPE.

Dans ce contexte, le projet de loi de Simplification de la vie économique[142], déposé au Sénat le 24 avril 2024, et dont l'adoption rapide a été réitérée dans la déclaration de politique générale du Premier Ministre du 14 janvier 2025 vise à réduire les délais administratifs en simplifiant les démarches et en rationalisant les normes pour les entreprises. Parmi ses 26 mesures, l'article 15 propose de classer les *data centers* comme projets d'intérêt national majeur, conformément à la loi sur l'industrie verte du 23 octobre 2023. Ce statut permettrait aux préfets de modifier les plans locaux d'urbanisme pour accélérer les démarches. **Cependant, le texte reste flou sur les problématiques de lisibilité et de stabilité juridique, laissant incertain son impact réel sur la réduction des délais**. Par ailleurs, l'article 18 introduit des

[140] *Étude d'impact et baromètre France Datacenter/EY – 2024.*

[141] *Ce mécanisme repose sur l'attribution de quotas d'artificialisation aux collectivités locales, calculés en fonction de l'artificialisation passée. Un quota d'artificialisation correspond à la limite fixée pour la création ou la transformation de surfaces naturelles ou agricoles en surfaces artificialisées (bâtiments, routes, infrastructures, etc.) dans un territoire donné.*

[142] *Texte n° 550 (2023-2024) de M. Bruno Le Maire, ministre de l'économie, des finances et de la souveraineté industrielle et numérique, déposé au Sénat le 24 avril 2024.*

modifications aux avances sur mesures compensatoires *via* un système de « crédit revolving[143] ». La nouvelle version impose des conditions de remboursement plus restrictives, réduisant l'attractivité de ces avances pour les entreprises. Cette rigidité pourrait freiner les investissements dans des projets pourtant essentiels à l'attractivité économique du territoire, limitant l'effet escompté de ce projet de loi.

Ces mesures ont été rendues urgentes par le contexte de concurrence mondiale et européenne exacerbée. **Le Royaume-Uni a déjà annoncé un plan doté de 50 propositions qui vise à faire du pays un « *leader* en IA » par le biais de la création de « zones de croissance de l'IA » où des permis de construire seront délivrés plus rapidement pour construire des *data centers* dédiés à l'IA**. Trois entreprises technologiques – Vantage Data Centres, Kyndryl et Nscale – se sont déjà engagées à investir 14 Mds £, soit 17 Mds €, au Royaume-Uni, pour développer notamment des centres de stockage des données, a indiqué le Gouvernement. Leurs projets devraient, selon lui, permettre de créer près de 13 000 emplois. La première « *zone de croissance de l'IA* » sera implantée à Culham, près de l'université d'Oxford (sud-est de l'Angleterre).

Recommandation 4

Capitaliser sur le lancement des 35 sites clés en main pour raccourcir les délais de construction de *data centers* dont l'intérêt économique et social est démontré en simplifiant les procédures administratives

[143] *Mécanismes de préfinancement permettant aux entreprises d'engager les travaux de compensation environnementale avant l'obtention des autorisations définitives.*

Les délais administratifs liés aux permis de construire et à la sortie de terre des *data centers* de grande capacité sont la principale raison de notre manque d'attractivité et nous privent de près de 100 Mds € d'investissements effectués par défaut dans des pays limitrophes, alors que notre électricité décarbonée devrait être un atout recherché.

Recommandation 4.1 : le projet de loi Simplification de la Vie Économique et les dispositions de type *fast track* dans la loi Industrie Verte en cours d'examen doivent doter d'un statut de projet d'intérêt national majeur les *data centers* de grande capacité et de proximité. Pour assurer l'efficacité de ce statut, certaines dispositions administratives doivent être intégrées au projet de loi qui veilleront à accélérer les procédures sans sacrifier la rigueur de sélection des projets les plus qualitatifs.

- **Regrouper les services environnement et urbanisme** au sein des DREAL pour accélérer le traitement des demandes administratives.

- L'un des critères retenus pour l'obtention du permis de construire d'un *data center* est aujourd'hui celui de la chaleur fatale dissipée dans l'environnement sans réutilisation. **Ce critère n'est pas toujours pertinent, bien qu'il soit important dans certains cas, et doit être remplacé par le ratio PUE (Power Usage Effectiveness) / WUE (Water Usage Effectiveness) afin de prendre en compte l'efficacité des techniques de refroidissement comme critère d'impact environnemental**.

- **Confier au Conseil d'État le traitement en premier et dernier ressort des recours contre l'implantation de *data centers* dans un territoire après l'octroi des permis, sur le modèle de l'éolien**

offshore[144]. En effet, les procédures de recours auprès des différents tribunaux peuvent prolonger de 7 ans les délais de construction des *data centers* une fois le permis de construire obtenu, ce qui représente une incertitude économique et industrielle insupportable et dissuade les investisseurs.

Recommandation 4.2 : renforcer l'acceptabilité sociale des projets de *data centers* situés dans des agglomérations de taille moyenne grâce à un « récit territorial » puissant. Pour cela, inciter fiscalement par le biais de subvention à la construction de « centres d'excellence technologiques » à proximité des nouveaux *data centers*. Pour les *data centers* assortis à ce type de centres, expérimenter une dérogation ciblée au principe de zéro artificialisation nette (ZAN) dans les communes peu urbanisées qui souhaiteraient accueillir ce type de projet pour favoriser leur développement économique.

2.3. TOUT AUSSI CRUCIAL, RAPPROCHER LES MÉTIERS DU RÉSEAU DES MÉTIERS DU TRAITEMENT DE DONNÉES POUR NE RIEN PERDRE DE L'EXCELLENCE DES INGÉNIEURS FRANÇAIS EN LA MATIÈRE

Le système éducatif français doit concevoir les infrastructures numériques comme les nouvelles voies de la prospérité et de la compétitivité du pays. Cela suppose d'agir à la fois sur les métiers en forte tension et d'adapter les formations, des plus élitistes aux plus opérationnelles, pour répondre aux besoins des entreprises, et de le faire au niveau européen.

[144] *Rapport du Conseil Général de l'Environnement et du Développement Durable, « La simplification des procédures d'autorisation applicables aux éoliennes en mer », juin 2021.*

En effet, la maîtrise des talents est une dimension cruciale. Il ne suffit pas de fixer des objectifs ambitieux ; il faut également convaincre les entreprises – utilisateurs finaux de mettre en place une véritable chaîne de responsabilités qui garantit de bout en bout (de l'implantation à l'usage) l'efficacité des infrastructures numériques utilisées. Cela passe nécessairement par la mobilisation des **DSI d'entreprises et de la communauté scientifique, ayant la charge d'expliquer la dimension véritablement stratégique des infrastructures (jusqu'à présent perçues comme une occupation ingrate et d'arrière-plan) et les approches désormais « packagées » des fournisseurs de service qui appellent un traitement véritablement global et une intelligence métier renouvelée. Cette nouvelle complexification de la gestion des infrastructures souvent perçue comme une perte de contrôle au profit des fournisseurs, peut au contraire être un facteur de rétention des talents en France et en Europe**.

a. Le plan Très Haut Débit a eu un impact positif pour développer des compétences en infrastructures réseau adaptées aux besoins des entreprises

Le secteur des infrastructures numériques devrait totaliser 33 000 créations nettes d'emplois sur la période de 2022 à 2030[145], ce qui représente une hausse de 5 %. **Ces créations devraient être largement portées par les projets IoT, qui représenteront à terme la part la plus importante avec près de 40 000 collaborateurs.** Près de 6 000 ingénieurs et 12 000 techniciens de maintenance réparateurs d'équipements sont attendus en renfort sur ces réseaux d'ici 2030. Les emplois dans les datacenters doubleraient également sur cette période, passant de 11 500 aujourd'hui à plus 28 000 en 2030. **Les familles de métiers les plus demandées sont les métiers de l'IT restant ainsi concentrées**

[145] *Le Monde informatique, mai 2023, « Infrastructures numériques : 33 000 créations d'emplois d'ici à 2030 ».*

autour de la sécurité, du *cloud*, de l'architecture réseau, des télécoms, de la gestion et du pilotage de projet, et du domaine développement, test et ops.

Depuis 2023, l'ANSSI a renforcé la visibilité et la certification des formations à travers des initiatives telles que SecNumedu et CyberEduc. Bien que des avancées aient été réalisées, il reste des efforts à fournir. Ces métiers sont accessibles avec des formations telles que le BUT Informatique, qui, depuis la réforme de 2019, inclut une licence professionnelle et une année en alternance. Cette approche favorise une insertion rapide dans le monde professionnel et s'avère particulièrement adaptée aux besoins des infrastructures numériques locales, telles que les *data centers* de proximité, en permettant aux étudiants de s'intégrer dans le tissu économique régional. **Le financement privé de l'apprentissage, soutenu par des dispositifs tels que la contribution unique à la formation professionnelle et à l'alternance (CUFPA), a aussi contribué à attirer des talents et à les former tout au long de leur carrière**.

Le plan THD a été moteur pour ouvrir de nombreuses formations initiales[146] sur le déploiement de réseaux et services de très haut débit avec une grande diversité de formations et de niveaux de porte d'entrée (niveaux 4 à 7). En réponse à la forte demande en déploiement de fibre optique, de nombreuses formations ont vu le jour sur l'ensemble du territoire. Cette initiative a permis d'offrir des parcours variés, adaptés aux niveaux de qualification allant du CAP (niveau 3) aux formations d'ingénieur (niveau 7), rendant accessible une palette de compétences aux demandeurs d'emploi, aux jeunes, et aux personnes en reconversion.

[146] *C. Collot, J. Hespel, A. Tarrerias, 22 mars 2023, Rapport complet : Étude prospective des besoins en emplois et compétences de la filière des infrastructures numériques à l'horizon 2030.*

b. Mais un rapprochement doit désormais s'opérer avec les métiers du traitement de données pour en tirer le plein potentiel

Le manque de passerelles entre les métiers de la connectivité, centrés sur le très haut débit, et ceux du traitement et du stockage des données liés à l'IA, l'edge et l'IoT demeure toutefois un obstacle majeur. Cette fracture freine l'adaptation des compétences à la complexité croissante des infrastructures numériques, où couches matérielles et logicielles s'imbriquent. Le contrat de filière des infrastructures numériques pour 2023-2025[147] met en lumière l'insuffisance de formations adaptées, encore largement orientées vers les infrastructures réseau, sans prendre en compte l'interdépendance croissante avec les solutions de traitement des données dans un continuum cloud-IoT-edge. **Ce plan vise donc à créer des passerelles entre métiers des infrastructures numériques, à développer de nouvelles formations et certifications, et à valoriser l'emploi et les compétences au sein de la filière**. À la suite d'une évaluation de la filière fibre optique à l'horizon 2023, le Comité stratégique de filière (CSF) a choisi de capitaliser sur les compétences déjà acquises, avec pour objectif d'identifier des passerelles de formation exploitables à court, moyen et long terme.

Pour y remédier, des initiatives comme le protocole d'engagement de développement de l'emploi et des compétences[148] (EDEC) ont été mises en place. **Signé en 2021, il vise à développer des passerelles professionnelles et à adapter les parcours aux nouveaux métiers – notamment ceux liés à la 5G et aux territoires connectés.** Grâce à ce protocole, 13 000 entreprises ont pu préserver 140 000 emplois

[147] *Conseil national de l'industrie. (2023). Contrat du Comité Stratégique de la filière Infrastructures Numériques 2023-2025, https://www.entreprises.gouv.fr/files/files/Entites/CNI/2023/contrat-de-filiere-infrastructures-numeriques.pdf.*

[148] *InfraNum. (2021, 2 décembre). Métiers des infrastructures numériques : Signature d'un protocole d'engagement de développement de l'emploi et des compétences (EDEC) avec la filière, https://infranum.fr/metiers-des-infrastructures-numeriques-signature-dun-protocole-dengagement-de-developpement-de-lemploi-et-des-competences-edec-avec-la-filiere/.*

directs, en grande partie dans des TPE-PME. **Cependant, le défi reste immense : l'attractivité des filières et l'adéquation entre formations et besoins industriels nécessitent des efforts concertés pour clarifier les perspectives de carrière et diversifier les compétences enseignées**. Le groupe de travail associé à ce protocole d'engagement s'est ainsi concentré essentiellement sur l'attractivité de la filière des infrastructures numériques en clarifiant les parcours de formation, les perspectives d'évolution, et la diversité des métiers pour soutenir le développement de ce secteur.

L'enjeu réside désormais dans une collaboration plus étroite avec les industriels, qui sont les premiers exposés à l'évolution des usages des infrastructures numériques, qu'il s'agisse des réseaux ou du traitement des données. Des établissements comme Telecom Paris, avec sa filière Grandes infrastructures numériques (GIN) ou la Cisco Networking Academy, illustrent cette dynamique en formant leurs étudiants à la maîtrise des paradigmes réseau, des architectures, des algorithmes, et des technologies émergentes. Ces formations combinent outils théoriques et méthodologiques pour analyser les réseaux actuels et anticiper leurs futures évolutions. De même, des initiatives telles que l'école des passerelles des métiers du numérique (PMN) offrent des cursus spécialisés, conçus en partenariat avec des entreprises et éditeurs, pour répondre aux besoins spécifiques du marché tout en renforçant l'employabilité et les perspectives de carrière des étudiants. **Cependant, ces modèles, bien qu'efficaces, restent une exception alors qu'ils devraient être la norme en France pour répondre aux besoins des entreprises**.

Encadré n° 18 • Focus sur le modèle du Centre National IMT, catalyseur du continuum de l'innovation entre mondes académique et industriel

Inauguré en novembre 2024, le Centre IMT « Réseaux et Systèmes pour la Transformation Numérique » met en place un ensemble d'activités nouvelles qui accélèrent les processus d'innovation, dans une logique de continuum entre les mondes académiques et industriels, en intégrant recherche, innovation et transfert technologique dans des domaines stratégiques tels que les réseaux et les infrastructures numériques, au service de domaines d'activités économiques tels que l'énergie, les transports, l'industrie 4.0 et la santé. En collaboration avec des partenaires économiques et publics, il conçoit des solutions de pointe, comme des infrastructures connectées hautement sécurisées pour les transports, dotées d'une intelligence distribuée pour répondre aux défis des réseaux véhiculaires. Dans le domaine de la santé, il se concentre sur la sécurisation des données médicales et les infrastructures critiques de télémédecine, soutenant ainsi les progrès de la santé connectée.

Le Centre joue également un rôle clé dans plusieurs programmes de recherche internationaux et nationaux, tels que le PEPR « Réseaux du futur » et le programme FRAMExG, pour lesquels l'institut Mines-Télécom est co-leader dans le cadre de la Stratégie Nationale sur les Réseaux du Futur. Grâce à ces engagements, le Centre contribue à l'accélération de l'innovation française dans le domaine du numérique, tout en contribuant au développement de standards globaux qui accompagnent les transformations numériques à l'échelle mondiale.

Le secteur du calcul intensif est très parlant car il est confronté à une pénurie de talents pour renforcer l'accompagnement des utilisateurs de supercalculateurs. Or, cet accompagnement est tout aussi essentiel que la puissance des machines elles-mêmes, sinon plus, pour maximiser l'impact de ces infrastructures. Disposer de machines performantes ne suffit pas ; il faut pouvoir exploiter pleinement leur potentiel grâce à un support technique adapté, des modèles optimisés et des jeux de données de haute qualité. **En Europe, l'enjeu n'est pas de multiplier des modèles de fondation génériques sans application directe, mais de les spécialiser en fonction des besoins métiers**. Par exemple, plutôt que de chercher à rattraper les leaders mondiaux avec des modèles de fondation développés de zéro, il serait plus pertinent de développer des modèles adaptés à des données sectorielles spécifiques, répondant à des usages stratégiques. **Cela souligne un défi crucial pour l'avenir du calcul intensif : former, fidéliser et attirer les talents nécessaires pour accompagner cette transition**. Il ne s'agit pas uniquement de fournir des machines, mais aussi de s'appuyer sur des experts capables de comprendre les besoins des organisations et de les traduire en solutions de calcul sur mesure. **Ce besoin est particulièrement visible dans des centres comme Jean Zay, où la montée en charge des utilisateurs a exigé un doublement des équipes de support, passées de 12 à 25 équivalents temps plein. Avec l'arrivée d'Alice Recoque, cette exigence d'accompagnement devra être anticipée pour garantir une exploitation optimale de la puissance de calcul disponible**.

Les centres de recherche font face à des défis majeurs pour fidéliser les talents essentiels à l'exploitation des supercalculateurs. **Limités par des plafonds d'emplois et souvent contraints de proposer des contrats courts de type CDD, ils peinent à offrir des perspectives professionnelles à long terme. À cela s'ajoute une rémunération nettement inférieure à celle des *hyperscalers*, parfois jusqu'à cinq fois moindre, ce qui rend ces profils hautement convoités**. Il est fréquent que ces experts débutent leur carrière dans un centre de recherche public avant

de rejoindre un *hyperscaler*, attirés par des opportunités financières et professionnelles plus attractives. **Cette situation met également en lumière un problème plus large : l'attractivité des carrières scientifiques, notamment pour les profils féminins, dans un secteur qui peine encore à diversifier ses recrutements**. De plus, certains acteurs privés renforcent leur avantage en intégrant directement la recherche dans leur chaîne de valeur. Certaines entreprises américaines peuvent offrir jusqu'à 1 M$ en crédits de calcul gratuits à des scientifiques, en contrepartie de l'utilisation de leurs outils propriétaires, augmentant ainsi leur dépendance. **Dans ce contexte, les centres publics européens peuvent se différencier en misant sur un accompagnement de qualité, un levier stratégique tout aussi important que la performance des machines elles-mêmes**. C'est l'accompagnement, qui permet aux organisations utilisatrices d'exploiter pleinement les ressources mises à disposition et d'en tirer quelque chose pour leur activité. **Il est crucial que l'Europe renforce cette stratégie pour se démarquer, car la valeur d'un supercalculateur ne réside pas seulement dans ses capacités techniques, mais dans la manière dont ces capacités répondent aux besoins concrets des utilisateurs**.

Recommandation 5

Lancer un projet « commando » pour développer des formations continues rapprochant les métiers de l'infrastructure numérique de réseaux de ceux de l'infrastructure numérique de traitement de données.

Recommandation 5.1 : créer des centres pluridisciplinaires inspirés du modèle IMT pour rapprocher les formations aux métiers du réseau de celles dédiées au traitement des données, en favorisant l'émergence de profils d'« ingénieurs maisons ». Ces centres permettraient de créer des passerelles entre formation initiale et continue, stimulant ainsi la mobilité entre filières et orientant les talents vers des projets industriels concrets.

Par exemple, il faut repenser la formation des équipes des centres de calcul qui accompagnent les entreprises utilisatrices. Aujourd'hui majoritairement axées sur des dimensions techniques ou généralistes, ces formations doivent intégrer une approche métier et usage, ou élargir les recrutements à des profils plus sectoriels.

Recommandation 5.2 : valoriser les engagements préalables entre l'entreprise et le salarié pourrait être envisagé pour assurer un minimum de fidélité. En effet, cela pose le problème de l'investissement privé pour la formation de ces ingénieurs qui seront par la suite fortement sollicités par les entreprises étrangères avec des moyens d'attractivité salariale nettement supérieurs aux nôtres.

3 La compétitivité française de nos infrastructures de réseau est à préserver et à exporter au niveau européen et mondial

3.1. LES ANTENNES RÉSEAU : DES TECHNOLOGIES À LA POINTE SOUFFRANT D'UN MANQUE D'INCITATIONS CIBLÉES QUI DÉTOURNENT D'UNE ADOPTION MASSIVE

a. La 5G : des investissements conséquents mais une adoption modeste

La 5G offre des fonctionnalités améliorées pour le grand public et l'industrie en complément des technologies de connectivité réseau existantes

La 5G a permis de mettre en place des fonctionnalités améliorées tant pour les applications grand public que pour les applications industrielles. Ces dernières générations de réseaux mobiles offrent des télécommunications à faible latence et à débit élevé. En effet, la 5G offre des débits 10 fois supérieurs à ceux de la 4G, et la 6G promet des débits 100 fois supérieurs à ceux de la 5G[149]. En outre, elles ouvrent de nouvelles fonctionnalités, comme le *network slicing* qui permet de créer plusieurs réseaux indépendants sur une même infrastructure physique de réseau pour répondre à des exigences variées. En somme, il s'agit d'attribuer les bonnes ressources à un utilisateur ou un type de trafic selon le besoin.

- Pour le grand public, la 5G offre des vitesses de téléchargement jusqu'à 10 Gbps et une latence très faible d'environ 1 milliseconde,

[149] *Polytechnique insights, mars 2022, « 5G, 6G : quels enjeux pour les nouveaux réseaux de télécommunication », https://www.polytechnique-insights.com/dossiers/digital/5g-6g/.*

grâce aux ondes millimétriques et à la technologie MIMO massive. La technologie MIMO massive utilise de nombreuses antennes pour multiplier les chemins de transmission indépendants, réduisant les interférences et augmentant simultanément la capacité, la vitesse et la fiabilité des réseaux sans fil. Cela garantit une connectivité fiable dans des environnements densément peuplés et permet des usages variés : streaming en ultra-haute définition, travail collaboratif sur des environnements partagés ou téléchargement rapide de fichiers volumineux (10 Go en 7 minutes contre 30 minutes en 4G).

- Pour les applications industrielles, la 5G se distingue par sa fiabilité dans les environnements électromagnétiques complexes : trains, voies ferrées, usines, stades, etc. Elle supporte des communications à faible latence (dites « URLLC ») et la connexion massive d'objets (dite « mMTC »), facilitant des innovations telles que les robots autonomes, les jumeaux numériques et les capteurs intelligents. L'intégration de l'IA aux réseaux 5G renforce ces applications avec des capacités d'apprentissage en temps réel.

La 5G et le WiFi sont à envisager de manière complémentaire et peuvent être utilisés conjointement selon le besoin. Par exemple, dans un stade, la 5G est utilisée pour assurer la connectivité de l'espace extérieur *via* la technologie massive MIMO (Multiple Input, Multiple Output), ce que le WiFi n'est pas en mesure de faire aujourd'hui en raison de la limitation des points d'accès (dits *« access points »*) qui sont limités en densité et en portée et des routeurs qui ne fonctionnent que dans des espaces intérieurs. **Bien que la nouvelle génération de Wi-Fi 7 offre des canaux de 320 MHz, soit le double des 160 MHz disponibles avec la 5G, permettant ainsi de réduire les interférences, cela ne suffit pas dans un stade. En effet, les centaines de milliers de connexions simultanées dépassent les capacités de gestion de trafic du Wi-Fi seul, notamment en termes de latence et de stabilité dans des environnements très denses**.

En revanche, le WiFi 7 est utile dans des environnements industriels dits « chargés », c'est-à-dire des sites où de nombreux appareils IoT communiquent simultanément, comme dans des usines connectées ou des entrepôts intelligents, car les fonctionnalités de « *preamble puncturing* », qui sont une capacité à transmettre des données en utilisant uniquement les parties non occupées d'un canal large, sans perturber les transmissions adjacentes, permettent de gérer les interférences dues au chevauchement des canaux dans des bandes de fréquences partagées. Le chevauchement des fréquences se produit lorsqu'un grand nombre d'appareils IoT utilisent simultanément des bandes de fréquences similaires dans un environnement restreint, créant ainsi des interférences. Cela arrive fréquemment dans les environnements industriels où des dizaines ou centaines d'appareils sont déployés dans un espace confiné et communiquent sur des bandes de fréquence partagées, comme le spectre 2,4 GHz ou 5 GHz utilisé par le WiFi. Ces interférences peuvent ralentir la transmission des données ou provoquer des pertes de connexion. **Néanmoins, le « *preamble puncturing* » ne prend pas en compte la criticité de la donnée lors de sa transmission dans les différents canaux, ce que fait la 5G *via* la technologie de *network slicing***. Cette dernière permet de créer des « tranches » virtuelles du réseau, dédiées à différents types de services avec des niveaux de performance et de sécurité adaptés. **Ainsi, dans des environnements industriels « sensibles », comme les usines produisant des biens critiques ou les infrastructures liées à l'énergie, la 5G est un complément essentiel du WiFi et de la 4G pour assurer une connectivité fiable, prioriser les flux critiques et garantir la continuité des opérations dans des conditions extrêmes**.

Tableau n° 4 • Comparatif des performances des technologies de connectivité réseau

Technologie	Latence	Débit	Usages
1G	~100 ms	~2,4 kbps	Communications vocales analogiques de base, sans transmission de données.
2G	~50-200 ms	~64 kbps	Introduction des SMS, appels vocaux numériques, transmission de données très basiques.
3G	~100-500 ms	~2 Mbps	Accès à Internet mobile, premiers usages multimédias (images, vidéos) et applications web.
4G	~50 ms	~100 Mbps à 1 Gbps	Streaming HD, jeux en ligne, IoT de base, développement massif des applications mobiles.
5G	~1 ms	~10 Gbps	IoT à l'échelle, connectivité dans des environnements denses (stades, festivals), véhicules autonomes, applications critiques dans les environnements industriels extérieurs, villes intelligentes.
6G	<0,1 ms	~1 Tbps (théorique)	Connectivité ultra-rapide et temps réel pour des applications exigeantes : chirurgie à distance, infrastructures d'énergie, couverture de zones blanches, voire spatiales, etc.

Les réseaux de 5G publics : de nombreux cas d'usages faisant appel à la 5G dite « standalone »

La distinction entre 5G non-standalone (NSA) et 5G standalone (SA) repose sur l'infrastructure utilisée et les fonctionnalités offertes.

- La 5G NSA s'appuie sur les infrastructures existantes de la 4G, notamment les antennes radio (RAN), et intègre partiellement des fonctionnalités 5G dans le cœur de réseau évolué de la 4G *(Evolved Packet Core)*. Cela permet d'exploiter les bandes basses de la 4G pour la couverture, tout en ajoutant des fonctionnalités spécifiques à la norme 5G pour offrir des débits améliorés et une meilleure latence.

En somme, la 5G NSA constitue une transition hybride entre la 4G et la 5G, permettant une adoption progressive tout en maintenant une dépendance au réseau 4G.

- En revanche, la 5G SA repose sur une infrastructure entièrement dédiée à la 5G, avec un cœur de réseau spécifique (5G Core) et des antennes radio (RAN) conçues pour les bandes de fréquences allouées à cette technologie. Cela inclut des fréquences hautes pour des débits élevés et des fréquences basses, parfois partagées avec la 4G, pour maximiser la couverture. La 5G SA offre ainsi des performances optimales, notamment pour des cas d'usage avancés comme les réseaux privés, les applications critiques en temps réel, tout en garantissant une latence ultra-faible et une capacité de réseau supérieure.

Pour utiliser des fonctionnalités de *slicing* dans des environnements publics ou privés, il est nécessaire de mettre en place une infrastructure 5G de type « *standalone* ». En effet, dans une configuration hybride (4G+5G), le *slicing* n'est pas possible car cette technologie nécessite une orchestration réseau fine et indépendante, qui n'est pas supportée par le cœur de réseau 4G. Le cœur du réseau dans une configuration NSA repose encore sur l'infrastructure 4G (*Evolved Packet Core* ou EPC), qui n'est pas conçue pour prendre en charge la segmentation dynamique et fine des ressources réseau. Le *slicing* exige un cœur 5G (5G Core ou 5GC), car ce dernier utilise des protocoles spécifiques (comme SBA, *Service-Based Architecture*) permettant de gérer et isoler différentes tranches réseau en fonction des besoins, garantissant des performances et des niveaux de sécurité distincts pour chaque usage.

Néanmoins, il est possible d'introduire progressivement les fonctionnalités SA en s'appuyant sur un réseau public hybride de type NSA. Pendant les Jeux Olympiques de Paris 2024, la 5G SA et NSA ont coexisté pour offrir des services innovants adaptés à des environnements complexes et densément peuplés. Par exemple, un module a

été développé dans l'application officielle des JO pour fournir un accès précis et de qualité à différents flux audio, même dans des lieux bondés. Ce service, réalisé avec la *startup* Odiho, a permis à chaque utilisateur d'accéder librement à la diffusion de contenus audio en direct via son smartphone, en scannant un QR code. Grâce à cette innovation, les spectateurs ont pu bénéficier de flux audio variés, tels que l'audio description, les commentaires en direct fournis par l'OBS *(Olympic Broadcasting Services)* et les annonces du speaker du stade. Les JO de Paris 2024 ont également été l'occasion de tester des solutions inclusives pour les personnes en situation de handicap. Par exemple, l'application Touch 2 See a permis aux spectateurs malvoyants ou non-voyants de suivre les événements sportifs grâce à une tablette interactive reproduisant le terrain en miniature avec un curseur magnétique. Ce système combinait des vibrations et des retours haptiques – une technologie qui permet de transmettre des informations via des sensations physiques, comme des vibrations ou des pressions –, afin de faire ressentir l'intensité et les actions du jeu par le toucher. **Cette expérience a été rendue possible par la disponibilité optimale du réseau 5G hybride NSA, illustrant comment la 5G hybride (NSA/SA) peut servir de tremplin vers des usages plus avancés tout en répondant à des défis opérationnels immédiats**.

Les réseaux de 5G privés : un potentiel inexploité pourtant nécessaire à la compétitivité future des entreprises françaises et à la constitution d'une offre souveraine

Pour les usages industriels, une partie de l'incertitude en matière de déploiement d'environnements 5G SA provient du marché de l'IoT lui-même. Historiquement, des technologies telles que LoRa *(Long Range)* et Sigfox ont été développées pour offrir une connectivité à faible coût et faible consommation énergétique, adaptée à des objets connectés utilisant de petits volumes de données. LoRa, principalement portée par l'Alliance LoRaWAN, a été conçue pour des réseaux longue

portée non cellulaires, pour la gestion d'infrastructures connectées en temps réel, de type compteur d'eau. De son côté, Sigfox proposait un protocole LPWAN *(Low-Power Wide-Area Network)* propriétaire, axé sur des usages similaires, notamment dans le suivi logistique ou les objets connectés industriels simples. **Ces technologies ont toutefois été mises à mal par l'émergence de la 5G elle-même, en raison de sa couverture plus large et une intégration native des besoins IoT dans les standards des télécommunications**. La faillite de Sigfox en 2022 illustre les difficultés pour des acteurs spécialisés à rester compétitifs face à des solutions plus polyvalentes et intégrées comme la 5G. **En outre, les usages IoT avancés au-delà des objets connectés classiques comme les écouteurs, montres connectées ou équipements de domotique peinent encore à être explicités de manière concrète**. Les promesses de la 5G pour des environnements industriels, tels que l'automatisation des chaînes de production, la maintenance prédictive, ou la robotique collaborative en temps réel, nécessitent un développement à grande échelle qui reste encore en phase expérimentale. **Cela alimente une certaine hésitation chez les industriels, qui doivent justifier des investissements importants dans un contexte où les cas d'usage concrets et rentables de la 5G SA ne sont pas encore pleinement établis**.

L'adoption des fonctionnalités avancées de la 5G SA est particulièrement exigeante pour les acteurs industriels, car aucun contrat ne peut être signé sur une infrastructure NSA étant donné que cette dernière ne garantit pas les performances nécessaires à des applications critiques (comme la latence ultra-faible ou le *slicing* réseau). Le passage à la 5G SA représente un investissement initial plus élevé que le WiFi ou la 4G, notamment en raison de la mise à niveau complète des infrastructures réseaux, y compris le déploiement de nouvelles antennes et d'un cœur 5G dédié (5G Core). Une enquête récente, *Equinix Global Tech Trends Survey* 2023, révèle que près de 50 % des responsables informatiques identifient la hausse des dépenses d'exploitation comme le principal obstacle à l'adoption de nouvelles technologies comme la 5G. **Ce qui**

change véritablement avec la 5G, c'est qu'elle force les entreprises à faire un choix stratégique entre des infrastructures fixes et mobiles. Contrairement à la 4G, qui s'installait en complément des réseaux fixes, la 5G SA offre la possibilité de remplacer certaines infrastructures filaires, ce qui nécessite un effort d'acculturation des entreprises pour comprendre ces nouveaux usages et leurs avantages.

Pourtant, dans certains cas, le déploiement de la 5G privée en milieu industriel (notamment dans des environnements extérieurs complexes) peut être plus économique que le déploiement d'une infrastructure WiFi. Par exemple, la 5G SA peut être déployée sans infrastructure 4G préexistante, ce qui réduit les coûts dans des environnements ne disposant pas déjà d'un réseau dense. **Un exemple concret est celui de l'entreprise Hub One, qui utilise un réseau privé 5G dans le cadre d'un projet avec le port autonome du Havre**. Ce réseau permet de gérer la connectivité des véhicules autonomes utilisés pour la logistique portuaire et d'assurer la surveillance en temps réel des infrastructures critiques. Grâce à la faible latence et à la robustesse du réseau 5G, le port peut optimiser ses opérations, tout en réduisant les coûts liés à la maintenance d'un réseau WiFi étendu et dense, qui serait moins adapté à ce type d'environnement extérieur.

Les 115 projets de 5G industrielle en cours en France en 2024[150] se concentrent principalement sur trois domaines : le contrôle qualité, la gestion des coûts énergétiques et la voiture connectée. Pour le contrôle qualité et la gestion des coûts énergétiques, le projet 5G Steel[151], mené par ArcelorMittal, en partenariat avec Ericsson et Orange Business, déploie la 5G dans des sites industriels en régions Hauts-de-France et Grand Est pour améliorer la fiabilité du processus de réception des aciers destinés au recyclage. Grâce à une solution logicielle

[150] *ARCEP, « Tableau de bord des expérimentations 5G industrielles et innovantes en France », avril 2024, expérimentations menées sur plusieurs fréquences: 2,6 Ghz, 3,8 Ghz et 26 Ghz.*

[151] *Orange, Communiqué de presse, 1er Février 2023, 5G Steel : les premiers cas d'usage très haut débit pour décarboner l'industrie de l'acier et travailler en mobilité sur un site industriel.*

supportée par un réseau 5G SA, l'application permet une inspection visuelle des matériaux, éliminant ainsi les risques liés à l'utilisation de supports papier. Cette innovation réduit significativement les délais de traitement et de pilotage des activités, en particulier dans les cas de litiges fournisseurs où des photos en temps réel peuvent être prises et partagées instantanément. Dans le domaine de la voiture connectée, le projet 5G Open Road, permet d'améliorer l'assistance à la conduite des véhicules automatisés. Ce consortium européen, composé de huit partenaires clés (Valeo, Stellantis, Renault, Lacroix City, Capgemini, Nokia, Bouygues et Vedecom), s'appuie sur des réseaux 5G pour sécuriser et fluidifier le trafic aux intersections équipées. En connectant des plateformes et des algorithmes en temps réel, ce projet permet aux véhicules d'accélérer leur prise de décision à des carrefours complexes, contribuant ainsi à améliorer la fluidité et la sécurité du trafic.

Ces expérimentations restent toutefois limitées aux grandes entreprises et aux projets de type *greenfield*, alors qu'elles devraient également inclure les PME-ETI locales, les zones moins denses et les projets de type *brownfield*. Les projets *greenfield* impliquent la création d'infrastructures numériques ou industrielles entièrement nouvelles, sans contraintes liées à des installations existantes. Ils permettent d'intégrer les technologies les plus récentes et offrent une flexibilité maximale, mais nécessitent des investissements importants et des délais plus longs. Un exemple typique est la construction de nouveaux *data centers* dans des zones émergentes. À l'inverse, les projets *brownfield* modernisent ou réutilisent des infrastructures existantes, ce qui est souvent moins coûteux et plus rapide. Cependant, cette approche est limitée par les contraintes des équipements en place. Elle est particulièrement adaptée aux environnements équipés, comme la mise à niveau des réseaux télécoms pour supporter la 5G, maximisant ainsi l'efficacité des ressources existantes.

En France, les exemples à grande échelle, comme le projet de la SNCF à Rennes dédié à la maintenance prédictive[152] des infrastructures

ferroviaires, restent rares, reflétant une faible pénétration de la 5G industrielle à l'échelon local. Pourtant, il existe des cas concrets de projets de 5G dans des zones rurales. Par exemple, les tests menés par Vodafone au Royaume-Uni dans le domaine de l'agriculture intelligente démontrent le potentiel de cette technologie pour moderniser des secteurs clés. Ces expérimentations incluent la surveillance en temps réel des cultures via des capteurs IoT connectés à un réseau 5G, permettant une optimisation des ressources comme l'eau et les engrais, tout en réduisant les coûts d'exploitation. **Dans ce type d'usages, la 5G offre une meilleure sécurité, car elle permet une transmission de données chiffrées avec une faible latence, ce qui réduit les risques de vulnérabilités ou d'interruptions dans des processus critiques comme la gestion des équipements agricoles ou des drones autonomes**. Elle utilise également moins d'énergie que d'autres solutions, car elle peut fonctionner de manière optimisée grâce à des antennes plus efficaces et à la capacité d'allouer dynamiquement des ressources réseau uniquement lorsque cela est nécessaire *(network slicing)*, ce qui réduit la consommation énergétique globale du réseau et des appareils connectés.

Encadré n° 19 • Gigalis, un exemple d'usage régional de la 5G pour des besoins stratégiques

Gigalis, groupement d'intérêt public en Pays de la Loire, illustre comment une approche régionale peut s'appuyer sur la 5G pour répondre à des besoins numériques spécifiques. Par exemple, dans le secteur industriel, Gigalis utilise la 5G privée pour optimiser les chaînes de production en temps réel, grâce à une

[152] *Rapport de la mission 5G industrielle, Philippe Herbert, président de la Mission 5G industrielle avec la collaboration de Lucas Gravit, Direction générale des Entreprises, en qualité de rapporteur, mars 2022.*

connectivité ultra-fiable et une latence minimale. Cette application permet aux entreprises locales de moderniser leurs infrastructures tout en renforçant leur compétitivité.

Au-delà des besoins industriels, Gigalis déploie également la 5G pour des projets critiques, comme la gestion des hôpitaux connectés ou des infrastructures de transport intelligentes. Ces initiatives montrent comment des réseaux dédiés, opérés localement, peuvent garantir une résilience accrue et une sécurité renforcée, en particulier pour des données sensibles.

En associant connectivité, souveraineté et cybersécurité, Gigalis propose un modèle concret de transformation numérique régionale, s'appuyant sur des technologies avancées pour renforcer l'efficacité et l'autonomie des acteurs publics et privés. Cet exemple montre que des initiatives locales peuvent jouer un rôle clé dans l'adoption stratégique de la 5G, tout en répondant aux enjeux de souveraineté numérique.

Conscient de ces enjeux, le Gouvernement a intégré dans sa stratégie numérique 2024-2030 plusieurs mesures visant à soutenir l'adoption de la 5G industrielle. Ces mesures incluent la facilitation de l'accès à des fréquences dédiées, afin de réduire les frais de redevance d'utilisation, et un plan de soutien aux industriels souhaitant déployer des réseaux 5G mobiles privatifs. **Ces actions répondent aux constats de la mission sur la 5G industrielle, commandée par la DGE en 2022, qui avait mis en lumière plusieurs freins institutionnels et techniques :** accès limité aux fréquences, manque de maturité de l'écosystème, disponibilité insuffisante des équipements et services, et difficultés à mobiliser les compétences nécessaires. La mission avait également relevé des interrogations sanitaires, environnementales et

sociétales, qui continuent de peser sur les décisions des acteurs industriels. **Aujourd'hui, les conclusions de cette mission servent de feuille de route pour guider les politiques publiques et structurer un écosystème industriel autour de la 5G**. Cela inclut des actions concrètes pour lever les incertitudes, renforcer les capacités techniques des entreprises et encourager les partenariats public-privé dans des projets innovants. **Cette mission peut également s'appuyer sur le programme de la Commission européenne *Connecting Europe Facility***, doté de 34 Mds €, pour soutenir trois secteurs : le transport (26 Mds €) avec des connexions transfrontalières et la mobilité durable ; l'énergie (6 Mds €) avec des réseaux intelligents ; et le numérique (2 Mds €) pour déployer la 5G, des infrastructures interconnectées (*cloud*, câbles sous-marins) et augmenter les capacités réseau des *data centers* européens (jusqu'à 1 Tb/s d'ici 2030). **Ce programme s'appuie essentiellement sur des projets pilotes incluant la 5G dans les « communautés intelligentes » et le long des corridors de transport**.

Une piste prometteuse pour cette mission serait d'utiliser la 5G privée comme enabler du continuum cloud-edge-réseau-IoT. Cette approche permettrait de répondre aux besoins de sécurité et de souveraineté sur les données sensibles tout en couvrant les usages critiques (cf. graphique n° 9). En offrant une maîtrise du réseau « bout en bout », la 5G privée garantit le traitement des flux de données dans des environnements sécurisés et réduit les risques liés aux interférences ou aux cyberattaques. **Cette technologie pourrait renforcer la compétitivité des entreprises françaises, notamment dans des environnements industriels sensibles, et consolider la souveraineté européenne dans les infrastructures numériques**. Dans le secteur spatial, la 5G privée offre également des opportunités, comme le montrent des projets de satellites en orbite basse. OneWeb, partiellement détenue par Airbus, associe déjà constellations LEO et 5G pour offrir une connectivité haut débit, y compris dans les zones peu couvertes. Des *startups* comme Kinéis ou New Space explorent également des solutions innovantes en combinant spectre 5G et satellites, sur des modèles inspirés

de Starlink. D'autres pays, comme les États-Unis et l'Inde, avancent déjà sur ces sujets. **En septembre 2024, la Maison Blanche a annoncé un partenariat stratégique avec l'Inde pour développer des projets pilotes en 5G dans des secteurs clés tels que la santé, l'agriculture et la logistique**. Pour rester compétitive, la France doit accélérer son adoption de ces technologies et structurer ses propres projets, en s'appuyant sur son écosystème technologique et industriel.

Les solutions Platform-as-a-Service (PaaS) : le levier incontournable pour capter la valeur ajoutée de la 5G

Le paysage des infrastructures numériques de connectivité est aujourd'hui caractérisé par un phénomène dit de « portage » des fonctions autrefois gérées par du matériel *(hardware)* vers des logiciels *(software)*. Cela signifie que des tâches réseau, comme la gestion des antennes ou du signal, peuvent désormais être exécutées sur des serveurs standards (soit des solutions de type PaaS) au lieu d'équipements spécifiques (comme les antennes). Ce nouvel état de fait est amené à prédominer en raison de l'augmentation de la puissance de calcul et de la standardisation désormais plus précise de ces fonctions et de leurs interfaces. Des fonctions réseau autrefois propriétaires peuvent être virtualisées et optimisées en temps réel pour répondre à des besoins spécifiques, notamment industriels ou critiques.

C'est dans ce cadre que se sont développées les initiatives de type Open RAN (ou O-RAN) regroupant opérateurs télécoms et acteurs du cloud. Open RAN peut se comprendre comme un concept visant à ouvrir et standardiser les interfaces et composants du réseau d'accès radio (RAN). Par exemple, l'initiative O-RAN Alliance, créée par des acteurs majeurs des télécoms (comme AT&T, Deutsche Telekom, Orange et Telefonica), élabore des spécifications techniques pour permettre l'interopérabilité entre les différents composants du réseau RAN, tels que les antennes, les unités radio et les logiciels de contrôle.

Ces spécifications ne sont pas des normes officielles (comme celles définies par l'ETSI ou le 3GPP), mais elles s'appuient sur ces normes pour favoriser un écosystème ouvert et modulaire. De nombreuses initiatives et projets pilotes sont menés dans le monde pour tester et déployer des réseaux Open-RAN. Ces projets incluent des tests pour les réseaux 4G/5G, notamment par des opérateurs comme Rakuten au Japon ou Dish Network aux États-Unis. Cette approche facilite l'interopérabilité en permettant à des équipements de différents fournisseurs de fonctionner ensemble, réduisant ainsi la dépendance à un acteur unique. Elle diminue les coûts grâce à l'utilisation de logiciels *open source* et de matériels standards, tout en offrant une flexibilité pour adapter rapidement les réseaux à de nouveaux besoins. **Cette transition vers une virtualisation des fonctions réseau et une standardisation fine transforme profondément les infrastructures de télécommunication, ouvrant des opportunités tout en posant des défis en matière de sécurité et de coordination**.

Dans le contexte de l'émergence de la 5G, les États-Unis ont adopté une stratégie visant à reprendre un rôle clé dans les télécommunications en s'appuyant sur leurs avantages comparatifs dans le secteur du logiciel, notamment la *Silicon Valley*. Plutôt que de concurrencer directement les grands équipementiers télécoms, ils se sont positionnés sur des segments à forte valeur ajoutée, comme les plateformes PaaS, les infrastructures *cloud-native*[153] et les logiciels pour systèmes distribués. Cette stratégie leur permet de capter la valeur ajoutée de la 5G et de se préparer aux futures générations de connectivité.

Contrairement à l'Europe, où les acteurs risquent d'être limités à la fabrication d'antennes, les États-Unis exploitent leurs forces dans le développement logiciel et les infrastructures du continuum cloud-edge-IoT-réseaux aujourd'hui dominées par les *hyperscalers*. Pour

[153] *Architecture conçue spécifiquement pour tirer parti des capacités du cloud (scalabilité, flexibilité, et automatisation) en utilisant des technologies comme les conteneurs, les microservices et l'orchestration pour exécuter des applications de manière distribuée.*

y parvenir, les entreprises américaines ont massivement investi les initiatives de type Open RAN et Open Core[154], qui favorisent des infrastructures ouvertes et *cloud-native*. Ces modèles permettent d'exécuter des logiciels standards sur des infrastructures matérielles classiques, accessibles aux grands acteurs du *cloud* américains. Cette ouverture a permis aux États-Unis de se repositionner dans un secteur où ils avaient perdu leurs équipementiers historiques (Lucent Technologies, Motorola, Nortel). **En misant sur des infrastructures centrées sur les logiciels et les capacités *cloud*, les États-Unis ont recentré l'innovation en matière de connectivité autour de leurs forces traditionnelles : les logiciels de type *cloud/edge* et les solutions PaaS**. Les capacités radio, autrefois intégrées dans des équipements propriétaires, sont désormais concentrées dans les antennes, tandis que des acteurs comme Nokia ou Ericsson s'associent aux *hyperscalers* pour optimiser les architectures 5G. L'accord signé en novembre 2024 entre Nokia et Microsoft[155] pour fournir des commutateurs dans les *data centers* illustre ce modèle : **les *hyperscalers* deviennent incontournables pour intégrer les services de télécommunications dans des infrastructures *cloud* et fournir des équipements de routage IP**. Cette stratégie centrée sur le logiciel et le *cloud* permet aux États-Unis de transformer un ancien désavantage en atout stratégique, consolidant leur leadership dans la chaîne de valeur des infrastructures numériques.

La domination américaine sur les segments stratégiques des télécommunications soulève la question du rôle que l'Europe, et particulièrement la France, peut jouer dans cette reconfiguration. **Les acteurs français et européens disposent pourtant d'une expertise unique dans la gestion intégrée des systèmes matériels et logiciels, notamment dans la conception systémique et l'orchestration d'infrastructures complexes**. Malgré cela, ils peinent à se positionner sur les segments à

[154] Par exemple, 64 % des entreprises de l'Open RAN Policy Coalition sont américaines.

[155] Zone Bourse, « Nokia signe un accord pluriannuel avec Microsoft », novembre 2024.

forte valeur ajoutée, comme le développement de solutions logicielles et l'optimisation des architectures *cloud-native*, laissant les *hyperscalers* américains dominer ces domaines clés. Pour éviter de rester cantonnée au rôle de fabricant d'équipements, la France doit adopter une stratégie proactive. Cela passe par le développement de logiciels adaptés, l'optimisation des infrastructures existantes et des collaborations renforcées avec d'autres acteurs européens. **Investir plus largement dans des initiatives comme Open RAN et Open Core serait un levier essentiel pour regagner en compétitivité. Par exemple, ouvrir les réseaux *via* des APIs standardisées permettrait de créer des écosystèmes innovants en collaboration avec des développeurs d'applications et des entreprises technologiques**. Cette approche pourrait stimuler l'émergence de nouveaux services tout en maximisant la valeur des infrastructures déjà déployées. Cependant, cette ouverture pose aussi des défis importants, notamment en matière de sécurité. En intégrant des solutions tierces, les opérateurs risquent de perdre le contrôle sur certaines fonctions stratégiques du réseau. Par exemple, une dépendance excessive envers des fournisseurs de logiciels non européens pourrait accroître la vulnérabilité face aux cyberattaques ou à des ruptures d'approvisionnement critiques. Une gestion rigoureuse des partenariats et un cadre réglementaire solide seront donc nécessaires pour limiter ces risques tout en assurant la souveraineté technologique européenne.

Les solutions Paas jouent un rôle clé dans cette reconfiguration stratégique, en fournissant une plateforme centralisée pour le développement, le déploiement et la gestion des applications réseaux. En mutualisant les infrastructures et en automatisant des tâches complexes, elles permettent de réduire les coûts opérationnels et de raccourcir significativement les délais de mise sur le marché des nouvelles solutions. Dans le cadre d'initiatives comme Open RAN, ces plateformes ouvrent des possibilités concrètes : par exemple, l'optimisation dynamique de la couverture réseau en fonction des besoins des utilisateurs ou la création de réseaux privés pour des secteurs industriels critiques,

comme les usines connectées ou les ports autonomes. **Des exemples compétitifs de solutions PaaS incluent des initiatives européennes comme celles développées par Orange avec son programme de virtualisation des réseaux, ou encore des collaborations comme celle entre Nokia et des acteurs du *cloud* pour déployer des infrastructures *cloud-native* optimisées**. Ces initiatives montrent qu'en misant sur les plateformes PaaS, l'Europe peut non seulement rattraper son retard, mais aussi se positionner sur des segments stratégiques de la 5G et de la 6G, en alignant innovation technologique et contrôle souverain des infrastructures critiques.

b. La 6G : un exemple de freins normatifs à une commercialisation compétitive

Une course technologique engagée en matière de 6G

Bien que la 6G soit encore en développement, certains champions nationaux, à l'image du Sud-coréen Samsung, estiment que la première commercialisation pourrait être effective en 2028, et le déploiement à l'échelle en 2030[156]. Des pays avancés sur le plan technologique comme la Corée du Sud ou Singapour explorent en outre la combinaison 5G et 6G pour construire des laboratoires industriels de villes intelligentes. **En 2023, la Chine, les États-Unis et la Corée représentaient ensemble la majorité des brevets 5G, avec plus de 65 000 familles déposées sur les 75 000 déclarées à l'ETSI**[157].

La France prépare l'arrivée de la 6G avec une enveloppe de 65 M€ pour les réseaux du futur dans le cadre du plan France 2030 qui servira à s'impliquer dans la standardisation à venir et à renforcer la

[156] *SFR Actus, « Samsung se prépare à l'arrivée de la 6G », juillet 2020.*

[157] *Questel, « Brevets essentiels de la norme 5G : la réalité derrière les chiffres », janvier 2024.*

coopération européenne. Ce programme[158], piloté par le CEA, le CNRS et l'Institut Mines-Télécom, vise à soutenir l'excellence scientifique et l'innovation industrielle dans le domaine des réseaux de nouvelle génération.

Le programme prévoit dix projets ciblés et trois vagues d'appels à projets ou manifestations d'intérêt, couvrant l'ensemble de la chaîne de valeur des réseaux du futur. Il s'articule autour de quatre axes principaux :

1. développer les usages de la 5G pour renforcer la compétitivité économique ;
2. élaborer des solutions souveraines françaises ;
3. consolider la recherche et le développement pour les futures générations de réseaux ;
4. renforcer l'offre de formation pour attirer des talents internationaux.

Dans la même dynamique, l'Europe a financé le *Smart Networks and Services Joint Undertaking* (SNS JU[159]) à hauteur de 1,8 Md € de 2021 à 2027[160] pour sécuriser les positions européennes le long de la chaîne de valeur 6G. L'objectif est de rapprocher les acteurs (académiques, industriels, TPE-PME) et de nouer des partenariats avec des pays alliés (Japon, États-Unis). Ce plan vient en complément d'initiatives nationales plus ou moins matures, à l'image de la Finlande (leader du SNS JU) et de son programme 6G-Finland[161] ou de la plateforme « France 6G » lancée récemment pour coordonner les efforts publics et privés. L'enveloppe dédiée reste toutefois intégrée au sein du projet de recherche plus large « Réseaux du Futur », doté de 65 M€, investissement

[158] *CNRS, « France 2030 : lancement du programme de recherche "Réseaux du futur" », juillet 2023.*

[159] *Le SNS JU « vise à assurer à l'Europe une position de leader industriel dans les domaines de la 5G et de la 6G », https://smart-networks.europa.eu/.*

[160] *Parlement Européen, "The path to 6G", janvier 2024.*

[161] *Site : https://6gfinland.fi/.*

timide face aux 700 M€ dédiés à la 6G par l'Allemagne[162] La 6G est ainsi devenue, avec l'IA et le quantique[163], une priorité des discussions du *EU-US Trade Technology Council*[164] (TTC), et est considérée comme une technologie d'importance vitale « pour la prospérité globale et la sécurité », ayant entraîné la déclaration d'un engagement d'échange réciproque d'informations et d'opportunités entre les deux groupements d'États. Cela s'incarne dans des projets comme Hexa-X (2021-2023), financé par la Commission européenne dans le cadre du partenariat public-privé 5G Infrastructure PPP, et rassemblant une trentaine d'organisations – opérateurs télécoms, géants technologiques, instituts de recherche et universités – autour de cas d'usage 6G innovants tels que l'imagerie et la localisation de haute précision, les robots industriels ultra-coordonnés ou encore les services immersifs à très haute bande passante. Dans la continuité, le projet Hexa-X-II, doté d'une enveloppe de 900 M€, soutient 63 projets, avec 16 nouvelles initiatives prévues pour début 2025, dont le projet SUSTAIN-6G, axé sur la durabilité de la 6G (environnementale, économique et sociétale). Le cadre concurrentiel est toutefois rude, la Chine exploitant déjà le spectre de la bande « Upper 6 Ghz »[165] (6 425-7 125 MHz), et les États-Unis progressant dans l'évaluation de la compatibilité 6G et services existants dans le cadre de la WRC-27 AI 1.7.

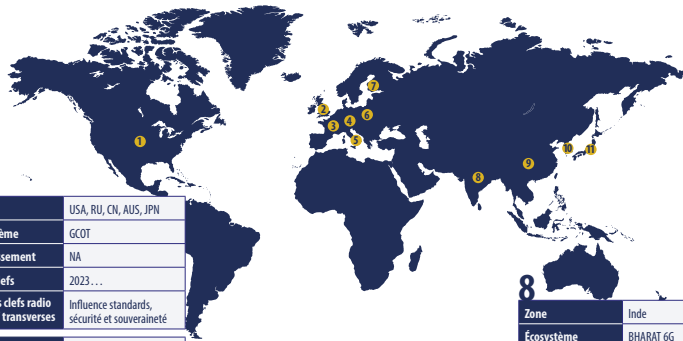
[162] Phonehouse, « 6G : l'Allemagne débloque 700 millions d'euros de financement », avril 2021.

[163] Commission européenne, « 6G Outlook », mai 2023.

[164] Commission européenne, "EU-Japan Summit: strengthening our partnership", mai 2022.

[165] RCRwireless news,"China claims world's first '6G' field test network", juillet 2024.

Graphique n° 32 • Planisphère des initiatives 6G dans le monde



1

Zone	USA, RU, CN, AUS, JPN
Écosystème	GCOT
Investissement	NA
Dates clefs	2023…
Technos clefs radio / cœur / transverses	Influence standards, sécurité et souveraineté

Écosystème	PAWR
Investissement	100 M $
Dates clefs	NA
Technos clefs radio / cœur / transverses	Open RAN, SW defined, satellite / Edge / Photonique

Écosystème	5G Americas
Investissement	1,6 Md €
Dates clefs	2022-2032
Technos clefs radio / cœur / transverses	Thz, IA, Open RAN / IA, flexibilité versus usage / Convergence virtualisation

Écosystème	ATIS next G alliance
Investissement	NA
Dates clefs	NA
Technos clefs radio / cœur / transverses	mmWave, THz, MIMO / IA IA / composants

2

Zone	Royaume-Uni
Écosystème	Open Network
Investissement	250 M £ (287 M €)
Dates clefs	2022-2025
Technos clefs radio / cœur / transverses	Open RAN

3

Zone	France
Écosystème	France 2030
Investissement	735 M €
Dates clefs	2021-2027
Technos clefs radio / cœur / transverses	THZ, mmWave / Cloud, sensing / Composants, cyber

4

Zone	Union Européenne
Écosystème	6G SNS - IA
Investissement	372 M €
Dates clefs	2022-2024
Technos clefs radio / cœur / transverses	THz, MIMO / IA, deepedge / Optique, NTN, IA

5

Zone	Italie
Écosystème	Restart
Investissement	2,1 M €
Dates clefs	2023-2025
Technos clefs radio / cœur / transverses	Radio intelligente / Cœur configurable / Photonique

6

Zone	Allemagne
Écosystème	6G platform
Investissement	700 M €
Dates clefs	2021-2026
Technos clefs radio / cœur / transverses	mmWave, THZ, RAN adaptatif / Plateformes ouvertes / Sensing network, IA, sécurité

7

Zone	Finlande
Écosystème	6G flagship
Investissement	251 M €
Dates clefs	2018-2026
Technos clefs radio / cœur / transverses	THz, latence, Tbits / Distribution, IA

8

Zone	Inde
Écosystème	BHARAT 6G
Investissement	1,1 Md €
Dates clefs	2023-2030
Technos clefs radio / cœur / transverses	Bandes, THz, surfaces intelligentes / Plateformes multi usages / Terminaux/ objets (modem)

9

Zone	Chine
Écosystème	ILT-2030
Investissement	30 Md $ (vision US)
Dates clefs	2021-2030
Technos clefs radio / cœur / transverses	MIMO, THz, surfaces intelligentes / NTN, distribution / virtualisation / Jumeau, multisensoriel

10

Zone	Corée du Sud
Écosystème	K-Network
Investissement	500 M € (200 M € à moyen terme)
Dates clefs	2030
Technos clefs radio / cœur / transverses	Open RAN extreme MIMO, THz / NTN / IA, sécurité

11

Zone	Japon
Écosystème	Beyond5G
Investissement	1,83 Md € (500 M € à moyen terme)
Dates clefs	2021-2030
Technos clefs radio / cœur / transverses	Latence, connectivité / Virtualisation open, NTN / Photonique, sécurité

Source : Situation des écosystèmes 6G étrangers, Filière Infrastructures Numériques, 2024.

La nécessité d'une stratégie de normalisation européenne alignée avec les réalités du terrain

Pour renforcer la compétitivité des acteurs européens, il est essentiel de développer une stratégie de normalisation alignée avec le déploiement des solutions de connectivité, afin de faciliter leur commercialisation une fois qu'elles atteignent des niveaux de maturité technologique (TRL) élevés. Les brevets essentiels aux normes (BEN[166]) jouent un rôle central dans ce processus, car ils couvrent les composants indispensables pour garantir l'interopérabilité des dispositifs conformes aux normes 5G ou 6G. Les détenteurs de BEN peuvent percevoir des redevances de la part des entreprises utilisant leurs technologies, consolidant ainsi leur position stratégique et financière. **Encourager davantage d'acteurs européens, et notamment français, à déposer des BEN permettrait de positionner les innovations locales dans les spécifications mondiales, favorisant leur adoption à grande échelle tout en assurant une indépendance technologique**. Contribuer activement à la standardisation grâce aux BEN, c'est aussi avoir l'opportunité de définir les règles du jeu, en plaçant les technologies européennes au cœur des déploiements globaux de la 5G aujourd'hui et de la 6G demain.

[166] *Commission européenne, Questions et réponses sur les brevets essentiels liés à une norme, 27 avril 2023, https://ec.europa.eu/commission/presscorner.*

Encadré n° 20 • Précisions sur le fonctionnement des BEN

Les titulaires de brevets ont le droit d'exclure d'autres personnes de l'utilisation de leur technologie brevetée et, à l'inverse, ils peuvent percevoir des redevances de licence lorsqu'ils autorisent l'utilisation de leur technologie brevetée.

Les BEN sont des brevets ordinaires. Toutefois, il existe deux considérations importantes les concernant : (1) ils sont techniquement essentiels à la mise en œuvre d'une norme volontaire, élaborée collectivement, et peuvent faire l'objet d'un engagement FRAND *(Fair, Reasonable, and Non-Discriminatory)*, c'est-à-dire un engagement à concéder des licences à des conditions justes, raisonnables et non discriminatoires ; et (2) comme les normes sont ouvertes, les technologies protégées par les BEN sont, de fait, régulièrement implémentées par des utilisateurs qui ne prennent pas de licence au préalable.

Le montant des redevances de BEN dépend ainsi de multiples facteurs dont la valeur de l'innovation couverte par le brevet dans le contexte de son utilisation, du type de produit dans lequel la technologie est intégrée et d'accords éventuels entre détenteurs et utilisateurs de brevets, qui peuvent inclure des tarifs fixes, des pourcentages de vente ou des plafonds.

Les BEN s'inscrivent dans un écosystème d'ensemble avec des entreprises innovantes qui apportent les résultats de leurs investissements en R&D sous forme de contributions pour développer des normes ouvertes et créer de nouveaux produits basés sur les technologies brevetées. Elles accordent ensuite des licences sur leurs BEN à des conditions FRAND à tout utilisateur qui en fait la demande.

Néanmoins, le projet de réforme des BEN présenté en avril 2023 par la Commission européenne, bien qu'animé par des intentions louables, risque de pénaliser la compétitivité des entreprises européennes en 6G. Ce texte vise à encadrer les redevances liées aux BEN pour offrir une plus grande prévisibilité, favoriser l'adoption des technologies brevetées, et éviter que des tarifs de licence imprévisibles ou excessifs freinent l'adoption.

Cependant, ce projet de réforme se heurte à quatre difficultés majeures qui pourraient pénaliser la compétitivité des entreprises européennes à la pointe en 6G.

1. **Des coûts réglementaires disproportionnés :** le projet de réforme impose des charges administratives supplémentaires, notamment par le biais de procédures en quatre étapes (enregistrement, contrôles d'essentialité, conciliation et calcul des redevances), qui alourdissent inutilement les processus des entreprises. Ces règles, redondantes avec les rôles déjà assurés par l'ETSI et les tribunaux européens, ralentissent l'accès à la justice et augmentent les coûts pour les entreprises. Selon une étude commandée par la DG GROW[167], les problèmes liés aux licences de BEN ne sont pas suffisamment importants pour justifier une telle intervention. Des organismes de normalisation comme l'AFNOR et le CEN-CENELEC partagent ces préoccupations, craignant une pénalisation de l'innovation européenne.

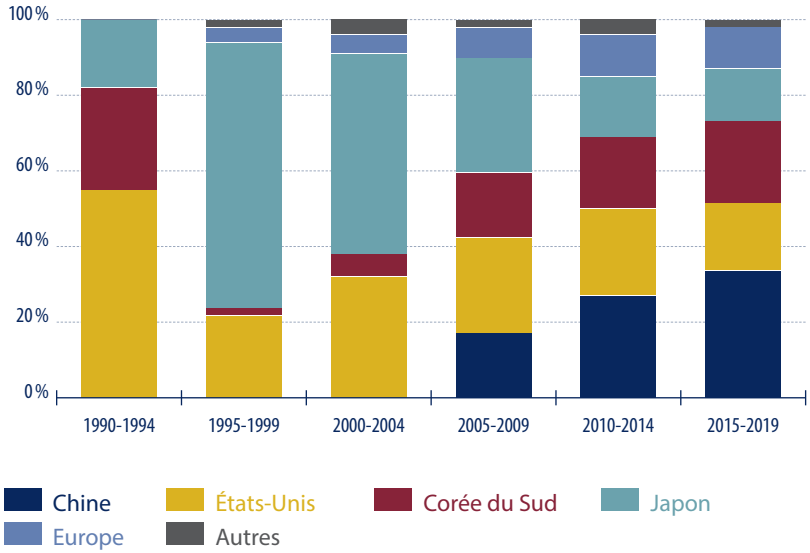
2. **Une approche quantitative et non qualitative :** le projet de réforme privilégie une logique de « guerre de chiffres » en valorisant le nombre de brevets déposés plutôt que leur qualité intrinsèque. **Cette approche avantage des pays comme la Chine, qui excelle dans la multiplication des dépôts, au détriment des entreprises européennes, souvent responsables de technologies**

[167] *Commission européenne, 2023, "Economic impact assessment on Standard Essential Patents (SEPs)", https://op.europa.eu/en/publication-detail/.*

fondamentales mais moins prolifiques en termes de nombre de brevets. Les contrôles d'essentialité proposés, bien qu'utiles pour vérifier la pertinence technique des brevets, ignorent leur véritable valeur. En réalité, tous les brevets essentiels ne se valent pas : certains introduisent des avancées technologiques majeures, tandis que d'autres se contentent d'améliorations mineures ou d'assurer l'interopérabilité. **La qualité d'un brevet ne peut être appréciée que de manière spécifique, par des évaluations indépendantes comme les examens de validité menés par l'Office européen des brevets (OEB) ou les tribunaux spécialisés**. Par exemple, un brevet valide dans plusieurs juridictions et utilisé avec succès dans des actions contre des contrefacteurs constitue un indicateur clair de sa valeur. **Une procédure réglementaire standardisée ne peut se substituer à ce type d'analyse détaillée**. En négligeant cette dimension qualitative, la réforme fragilise des entreprises européennes de pointe comme Nokia ou Ericsson, qui investissent massivement en R&D pour conserver leur compétitivité. Ces acteurs consacrent jusqu'à 19 % de leur chiffre d'affaires à la recherche et développement – soit près du double de la plupart des grandes entreprises technologiques[168]. Ils dépendent de revenus de licences équilibrés pour financer leurs innovations. À l'inverse, des redevances mal calibrées pourraient compromettre leur capacité à développer les technologies coûteuses mais essentielles qui soutiendront les réseaux mobiles de demain.

[168] *Rapport Annuel de Nokia, 2023, p. 75.*

Graphique n° 33 • Part de brevets déposés par région depuis 1990



Source : Politico Research and Analysis, Standards at stake : the EU's plan to balance SEP licensing and innovation, 16 October 2024, p. 6.

3. **Une mise en œuvre irréalisable :** le projet de réforme impose que les redevances soient fixées avant même la mise en œuvre des normes, une chronologie inadaptée qui va à l'encontre des pratiques actuelles de standardisation. En effet, la valeur réelle des BEN ne peut être estimée qu'une fois la norme déployée et testée dans des conditions pratiques. **Surtout, elle ne répond pas aux véritables obstacles rencontrés par les PME, TPE et ETI, qui souffrent davantage d'un manque de compétences et de ressources pour gérer les licences BEN que de litiges autour de ces brevets**. En effet, l'immense majorité des négociations de BEN

– près de 95 à 98 %[169] – se concluent à l'amiable, ce qui démontre que les conflits dans ce domaine restent exceptionnels. Une enquête menée par la DG GROW auprès de 3 192 PME a révélé que seules 39 entreprises (1,15 %) ont signalé des difficultés liées aux BEN[170]. **Avant d'ajouter des procédures administratives, il faut donc plutôt inciter les entreprises à participer à la normalisation *via* les BEN**. En effet, les entreprises qui investissent le plus en R&D en matière de 6G s'appuient aussi sur des écosystèmes dynamiques où les PME jouent un rôle crucial, que ce soit comme partenaires ou comme utilisatrices de ces innovations. **Plutôt que d'imposer des réglementations complexes et coûteuses, le tissu de PME-ETI a besoin de soutien concret, notamment en matière d'accompagnement et d'expertise pour comprendre et gérer les licences de BEN**. En alourdissant les procédures et les obligations, la réforme risquerait de freiner leur participation à l'élaboration des BEN.

4. **Un impact géopolitique sous-estimé :** le cadre proposé ignore les dynamiques géopolitiques essentielles. **Des acteurs non européens, comme Huawei ou ZTE, bénéficient de subventions publiques massives, tandis que les entreprises européennes dépendent uniquement de leurs revenus de licences pour financer leurs innovations**. L'analyse d'impact de cette réforme souligne un risque de suppression des redevances FRAND *(Fair, Reasonable, and Non-Discriminatory)* qui affaiblirait la position concurrentielle de l'Europe face à des géants comme Samsung, qui peut subventionner ses réseaux grâce à d'autres activités.

En l'état, cette réforme risque d'étouffer la capacité des entreprises européennes à investir dans la R&D en matière de 6G et autres technologies critiques[171]. Alors que la connectivité avancée reste l'un

[169] *For Nokia, 2% of patent licensing agreements involve litigation. See Licensing principles I Nokia.*

[170] *Impact Assessment Report, SWD(2023) 124 final, p. 63.*

[171] *Politico Research and Analysis, Standards at stake: the EU's plan to balance SEP licensing and innovation, 16 octobre 2024.*

des rares domaines où l'Europe a le potentiel de jouer un rôle de premier plan[172], des politiques mal calibrées pourraient affaiblir cette position. Une révision profonde, fondée sur une analyse pragmatique des impacts économiques et géopolitiques, est indispensable pour préserver l'innovation et la compétitivité européennes dans les technologies de demain.

Recommandation 6

Accélérer le déploiement de la 5G en milieu industriel, au moins sur les projets *greenfield*, en ciblant résolument les besoins des entreprises utilisatrices (TPE-PME-ETI).

La mise à l'échelle des infrastructures numériques est une réalité, avec la 5G appelée à remplacer progressivement la 4G. Si cette transition semble assurée à court et moyen terme, elle reste incertaine sur le long terme. En effet, les entreprises seront appelées à faire des choix et des investissements sur des déploiements encore coûteux mais dont la valeur reste à démontrer : 5G versus Wifi, 6G et OpenRAN par exemple. Cet inconfort ne doit pas empêcher la préparation du tissu économique à l'adoption généralisée de la 5G. Très classiquement, le déploiement privé suit le déploiement public et nous y sommes pour les usages quotidiens (JO de Paris, transport, vidéosurveillance…). **Il est essentiel de démocratiser l'accès à la 5G pour toutes les entreprises et zones géographiques (stades, usines, ports, zones rurales), tout en se préparant à la 6G, qui mobilise dores et déjà la Californie et nous invite à une réflexion de suiveur agile ou de parieur technologique**.

[172] *Digital Europe, The EU's critical tech gap: Rethinking economic security to put Europe back on the map, p. 29. Voir aussi Draghi Report.*

Recommandation 6.1 : recentrer les 735 M€ alloués à la mission 5G Industrielle dans le cadre de France 2030 autour de deux objectifs clés.

D'une part, l'accélération de la commercialisation de fonctionnalités 5G avancées – comme le *slicing* – à moindre coût. Cela nécessite de mettre en œuvre un cadre structuré de subventions publiques pour soutenir les entreprises devant adapter leurs outils de production à ces nouvelles technologies, et ainsi réduire leurs coûts d'adoption. Ces aides financières doivent cibler à la fois les grandes entreprises, qui peuvent servir de locomotives, et les ETI-PME innovantes, qui jouent un rôle clé dans les chaînes de valeur industrielles. Elles doivent être fournies en impliquant les collectivités locales en leur donnant pour instruction de privilégier la commande publique de fonctionnalités 5G avancées afin de renforcer la cohérence du dispositif. Pour cela, il serait opportun de développer des projets pilotes en 5G industrielle, en collaboration avec des industriels régionaux, pour démontrer la valeur ajoutée de la 5G dans des environnements à forte valeur ajoutée (usines, ports, agriculture de précision).

D'autre part, l'utilisation de réseaux privés comme *enablers* des usages identifiés dans le cadre du continuum souverain cloud-edge-réseaux IoT. Les réseaux privés 5G sont aujourd'hui la seule solution technologique capable d'assurer une transmission sécurisée et adaptée à la criticité des données, tout en garantissant une intégration fluide et cohérente entre les différents canaux du continuum.

Recommandation 6.2 : mettre en place des partenariats publics privés à l'échelle européenne avec les pays à la pointe en matière de connectivité réseau, tels que la Corée du Sud,

le Japon, l'Inde et Singapour pour faire émerger des offres compétitives de solution de type PaaS et les intégrer aux initiatives Open-RAN. Cela permettrait de soutenir l'émergence d'un écosystème européen capable de capter la valeur effective de la 5G et des générations suivantes et de ne pas se cantonner à la fabrication d'antennes. Cela favoriserait aussi la diffusion de solutions *open source* de qualité pour encourager le développement de sociétés européennes spécialisées dans l'*open source* communautaire.

Recommandation 6.3 : intégrer à l'agenda de la prochaine Commission européenne une consultation approfondie des industriels et des experts sur les impacts du nouveau système de redevances sur les brevets essentiels aux normes (BEN), afin d'éviter des freins à l'innovation et de la calquer sur les besoins du terrain.

3.2. LES CÂBLES : UN OUTIL DE RÉSILIENCE À MIEUX PROTÉGER

a. Une politique d'installation des câbles terrestres et aériens qui fait fi de la géographie et des coûts

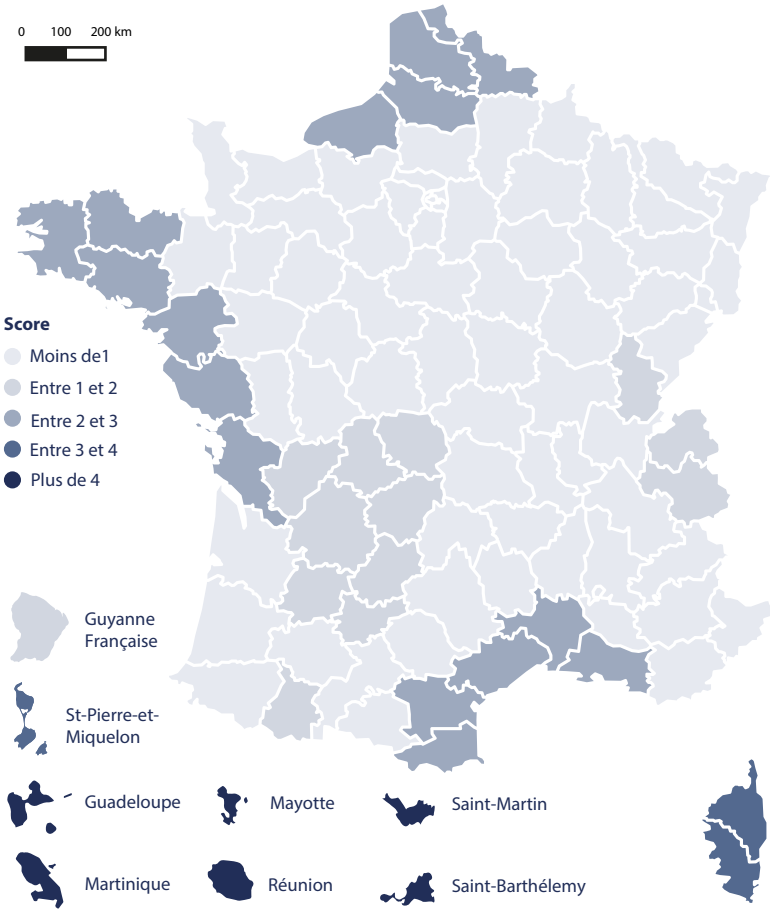
Il n'existe pas de politique publique cohérente pour enfouir les câbles terrestres, alors même que cela augmenterait la résilience et la durée de vie du réseau

Si le plan Très Haut Débit a considérablement renforcé l'accès à la connectivité dans certaines zones, l'enfouissement des réseaux filaires reste absent d'une politique publique cohérente et unifiée. Aujourd'hui, plus de 50 % des réseaux en France sont encore aériens, exposés à des risques majeurs tels que sabotages, accidents routiers, catastrophes naturelles ou aléas climatiques. La crise du Covid a révélé une double vulnérabilité : d'une part, une sous-estimation des risques de sabotages et d'attaques coordonnées, comme en témoignent les interruptions massives de connectivité provoquées cet été par des actes de sabotage ciblés[173] ; d'autre part, des infrastructures particulièrement sensibles aux aléas climatiques, illustrées par les dégâts causés par la tempête Alex dans les Alpes-Maritimes. À cela s'ajoute la transition d'un réseau historique centralisé vers un réseau morcelé, exploité par une pluralité d'acteurs locaux aux régimes divers. L'enfouissement des lignes pourrait non seulement réduire ces vulnérabilités, mais également prolonger la durée de vie de certaines infrastructures de 20 ans, contribuant à leur pérennité. **Dans ce contexte, les réseaux FttH affichent une résilience jugée seulement moyenne, en raison de la fragilité croissante des infrastructures aériennes face aux vents violents et aux incendies**.

[173] *Le Monde, 13 août 2023, « Les opérateurs de télécoms craignent les sabotages du réseau de fibre optique », https://www.lemonde.fr/.*

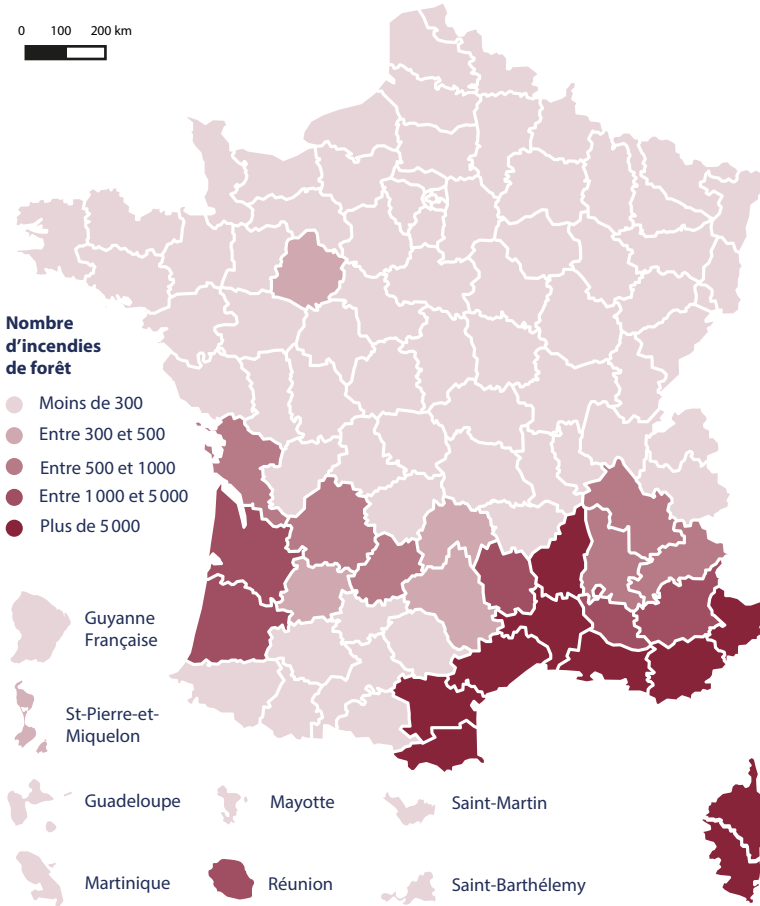
Graphique n° 34 • Cartographie des risques, zone de vent

(valeur moyenne par département)



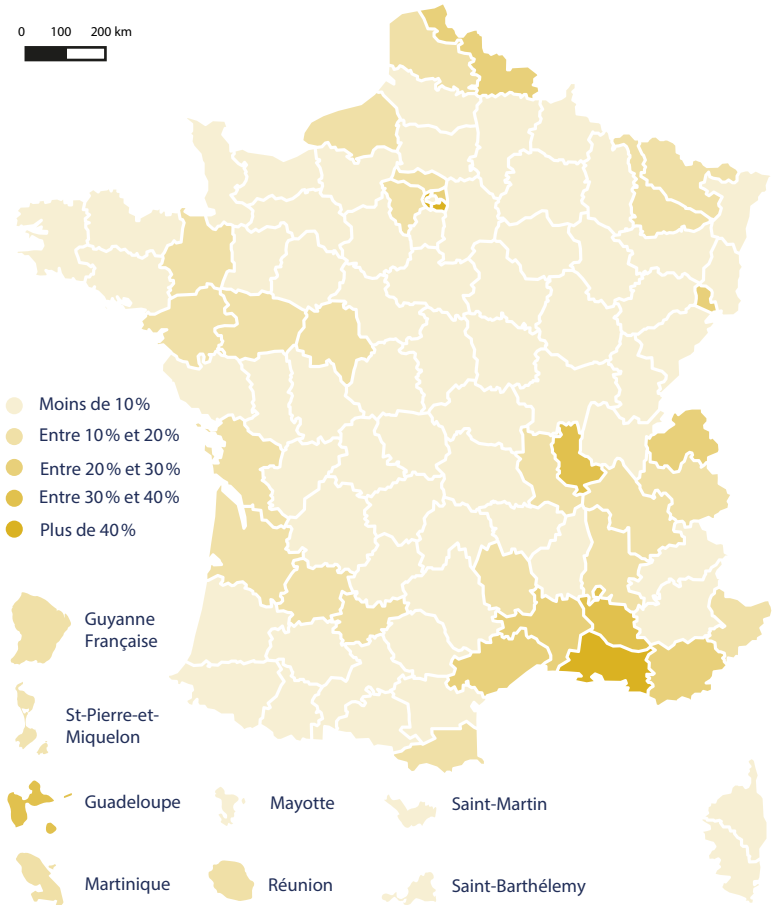
Source : InfraNum et Banque des Territoires, Résilience des Réseaux FttH, 2023.

Graphique n° 35 • Cartographie des risques, incendie de forêt



Source : InfraNum et Banque des Territoires, Résilience des Réseaux FttH, 2023.

Graphique n° 36 • Taux de communes dans les territoires à risque important d'inondation



Source : InfraNum et Banque des Territoires, Résilience des Réseaux FttH, 2023.

Des scénarios d'enfouissement qui appellent des arbitrages publics

Pour chiffrer les coûts associés à un enfouissement raisonné des câbles terrestres et aériens, InfraNum et la Banque des Territoires se sont appuyés sur un échantillon anonymisé et représentatif de 15 départements.

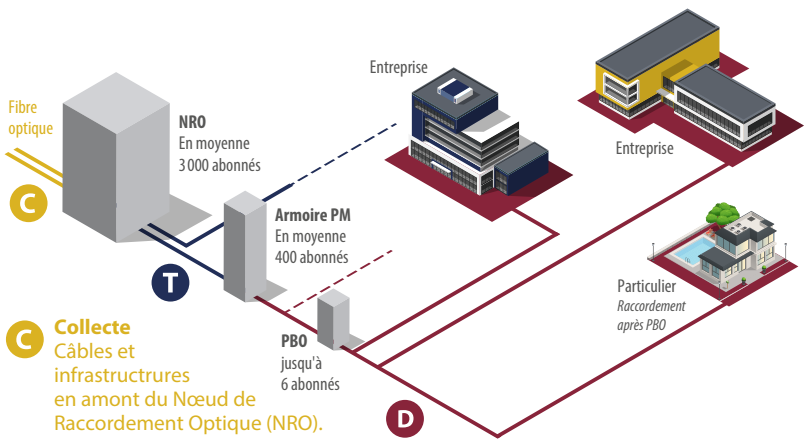
La mesure identifiée comme prioritaire est celle de l'enfouissement du réseau de collecte, qui se réfère aux câbles et infrastructures en amont du nœud de raccordement optique (NRO). Ce réseau est critique car il constitue la colonne vertébrale de la connectivité : toute coupure ou dysfonctionnement peut entraîner des interruptions massives des services sur de vastes territoires, affectant la continuité numérique des entreprises et des services publics. Les câbles des NRO collectent les flux de données provenant des réseaux de distribution locaux pour les acheminer vers les réseaux principaux ou de transport (backbone). **Le coût estimé pour cette mesure est de 630 M€[174], correspondant à l'enfouissement de 11 100 km de réseau**. Le réseau de distribution, quant à lui, désigne les infrastructures reliant les points de mutualisation (PM) aux utilisateurs finaux (entreprises, foyers, etc.). Bien qu'il soit essentiel pour assurer la connectivité à l'échelle locale, il est moins critique car une interruption affecte généralement un nombre plus restreint d'utilisateurs et est plus rapide à réparer. Les priorités d'enfouissement pour ce type de réseau dépendent des spécificités locales, comme les zones à risque de vents violents, d'incendies ou d'inondations.

Vient ensuite le réseau de transport avec des infrastructures critiques comme les nœuds de raccordement optique (NRO) et les points de mutualisation (PM) situés dans des zones vulnérables. Les NRO permettent de regrouper et de répartir les flux de données à l'échelle régionale, tandis que les PM assurent la distribution des services numériques vers des zones spécifiques. L'étude estime qu'un investissement

[174] *Source : Résilience des réseaux FTTH, Infranum, Banque des territoires 2023.*

supplémentaire de 410 M€[175] serait nécessaire pour sécuriser ces nœuds, notamment par leur protection contre les risques environnementaux. Par ailleurs, 560 M€[176] supplémentaires sont requis pour sécuriser les chambres souterraines, qui servent d'espaces de connexion et de maintenance pour les câbles optiques. Ces chambres, souvent exposées à des inondations ou à des infiltrations, sont essentielles pour maintenir l'intégrité du réseau. **Cela porte le total nécessaire à l'enfouissement et à la sécurisation complète du réseau de collecte à environ 1,7 Md €**.

Graphique n° 37 • Schéma explicatif, collecte, transport, distribution



C **Collecte**
Câbles et infrastructrures en amont du Nœud de Raccordement Optique (NRO).

T **Transport**
Câbles et infrastructrures engtre le Nœud de Raccordement Optique et le Point de Mutualisation (NRO/PM).

D **Distribution**
Câbles et infrastructrures partant du PM jusqu'au Point de Raccordement de Branchement Optique (PBO) inclus.

Source : https://thd42exploitation.fr/comment-avoir-la-fibre-optique-chez-moi.

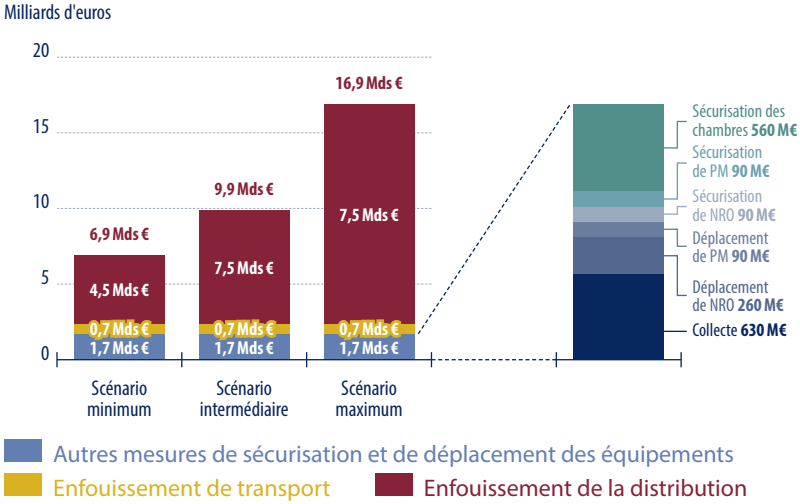
[175] *Ibid.*

[176] *Ibid.*

Les scénarios prennent ensuite en compte la puissance, et a fortiori la criticité des câbles concernés ; avec un focus sur les câbles au format 14 FO pour la distribution inter-bourg, qui sont des câbles optiques à 14 fibres permettant de relier des zones rurales ou périurbaines aux infrastructures principales, et qui sont les plus critiques car ils assurent la continuité de service dans des territoires où les solutions alternatives en cas de rupture (comme le mobile ou les réseaux secondaires) sont souvent inexistantes ou insuffisantes. Cela donne un scénario minimum d'enfouissement à 6,9 Mds € et un scénario maximum à 16,9 Mds €, **le strict minimum étant estimé à 1,7 Md € pour la sécurisation des chambres, des PM, des NRO (ainsi que leur déplacement) et le réseau de collecte. Par ailleurs, les besoins en main-d'œuvre sont chiffrés à 5 Mds €** par l'étude[177] car l'enfouissement nécessite des équipes qualifiées pour les travaux de génie civil, la pose des conduits et câbles, ainsi que la sécurisation des sites, dans un contexte où le marché de l'emploi dans ces secteurs est déjà sous tension.

[177] *Source : Résilience des réseaux FTTH, Infranum, Banque des territoires 2023.*

Graphique n° 38 • Les trois scénarios pour sécuriser les réseaux FTTH en France



Source: InfraNum et Banque des Territoires, Résilience des Réseaux FttH, 2023.

Des articulations nécessaires avec le plan national d'adaptation au changement climatique (PNACC)

Une politique d'enfouissement raisonnée des câbles terrestres doit être adaptée aux caractéristiques de chaque zone, car un enfouissement systématique est coûteux, n'est pas toujours approprié et ce pour une garantie de sécurité non nécessairement prouvée. Dans les régions fréquemment touchées par des inondations, les câbles souterrains sont vulnérables aux dommages causés par l'eau, ce qui peut entraîner des courts-circuits, la corrosion des matériaux et des difficultés d'accès pour les réparations. Par exemple, les inondations dans le Pas-de-Calais ou celles causées par la tempête Xynthia en 2010 ont mis en évidence les défis importants liés à l'entretien et à la réparation des infrastructures souterraines dans de telles conditions. De plus, les sols

sujets à des affaissements, des glissements de terrain ou à des phénomènes de retrait-gonflement d'argile (RGA), qui désignent l'expansion et la contraction des sols argileux en fonction de l'humidité, peuvent compromettre l'intégrité des câbles enfouis. Ces mouvements de terrain peuvent exercer des pressions irrégulières sur les infrastructures souterraines, entraînant des fissures ou des ruptures de câbles et nécessitant des réparations complexes et coûteuses.

Le plan national d'adaptation au changement climatique (PNACC[178]) est un bon levier pour poursuivre la politique d'enfouissement car il permet de l'inscrire dans une dynamique territoriale plus unifiée. Dans le cadre de la consultation publique sur le Plan National d'Adaptation au Changement Climatique, 14 mesures principales et 51 complémentaires ont été proposées pour préparer le territoire national à un scénario de +4°C. Parmi celles-ci, la mesure 32, portant sur la résilience des services de communication électronique, appelle à une stratégie de long terme pour assurer la continuité et la qualité du service. Toutefois, cette mesure ne peut être pleinement efficace qu'en synergie avec la mesure 7, dédiée aux risques d'incendies de forêt, et la mesure 19, qui intègre les enjeux climatiques dans la prévention des risques technologiques. De même, la mesure 24, sur l'intégration des enjeux climatiques dans les normes techniques, doit être appliquée avec pragmatisme pour ne pas alourdir les opérations en cours.

Bien que cette consultation soit un bon point de départ, elle présente trois limites majeures qu'il convient de pallier :

1. **Des choix budgétaires flous et une temporalité inadéquate :** les coûts des mesures ne sont pas priorisés en fonction de leur impact stratégique, et le début des opérations prévu pour 2026 est trop tardif pour répondre à l'urgence climatique.

[178] *A. Pannier-Runacher, ministre de la Transition écologique, de l'Énergie, du Climat et de la Prévention des risques, octobre 2024, « présentation du plan national d'adaptation au changement climatique », https://www.ecologie.gouv.fr/.*

2. **Une focalisation excessive sur les risques hivernaux :** le PNACC met l'accent sur les tempêtes hivernales, négligeant les principaux risques identifiés par InfraNum et la Banque des Territoires, tels que les vents violents, les incendies inter-bourg et les inondations. Ces menaces, particulièrement critiques en été, nécessitent une approche plus équilibrée.
3. **Une absence d'intégration technologique proactive :** aucune mesure ne prévoit explicitement de tirer parti des technologies avancées pour prévenir les risques climatiques, alors que le projet pilote de la DGE pourrait être un levier pour installer des capteurs intelligents lors des travaux d'enfouissement des points névralgiques du réseau.

Recommandation 7

Sécuriser les nœuds critiques de distribution des câbles terrestres par une politique d'enfouissement raisonnée des câbles terrestres et aériens.

Les câbles sont devenus des équipements vitaux et vulnérables, de plus en plus soumis aux menaces de sabotages et aux aléas climatiques. Il est donc nécessaire de les sécuriser afin de sécuriser le réseau. Le coût estimé d'une telle sécurisation se situe entre 6 à 17 Mds € ce qui représente un montant difficilement accessible compte tenu de nos marges de manœuvre budgétaires. Le choix est donc de proposer une sécurisation prioritaire des réseaux de collecte de données les plus critiques dits NRO.

Pour y parvenir, intégrer dans le plan national d'adaptation au changement climatique (PNACC) une politique d'enfouissement raisonnée à partir d'une identification précise des réseaux de collecte des données et des risques sur les nœuds critiques de distribution.

Recommandation 7.1. : concentrer les efforts financiers sur le réseau de collecte des données (NRO) pour enfouir les points névralgiques des réseaux de communications électroniques. Face aux contraintes budgétaires, il serait inefficace de tenter simultanément l'enfouissement du réseau de distribution, qui dessert un nombre plus restreint d'utilisateurs et est plus rapide à réparer. En priorisant le réseau de collecte, notamment dans les zones vulnérables identifiées (forêts, zones à risque d'inondations), il est possible d'obtenir un impact immédiat et significatif.

Recommandation 7.2 : inclure les menaces liées aux vents violents et aux incendies dans la hiérarchie des risques climatiques identifiée dans le PNACC (en mesure 32). Une approche annuelle est indispensable, avec des expérimentations dès les périodes estivales, où les infrastructures numériques sont exposées aux incendies et à la chaleur extrême. Aligner cette mesure avec la mesure 7 du PNACC dédiée aux incendies permettrait une gestion des risques plus cohérente.

Recommandation 7.3 : intégrer dans le PNACC l'utilisation des technologies avancées de surveillance pour anticiper les nouveaux risques climatiques (mesure 19). Ce type d'outils est susceptible de contribuer à l'adoption du edge computing sur le territoire français (conformément à la recommandation 1).

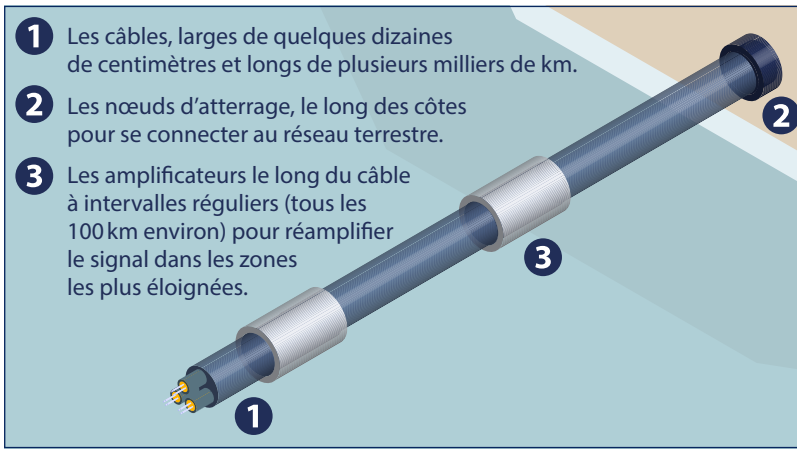
Veiller à associer les collectivités territoriales et autorités déconcentrées compétentes, ainsi que les parties prenantes du génie civil pour garantir la faisabilité de ces mesures, et à bien coordonner ces mesures avec les travaux publics.

b. Des enjeux majeurs de sécurité et de standard sur les câbles sous-marins

Les câbles sous-marins jouent un rôle structurant pour la connectivité ou la qualité de service et la sécurité des infrastructures numériques françaises, assurant le transport de 97 % des données mondiales. Actuellement, la majorité de ces câbles est contrôlée par de grands acteurs technologiques américains (GAMMA), et près de 70 % du trafic est généré par des fournisseurs de contenu tels que Google, Meta et Netflix. En effet, la présence d'entreprises sur plusieurs continents, mais également le fort déploiement des services *cloud* ont accentué les besoins d'interconnexion des entreprises, et avec eux l'implication progressive des *hyperscalers* dans l'acquisition de câbles sous-marins. Google a par exemple investi dans 13 chemins de câbles sous-marins Internet pour un total de 100 000 kilomètres en 2020[179].

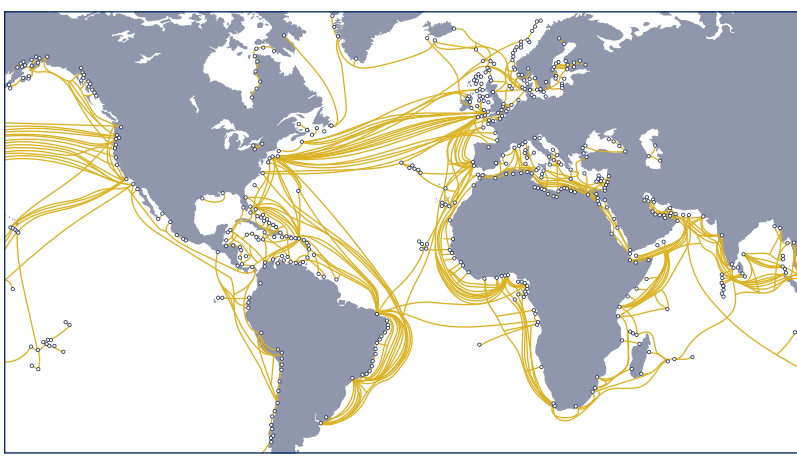
[179] *Zms cable, « L'ambition de Google de connecter le monde : la disposition des câbles Internet sous-marins ».*

Graphique n° 39 • De quoi est composé un câble sous-marin ?



Source : https://www.submarinecablemap.com/.

Graphique n° 40 • Carte mondiale des câbles sous-marins



Source : https://www.submarinecablemap.com/.

Des entreprises françaises de câbles sous-marins compétitives

La France a de nombreux atouts dans le domaine des câbles sous-marins, tant en ce qui concerne ses acteurs économiques que son environnement normatif. Sur le plan économique, la France conserve une position stratégique grâce à des acteurs comme Orange et Alcatel Submarine Networks (ASN), qui possèdent et exploitent une part significative des câbles transitant par le pays. Sur le plan normatif, la directive nationale du 13 novembre 2020 sur l'attractivité des câbles sous-marins a permis de réduire les délais administratifs de manière significative, facilitant l'implantation de projets de câbles sous-marins sur le territoire et contribuant à la compétitivité de la France dans ce secteur essentiel. Cette réglementation promeut un environnement favorable à l'expansion et à la modernisation des infrastructures. En effet, elle impose des études d'impact plus précises et encourage des tracés limitant les perturbations sur les écosystèmes marins. Par ailleurs, en sécurisant juridiquement les délais et les coûts de déploiement, elle réduit les incertitudes pour les investisseurs, rendant la France plus compétitive pour les consortiums internationaux.

Les entreprises françaises sont à la pointe en matière d'innovation technologique dans le secteur des câbles sous-marins. **Elles déploient notamment des technologies de *fiber sensing* qui permettent de détecter et de surveiller en temps réel des variations physiques telles que les vibrations, la température et la déformation des câbles**. Ces capacités offrent une prévention précoce des incidents potentiels et optimisent la gestion des infrastructures. Orange utilise par exemple le *fiber sensing* pour améliorer la surveillance de ses infrastructures de câbles sous-marins. **En intégrant cette technologie, l'entreprise peut détecter les moindres vibrations et changements de température le long de ses câbles, permettant ainsi de prévenir des dommages causés par des activités maritimes non autorisées, des mouvements sismiques ou d'autres perturbations**

environnementales. Cette capacité de détection précoce permet à Orange de réagir rapidement et de protéger l'intégrité de ses infrastructures critiques, renforçant la continuité et la sécurité des communications sous-marines. **Des entreprises comme Nexans ont aussi mis en place des plateformes avancées de surveillance qui consistent à recueillir, analyser et interpréter les données issues de capteurs installés le long des câbles sous-marins**. Cela permet d'identifier les anomalies rapidement et de coordonner des interventions préventives. Leur plateforme de données polyvalente sur le *cloud*[180] centralise la surveillance des données critiques, améliorant ainsi la réactivité face aux risques et renforçant la résilience des réseaux de câbles sous-marins.

La nécessité de renforcer la sécurité physique des câbles qui transitent par la France

Le rachat récent d'Alcatel Submarine Networks (ASN) par l'État français souligne l'importance stratégique des câbles sous-marins pour la souveraineté nationale et la résilience des infrastructures numériques. Ce rachat permet à la France de maintenir un contrôle direct sur une entreprise clé, essentielle à la conception, au déploiement et à l'entretien de ces infrastructures critiques. ASN, qui emploie 2 000 personnes dont 70 % en France, bénéficie désormais de moyens renforcés pour innover et répondre aux priorités stratégiques nationales, notamment dans un contexte de menaces géopolitiques croissantes. La flexibilité offerte par la participation résiduelle de Nokia (20 %) garantit également une possibilité de nationalisation complète si les circonstances l'exigent.

Les câbles sous-marins, essentiels à la connectivité mondiale, sont de plus en plus ciblés par des actes de sabotage et des tensions géopolitiques. Les récents incidents en mer Baltique, où deux câbles

[180] *Nexan perspectives, juin 2024, « La surveillance avancée au service de la fluidité du réseau de câbles sous-marins », https://www.nexans.com/fr/perspective/la-surveillance-avancee-au-service-de-la-fluidite-du-reseau-de-cables-sous-marins/.*

critiques (BSC et C-LION1 en novembre 2024) reliant plusieurs pays européens ont été endommagés, illustrent cette vulnérabilité. **Bien que les perturbations aient été mineures, les soupçons de sabotage par un navire sous pavillon chinois mettent en lumière trois enjeux majeurs :**

1. **La vulnérabilité physique des câbles face aux manipulations intentionnelles**.
2. **Les limites du droit maritime international, qui entrave les interventions rapides sans l'accord de l'État du pavillon du navire suspect**.
3. **L'importance géopolitique des câbles sous-marins dans un contexte de rivalités accrues entre grandes puissances**.

Dans ce contexte, deux priorités stratégiques se dégagent pour renforcer la sécurité des câbles sous-marins.

La première priorité consiste à investir dans des plateformes de surveillance intégrées capables de consolider des données provenant de différentes sources pour améliorer la détection des anomalies et la maintenance proactive. Aujourd'hui, l'écosystème de surveillance des câbles sous-marins est fragmenté, chaque fournisseur utilisant des logiciels propriétaires, ce qui complique la coordination, malgré les efforts quotidien de sécurisation et de détection des menaces entrepris par la Marine Nationale. **De plus, environ 20 % des incidents en haute mer restent inexpliqués[181], un chiffre préoccupant alors que les menaces provenant de groupes sponsorisés par des États, comme les *Advanced Persistent Threats* (APT)[182], se multiplient**.

[181] *Jizô AI, janvier 2024, « la France et son réseau de câbles sous-marins : enjeux et risques des infrastructures vitales du cyberespace », https://sesame-it.com/fr-fr/blog_articles/france-reseau-cables-sous-marins-enjeux-cyber.*

[182] *Dans le contexte présent, il s'agit de cyberattaques sophistiquées et prolongées menées par des groupes étatiques ou criminels visant à espionner, intercepter ou perturber les flux de données critiques transitant par ces infrastructures stratégiques, souvent via des compromissions physiques ou logicielles.*

Ces attaques ciblent particulièrement les câbles longue distance, dont la maintenance est entravée par leur éloignement et des conditions maritimes difficiles. En investissant dans des outils de surveillance centralisés et des capteurs sous-marins avancés, les capacités de réponse aux menaces pourraient être considérablement renforcées, tout en permettant une gestion plus efficace des infrastructures.

La deuxième priorité concerne le renforcement de la résilience des réseaux dans les territoires ultramarins, en garantissant systématiquement la présence d'au moins trois câbles sous-marins par territoire, notamment dans les DROM. Ces investissements, cruciaux pour la connectivité des territoires, ne doivent pas être guidés uniquement par la densité de population, mais par des impératifs stratégiques visant à sécuriser des infrastructures essentielles à l'économie, à la société et à la souveraineté numérique. Un soutien financier public est indispensable pour encourager ces déploiements, assurer la redondance des réseaux et limiter les risques de coupures prolongées qui pourraient isoler ces territoires stratégiques.

Deux pistes de solutions gagneraient à être explorées dans ce contexte.

- Une première solution consiste à renforcer la protection physique des infrastructures enterrées et des installations associées. **Cela inclut la sécurisation des chambres de tirage à l'aide de systèmes de verrouillage robustes, de capteurs de mouvement pour détecter toute intrusion, et de dispositifs de surveillance par drones pour un suivi en temps réel**. Ces mesures visent à protéger les points les plus vulnérables des câbles sous-marins, là où ils émergent ou se connectent au réseau terrestre.

- **L'établissement de back-ups spatiaux, en complément, pour pallier les interruptions des câbles sous-marins, notamment dans les territoires ultramarins**. En cas de sabotage ou de défaillance des réseaux terrestres, des satellites géostationnaires, comme

ceux d'Eutelsat, pourraient prendre le relais pour assurer la connectivité. Cependant, cette option présente des limitations : la capacité de desserte reste restreinte, car un nombre limité de satellites peut fonctionner simultanément, et les coûts sont significativement plus élevés (2 à 3 fois plus élevés que la fibre). De plus, la latence est supérieure à celle des systèmes comme Starlink, car les satellites géostationnaires se trouvent à une distance 50 fois plus grande.

Une présence française dans les instances de standardisation laissée aux industriels avec un soutien étatique insuffisant

La présence française dans les instances de normalisation des câbles sous-marins tient essentiellement aux acteurs industriels, mais elle reste délaissée par les acteurs étatiques. Bien que des organes comme l'ARCEP, la DGE et l'ANSSI soient ponctuellement présents dans les instances de normalisation internationales, les efforts de normalisation sont essentiellement supportés par les entreprises françaises de câbles sous-marins car ces dernières disposent des ressources techniques et des intérêts économiques directs pour défendre leurs positions.

Tableau n° 5 • Panorama de la présence française dans les instances de normalisation internationales de câbles sous-marins

Instance	Rôle	Présence française
Union Internationale des Télécommunications (UIT)	Normalisation des télécommunications, gestion du spectre radioélectrique et des orbites satellites.	Membre actif, élu au Conseil, représenté par l'ARCEP et des opérateurs comme Orange, qui participent aux travaux techniques. La France est le 4ème contributeur budgétaire.

Instance	Rôle	Présence française
International Cable Protection Committee (ICPC)	Promotion des meilleures pratiques pour la protection des câbles sous-marins.	Membres industriels français, dont Orange Marine, Total Energie One Tech, Axiom, Shom, Sopartech, ONATI, fortement impliqués dans les discussions techniques.
Groupes de travail de l'OCDE	Études sur les infrastructures numériques et leurs implications économiques et sécuritaires.	Participation par des représentants du gouvernement (Direction générale des entreprises) et des experts privés.
Commission CF2 de la Commission européenne	Renforcement de la sécurité et de la résilience des infrastructures critiques en Europe.	Participation active via l'ANSSI et des contributions d'opérateurs nationaux comme Orange et Bouygues Télécom.

Pourtant, il est fondamental de renforcer la collaboration des parties prenantes dans la gestion et la maintenance des câbles sous-marins car la majorité des projets sont soutenus et détenus par des consortiums internationaux[183]. Cela passe par une stratégie de normalisation plus pragmatique centrée sur la prévention de l'obsolescence des infrastructures et leur conformité aux évolutions technologiques. Des pays comme le Portugal[184] ont adopté des bonnes pratiques en la matière en mettant en place des mesures de protection proactives de leurs câbles sous-marins par le biais d'une collaboration étroite avec ces consortiums internationaux. Le pays a ainsi contribué à l'élaboration de normes techniques et de protocoles de sécurité pour ces infrastructures critiques. **Cette participation active a permis au Portugal de s'assurer que les projets de câbles sous-marins, tels que le projet de câble sous-marin PISCES**[185]**, respectent des standards élevés en matière de performance et de résilience**. De plus, en alignant ses réglementations nationales sur les meilleures pratiques internationales, le Portugal a facilité l'intégration de ses infrastructures

[183] *Ejercitos, Geopolítica de los cables submarinos de comunicaciones.*

[184] *Euronews, septembre 2024, « Guerres hybrides : la vulnérabilité des câbles sous-marins inquiète les experts ».*

[185] *Commission européenne, 3 décembre 2024,« Étude pour le système de câbles sous-marins reliant le Portugal, l'Irlande et l'Espagne (système de câbles PISCES) ».*

dans le réseau mondial, renforçant ainsi sa connectivité et sa sécurité numérique.

Recommandation 8

Valoriser l'atout stratégique que représentent les câbles sous-marins français *via* une stratégie intégrée combinant surveillance renforcée, investissements ciblés en Outre-mer et influence accrue dans les instances internationales.

La résilience des câbles sous-marins français est menacée par des actes de sabotage terroristes et militaires de plus en plus nombreux et sophistiqués. La France a un avantage compétitif majeur dans le secteur grâce à ses entreprises (ASN, Nexans et Orange Submarine Cable) qui fabriquent des câbles haute performance et nous assurent une maîtrise de bout en bout de notre infrastructure sous-marine (conception, fabrication, pose et maintenance). Pour maintenir cet avantage, dans le contexte géopolitique tendu que nous connaissons, deux priorités :

Recommandation 8.1 : investir, *via* France 2030, pour renforcer la résilience des câbles sous-marins en se concentrant sur les deux éléments ci-dessous.

1. **Les technologies de surveillance et de gestion innovantes, comme le *fiber sensing*, qui détecte les variations de tension et de température pour une maintenance préventive efficace**. Compléter ces solutions par des systèmes d'IA capables de différencier anomalies et menaces réelles, afin de détecter de manière proactive les risques de rupture ou de sabotage et d'optimiser les interventions sous-marines.

2. **Passer commande à des sociétés de satellites géostationnaires européennes, de type Eutelsat, pour sécuriser à moindre coût les câbles situés dans les territoires ultra-marins, afin de garantir redondance et fiabilité du système dans son ensemble**. Ces « back-ups spatiaux » auraient pour objectif de pallier les interruptions des câbles sous-marins, en cas de sabotage ou de défaillance des réseaux terrestres, pour renforcer leur sécurité physique.

Recommandation 8.2 : renforcer la participation des autorités publiques et des industriels français au sein des organismes stratégiques tels que l'UIT (Union Internationale des Télécommunications), l'ICPC *(International Cable Protection Committee)*, les groupes de travail de l'OCDE et la commission CF2 de la Commission européenne. À l'instar du modèle portugais, tirer parti de cette présence pour développer des partenariats avec les consortiums internationaux gestionnaires de câbles pour positionner la France comme un acteur central et incontournable dans l'installation, la gestion et la sécurisation des câbles sous-marins.

3.3. LES SATELLITES : UN POTENTIEL QUI DEMEURE INEXPLOITÉ FAUTE DE VISION ÉCONOMIQUE ET STRATÉGIQUE

Les satellites contribuent au système d'infrastructures numériques en relais des antennes et des câbles dans les zones dites blanches ou dans les zones aériennes. Ils représentent à la fois une composante indispensable d'une approche sécurisée et optimisée et un potentiel économique majeur insuffisamment exploité au moment même où la demande en connectivité augmente fortement.

Les communications par satellite jouent un rôle essentiel en connectant les populations vivant dans des zones blanches à Internet, en facilitant les télécommunications à l'échelle mondiale, et en fournissant des services critiques en cas de catastrophe naturelle ou d'urgence. Ces satellites forment un réseau interconnecté, communiquant entre eux avant de transmettre les données vers la Terre. Deux types d'instruments réceptionnent ces données spatiales : les téléports, qui sont des stations terrestres équipées d'antennes de grande taille et situées à des emplacements stratégiques pour assurer une liaison stable avec les satellites, et les antennes individuelles, comme celles utilisées par Starlink, qui permettent une connexion directe et locale pour les utilisateurs finaux, même dans des zones isolées. La transmission des données vers leur destination finale s'appuie sur les câbles terrestres et sous-marins, qui assurent une connectivité globale en transportant rapidement les flux d'information entre les continents et les infrastructures locales.

Les fonctionnalités associées sont désormais accessibles sur les appareils mobiles, les derniers modèles de smartphones fournissant des accès aux réseaux satellitaires pour l'envoi de SOS d'urgence, l'assistance routière dans une zone sans couverture réseau wifi/mobile ou le partage de position avec ses proches. SpaceX, par exemple, avec ses 6 000 en orbite basse (LEO) et un objectif de 12 000 d'ici 2027 *via* le service Starlink, offre une latence de 20 ms et un débit jusqu'à 220 Mbps pour un abonnement de 40 €/mois. Les acteurs du *New Space*, comme Amazon avec le projet Kuiper, et la collaboration Eutelsat-OneWeb[186], proposent aussi des constellations de masse à bas coûts. Amazon prévoit 3 000 satellites d'ici 2029, et Eutelsat-OneWeb vise une couverture mondiale. Les partenariats se multiplient, comme celui de Google et Starlink, où il est prévu que des datacenters Google Cloud hébergent des stations Starlink au sol, améliorant la connectivité dans les zones

[186] *France Inter, AFP, V. Vasseur, 25 juillet 2022, « Naissance d'un géant de l'internet depuis l'espace : le mariage entre Eutelsat et OneWeb en cinq questions ».*

difficiles. La Chine avance également avec China SatNet, déployant 369 satellites[187] en orbite basse.

Le secteur spatial reste étroitement lié à des enjeux de souveraineté, avec environ 80 % de son financement provenant de fonds publics, principalement *via* des commandes publiques stratégiques. Cette dynamique souligne la nécessité d'une stratégie européenne claire pour éviter des dépendances critiques.

La forte croissance de la demande de connectivité à l'échelle mondiale a profondément impacté le secteur des télécommunications spatiales. Cette croissance a fait émerger le mouvement du *New Space*[188], qui désigne un écosystème spatial fondé sur une approche plus agile, compétitive et commerciale que le secteur spatial traditionnel. Cette approche repose sur l'intervention accrue d'acteurs privés et la réduction des barrières à l'entrée. Elle trouve son origine dans la révolution des usages numériques terrestres, notamment l'internet des objets (smartphones, véhicules connectés, etc.) mais aussi dans la passion personnelle des entrepreneurs américains ayant fait fortune dans le secteur de la technologie, et qui la considèrent comme la dernière « brique » de l'infrastructure numérique. **Pour répondre à cette demande croissante et conquérir des parts de marché, de nouveaux acteurs se sont imposés en déployant des constellations de satellites, qui sont des ensembles coordonnés de satellites identiques visant une couverture quasi-complète de la planète**. La baisse des coûts d'accès à l'espace a joué un rôle crucial dans cette expansion. SpaceX, par exemple, a révolutionné le secteur avec des innovations de rupture, dont la plus marquante est la réutilisation des lanceurs. C'est dans cette perspective qu'Elon Musk a lancé le projet Starlink, visant à déployer une infrastructure de connectivité mondiale basée

[187] *Statista, janvier 2025, « Nombre de satellites en orbite dans le monde au 1er mai 2023, par pays opérateur ».*

[188] *Martin, G. (2015). NewSpace: The Emerging Commercial Space Industry. NASA Ames Research Center, https://www.earthdata.nasa.gov/s3fs-public/2023-11/newspace_nasa.pdf.*

sur une constellation de satellites en orbite basse, pour un coût estimé à 30 Mds $ sur dix ans[189], sa rentabilité n'étant à ce stade pas connue. L'objectif d'Elon Musk est d'intégrer la technologie dans la vie quotidienne : par le biais de The Boring Company pour les infrastructures, Tesla pour la mobilité, et Neuralink pour l'interface homme-machine. **Ce projet dépasse le cadre du déploiement technologique : il incarne une vision en matière d'infrastructures numériques mondiales, résilientes et interconnectées, capables de répondre aux grands défis sociétaux tels que la pauvreté, la famine, les maladies, la gestion des ressources limitées et la congestion urbaine**.

Les constellations de satellites reposent aujourd'hui sur la combinaison de nouvelles approches industrielles, l'exploitation et la valorisation des données en aval, ainsi que sur les commandes publiques. Cependant, leur modèle économique reste flou et leur rentabilité incertaine. Par exemple, la US Space Force a mis en place une approche de satellites à architecture hybride, consistant à combiner des constellations publiques et privées pour maximiser la résilience et l'efficacité des infrastructures spatiales. Un autre exemple concerne le programme Starshield de SpaceX, qui étend les capacités de Starlink en offrant des services de connectivité sécurisée spécifiquement dédiés aux gouvernements et aux applications militaires, notamment pour le traitement de données sensibles et le chiffrement renforcé des communications. Cette stratégie repose sur l'intégration de satellites commerciaux dans leurs systèmes de défense et de communication, réduisant ainsi les coûts en externalisant une partie des capacités à des acteurs privés. En contrepartie, des contrats pluriannuels garantissent un flux de revenus pour les opérateurs commerciaux, tout en assurant à l'État un contrôle stratégique sur l'allocation des fréquences et l'utilisation des données critiques. **Néanmoins, cette stratégie est incertaine car elle repose sur une dépendance accrue vis-à-vis des acteurs privés, ce qui peut poser des problèmes en cas de défaillance ou de conflit**

[189] *Reuters, juin 2021, "Musk says may need $30 bln to keep Starlink in orbit".*

d'intérêts. De plus, elle est difficilement réplicable en Europe en raison de la fragmentation du marché spatial européen et de l'absence d'une stratégie unifiée entre les États membres.

En Europe, le marché des satellites à usage B2C reste sous-développé, avec une présence limitée face à des acteurs comme Starlink, tandis qu'Iris² et OneWeb se concentrent sur les segments B2G et B2B. Malgré le développement précoce de OneWeb, l'Europe n'a pas pleinement investi le potentiel du B2C, à l'inverse de SpaceX. Les échecs des projets similaires dans les années 1990, menés par Microsoft ou Motorola, avaient conduit à considérer ce modèle comme non viable, freinant l'ambition européenne. Aujourd'hui, des initiatives comme l'accord entre Starlink et l'Italie illustrent la prise d'avance des acteurs non européens sur ce segment. Si ce choix peut se justifier économiquement, il soulève des questions stratégiques sur la souveraineté européenne et l'importance de préserver un contrôle sur des infrastructures critiques. **Les cas d'usage actuels pour la connectivité spatiale illustrent clairement son importance. Elle constitue une solution pour les 20 % de foyers non couverts par le plan fibre, où des offres comme Starlink, à environ 50 euros par mois, peuvent être plus avantageuses pour les derniers raccordements. De plus, la connectivité satellitaire est essentielle pour les véhicules connectés, les trains et les avions, notamment lorsqu'ils quittent les zones couvertes par la 5G ou le Wi-Fi**.

Pour les acteurs européens, la mise en place de constellations satellitaires vise principalement à répondre à des impératifs stratégiques de souveraineté et de sécurité, plutôt qu'à exploiter pleinement le potentiel commercial du New Space. Les priorités actuelles sont de garantir des moyens de communication sécurisés pour les informations sensibles et d'assurer un accès haut débit dans les zones blanches encore présentes sur le territoire. C'est dans cette optique que le projet de constellation souveraine Iris², doté de 12 Mds € [190] et initié

[190] *Le Monde, P. Jacqué (Bruxelles, bureau européen), mai 2024, « Europe spatiale : bataille de gros sous entre la Commission et les entreprises pour la future constellation Iris² ».*

en 2022, a été développé en partenariat avec l'ESA et des industriels européens.

Iris² rencontre toutefois des difficultés de mise en œuvre en raison des rivalités entre États membres et des divergences sur le financement et la répartition des compétences industrielles. Si le contrat de concession d'Iris² a été attribué au consortium SpaceRISE[191], la question des arbitrages autour du modèle économique de la politique spatiale commerciale reste entière. Pour l'instant, ce modèle économique spatial européen repose essentiellement sur de l'exploitation commerciale. **Il n'est toutefois pas viable dans un contexte où des entreprises comme Starlink ou Kuiper disposent de leviers financiers et industriels sans précédent pour passer à l'échelle**. À titre d'exemple, Starlink vise le déploiement de 42 000 satellites en orbite basse (LEO) d'ici à 2027 contre 290 d'ici à 2030 pour Iris²[192]. Les constellations satellites américaines bénéficient de fait de commandes massives du Pentagone, réparties entre SpaceX, ULA et Blue Origin qui, à elles seules, représentent parfois le double de budgets comme celui d'Ariane 6. Par ailleurs, les projets européens sont doublement pénalisés par le fait qu'ils ne peuvent pas réutiliser leurs lanceurs comme l'a récemment fait Starlink avec Super Heavy[193] et Falcon 9. Cette réutilisation est impossible, car elle mettrait en difficulté toute la chaîne de sous-traitants européens, qui ne pourrait plus compter sur un volume de commandes suffisant pour assurer leur rentabilité.

Dans ce contexte, le projet Ariane 6 a été lancé pour permettre un accès européen autonome à l'espace. Bien que non réutilisable, ce lanceur offre des atouts stratégiques, notamment une capacité d'emport supérieure à celle de Falcon 9 et une précision accrue

[191] *Le Monde, 2 novembre 2024, « Face à Starlink d'Elon Musk, l'Europe aura sa constellation Iris² ».*

[192] *Ibid.*

[193] *Le Figaro, octobre 2024, « Starship : SpaceX récupère le premier étage de sa mégafusée, une première historique ».*

grâce à son moteur Vinci. Ces caractéristiques en font un outil clé pour soutenir le déploiement de constellations satellitaires et attirer des contrats commerciaux comme celui signé avec Amazon pour le projet Kuiper. Cependant, la dépendance de l'Europe à un modèle fondé sur des volumes limités et des coûts de lancement encore élevés freine sa compétitivité face à des acteurs comme SpaceX.

Encadré n° 21 • Focus sur le programme Ariane 6

Ariane 6 a deux objectifs principaux, en soutien d'une infrastructure numérique spatiale européenne : garantir un accès à l'espace et une compétitivité pour les acteurs européens grâce à des lanceurs moins coûteux et réutilisables.

La réussite du lancement test d'Ariane 6 acte le recouvrement européen de son accès autonome à l'espace, momentanément interrompu depuis 2022, sous l'effet conjugué du dernier vol d'Ariane 5, de l'échec du premier vol commercial de Vega-C, et de l'interruption des missions sous le système russe de Soyouz.

Le lanceur européen n'est certes pas réutilisable, mais il admet plusieurs composantes qui en font un système parfaitement adapté au marché des constellations de satellites. Il est d'abord particulièrement fiable, un marqueur du programme Ariane, et extrêmement précis. Son moteur Vinci lui permet en effet de déposer des satellites sur des orbites spécifiques, ce qui est loin d'être le cas de tous les lanceurs. Également, Ariane 6 propose une capacité d'emport plus importante que Falcon 9, avec 10 tonnes en orbite basse pour la version à deux boosters (Ariane 62), et près de 20 tonnes pour celle composée de quatre boosters (Ariane 64).

En démontrant la capacité à réduire les coûts de lancement grâce à des innovations technologiques, notamment la réutilisabilité des lanceurs, Ariane 6 répond aux enjeux financiers contemporains pour rivaliser avec des concurrents internationaux tels que SpaceX. La diminution des coûts, l'augmentation des capacités d'emport et la plus grande polyvalence dans les mises en orbite ouvrent la voie à une plus grande fréquence de missions, cruciales pour le déploiement de satellites de télécommunication. Ariane 6 a d'ores et déjà vu son carnet de commande se remplir, et a signé un contrat historique en 2022 avec Amazon pour le déploiement de sa constellation Kuiper. Il prévoit 18 lancements sur trois ans pour le compte de l'entreprise américaine, et témoigne de la pertinence commerciale du lanceur.

Pour rester pertinent sur le marché des constellations, l'Europe doit identifier des leviers de différenciation. Le projet *Advanced Space Cloud for European Net zero emission and Data sovereignty*[194] (ASCEND), coordonné par Thales Alenia Space et financé par la Commission européenne, en est un exemple prometteur. **Ce projet explore la faisabilité de data centers spatiaux, exploitant l'énergie solaire disponible en orbite et les basses températures naturelles pour le refroidissement**. Ces infrastructures pourraient réduire considérablement l'empreinte carbone des *data centers* terrestres, tout en améliorant la latence des communications, un atout crucial pour des applications sensibles comme l'intelligence artificielle et l'IoT. Au-delà de leur dimension écologique, ces *data centers* orbitaux offriraient une résilience accrue face aux catastrophes naturelles et aux cyberattaques, tout en assurant une continuité des services critiques en cas de crise.

[194] *Thales Alenia Space, 12 décembre 2023, "ASCEND: a new alternative to terrestrial datacenters", https://www.thalesaleniaspace.com/en/news/ascend-new-alternative-terrestrial-datacenters.*

Cette approche illustre comment l'Europe peut dépasser la simple imitation des modèles américains pour se positionner comme un leader dans la transition numérique durable, en alliant innovation technologique, réduction de l'impact environnemental et renforcement de la sécurité des infrastructures numériques.

L'émergence d'acteurs comme Starlink et Kuiper, qui développent des solutions comme l'iPhone satellitaire, pose un défi majeur pour l'industrie des télécoms. Une fois cette technologie disponible, le positionnement historique des fournisseurs d'accès à internet (FAI) et des opérateurs télécoms traditionnels pourrait être durablement remis en question. **Historiquement, les opérateurs ont été perçus comme des infrastructures essentielles, mais leur ouverture à la concurrence a permis l'entrée de nouveaux acteurs. Aujourd'hui, sans opérateurs historiques dominants, l'équation change**. Cette situation de fait pose la question, au niveau européen, de mettre en place un système de FRAND *(Fair, Reasonable and Non-Discriminatory)*, qui pourrait être envisagé pour réguler l'accès à ces nouvelles infrastructures.

Dans ce contexte, deux enjeux structurants, qui mériteraient un rapport dédié, se dessinent pour la décennie à venir :

1. **Comment garantir un traitement équitable entre les fournisseurs de satellites proposant des services de télécommunication, comme Starlink et Kuiper, et les opérateurs de télécommunication traditionnels, qui sont soumis à des obligations réglementaires plus strictes ?**

2. **Comment coordonner efficacement le recours aux services satellitaires pour couvrir les zones blanches et gérer les opérations de sauvegarde, afin de poser les bases d'un déploiement réussi du projet de constellations satellitaires européen Iris² ?**

4 Le cadre normatif européen doit s'adapter aux nouvelles dynamiques technologiques mondiales

4.1. MIEUX EXPLOITER LA FRAGMENTATION DU MARCHÉ UNIQUE POUR DONNER AUX ACTEURS TECHNOLOGIQUES EUROPÉENS LES MOYENS DE S'IMPOSER SUR LE MARCHÉ MONDIAL

a. Le cadre réglementaire européen impose des obligations excessives, ce qui revient à un véritable auto-sabotage pour la compétitivité des acteurs

Dans son discours du 21 janvier 2025 à Davos, Ursula Von Der Leyen a rappelé que : *« À l'heure actuelle, le marché unique européen présente toujours bien trop de barrières nationales. Les entreprises doivent parfois composer avec 27 législations nationales.* ***En lieu et place de cela, nous proposerons aux entreprises innovantes d'exercer leurs activités dans toute l'Union suivant un ensemble unique de règles. Nous l'appelons le 28e régime****. Droit des sociétés, insolvabilité, droit du travail, fiscalité… un cadre unique et simple dans toute l'Union. Cela aidera à faire tomber les obstacles les plus courants qui empêchent de passer à l'échelon supérieur dans toute l'Europe. Car l'échelon continental est notre plus grand atout dans un monde de géants. »*

L'argument de la fragmentation du marché européen est souvent avancé comme un frein à l'émergence et à la consolidation d'acteurs compétitifs. Le rapport Letta[195], publié en avril 2024 sous le titre *More than a market*, proposait ainsi de réduire le nombre d'opérateurs télécoms en Europe de 100 à 20-30 et d'instaurer un *Pan-European Core*

[195] *E. Letta, avril 2024, Rapport "Much More Than a Market".*

Network, afin de rendre l'infrastructure réseau moins dépendante de son ancrage national et de permettre aux acteurs européens de changer d'échelle. Toutefois, cette proposition a suscité de vives oppositions, notamment de la part de l'Alliance européenne sur les Fibres Locales (ELFA) et de MVNO Europe, qui ont alerté sur le risque d'évincer les petits opérateurs, alors même qu'ils assurent encore 50 % des capacités du continent.

Le problème réside moins dans l'existence de cadres nationaux distincts que dans le manque de coordination et dans l'absence d'une approche pragmatique pour exploiter la diversité législative de l'UE à des fins de compétitivité. L'exemple de la gestion du spectre radioélectrique en est une parfaite illustration : bien que les appels d'offres pour l'attribution des fréquences relèvent de la compétence nationale, une coordination européenne existe déjà à travers le Comité du spectre radioélectrique pour les aspects techniques et le Groupe pour la politique en matière de spectre radioélectrique pour les enjeux stratégiques. De plus, cette gestion nationale n'a pas empêché l'émergence d'opérateurs pan-européens tels que Vodafone ou Orange, qui opèrent à travers plusieurs États membres.

Ce n'est donc pas tant l'existence de 27 cadres réglementaires qui pose problème que l'absence d'une harmonisation efficace sur les aspects critiques. La priorité devrait être mise sur la standardisation des infrastructures et des régulations techniques pour assurer une connectivité transfrontalière fluide, réduire les coûts d'exploitation et favoriser le développement d'écosystèmes industriels européens capables de rivaliser avec les géants internationaux. Par exemple, la France, l'Allemagne et le Royaume-Uni ont adopté des calendriers et des bandes de fréquences différents pour le déploiement de la 5G, freinant ainsi la mise en œuvre de projets communs et limitant le développement d'acteurs européens compétitifs capables de rivaliser avec des géants étrangers qui opèrent à plus grandes échelles.

Cette hétérogénéité réglementaire, pointée dans le livre blanc de février 2024 de la Commission européenne, ne se limite pas aux télécoms. Elle freine également l'essor de secteurs stratégiques comme l'ioT et la 6G, qui nécessitent une gestion cohérente des fréquences pour maximiser leur potentiel. **L'absence d'une approche coordonnée ne fragilise ainsi pas seulement le marché unique ; elle empêche aussi l'Europe de tirer parti de ses atouts en matière d'innovation et de diversité économique**. Plutôt que de chercher une uniformisation absolue, une coordination ciblée sur les enjeux structurants permettrait de transformer la diversité réglementaire européenne en avantage compétitif, au lieu de la laisser devenir un facteur de blocage.

Encadré n° 22 • Détails sur les rythmes d'attribution et de déploiement de bandes de fréquence en Europe et en France

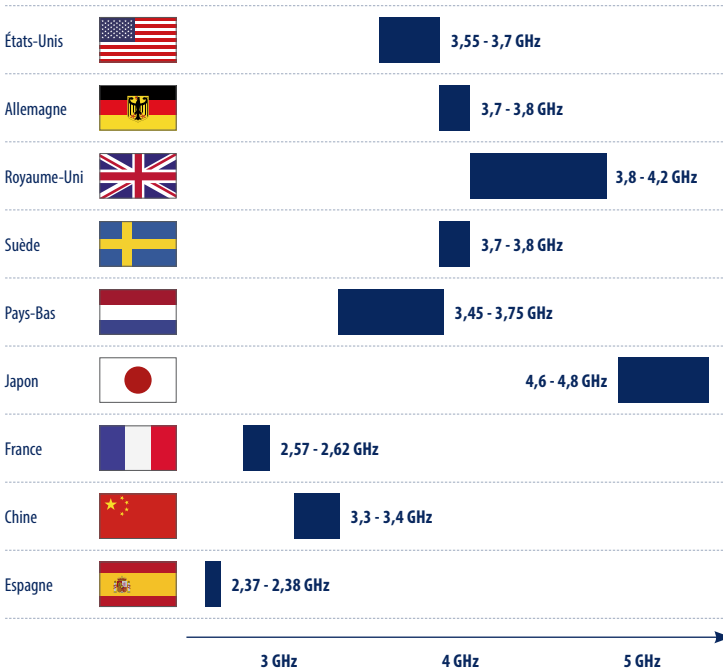
En Europe, les attributions et les déploiements des fréquences ont suivi des rythmes variables, et ont porté sur des fréquences différentes. L'Allemagne et le Royaume-Uni ont ainsi été les premiers pays à attribuer les bandes autour des 3,8 Ghz en 2019, suivis par la France et ses expérimentations industrielles. Malgré une volonté d'uniformisation dans l'utilisation des différentes gammes de fréquence de la part des instances européennes[196], les disparités demeurent. Les bandes pionnières de la 5G comme la bande 26 Ghz, sont désormais attribuées à hauteur de 73 % dans l'Union européenne[197]. L'observatoire de la 5G au niveau européen a par ailleurs souligné une faible demande pour la

[196] *Le Radio Spectrum Committee (RSC) a mandaté, dès 2021, le CEPT (Conférence européenne des administrations des postes et télécommunications) pour permettre l'utilisation partagée et sans friction de la bande 3.8-4.2 GHz.*

[197] *5G Observatory, "Biannual Report", juin 2024.*

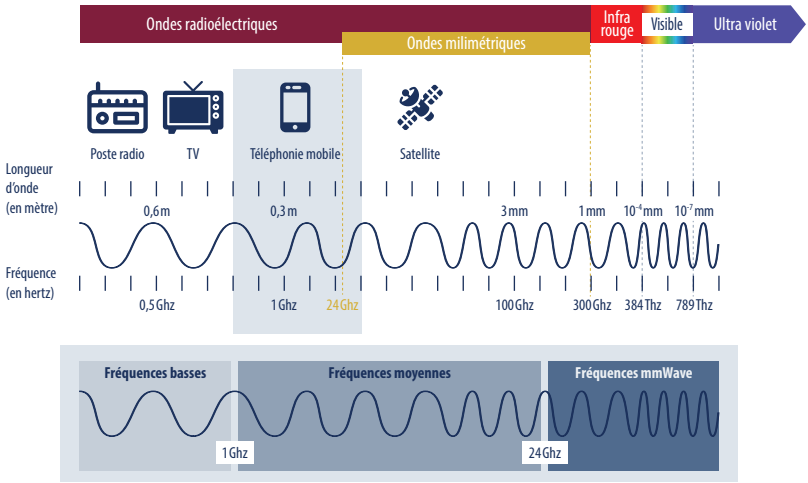
bande 26 Ghz, expliquant une appropriation et une activité juridique limitée du sujet par les gouvernements.

Graphique n° 41 • La variation des fréquences du spectre pour les verticales 5G dans différents pays

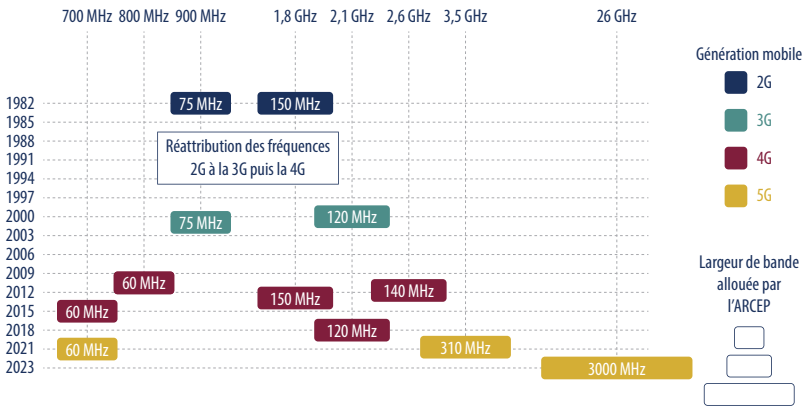


Source : Wavestone, 2021, Les ondes millimétriques et la 5G : une nouvelle dimension pour la téléphonie mobile.

Graphique n° 42 • Les ondes Radio électriques



Graphique n° 43 • Historique de l'attribution des fréquences mobiles en France



Source : Wavestone, 2021, Les ondes millimétriques et la 5G : une nouvelle dimension pour la téléphonie mobile.

La directive *Corporate Sustainability Reporting Directive* (CSRD) de 2024[198] incarne parfaitement cet auto-sabotage normatif, propre à l'Union européenne. Cette directive impose des contraintes supplémentaires aux entreprises européennes, en exigeant une transparence accrue sur les impacts environnementaux, sociaux et de gouvernance (ESG). Bien que ces obligations soient louables, elles accentuent le fossé avec les concurrents internationaux, notamment américains et chinois, soumis à des réglementations bien moins strictes. Cette asymétrie alourdit les coûts de conformité des entreprises européennes et limite leur flexibilité stratégique, posant la question de l'intégration de la sobriété énergétique comme levier de compétitivité. **Plutôt que de mettre l'accent prioritairement sur les risques et leur encadrement, le cadre européen devrait se concentrer sur des incitations efficaces pour favoriser l'émergence de bonnes pratiques sur le terrain**. Dans le cas de l'IA générative, une idée reçue continue de prévaloir : seuls les modèles volumineux garantiraient de bonnes performances. Cette logique, en érigeant des barrières à l'entrée, limite l'accès à cette technologie et détourne le débat de son véritable enjeu : la qualité des données d'entraînement. L'ouverture des données en Europe devient ainsi un levier essentiel pour promouvoir des modèles dits « frugaux », optimisés pour des tâches spécifiques, qui combinent performances élevées, coûts réduits et faible impact environnemental. **Plus largement, cette logique revient à imposer à des entreprises européennes des obligations de transparence, alors qu'il serait plus efficace d'orienter les comportements des fournisseurs d'infrastructures numériques étrangers. Une piste serait d'introduire un « bonus environnemental », sous forme d'allègements administratifs ou réglementaires, pour encourager l'adoption de services numériques éco-conçus et sobres en énergie. Cette approche permettrait aux entreprises européennes de mieux anticiper et réduire leur empreinte énergétique, tout en valorisant l'innovation durable**.

[198] *Commission européenne, Text of the CSRD (2022/2464/EU).*

Encadré n° 23 • Focus sur les différences entre le scope 1, scope 2 et scope 3

Scope 1 : émissions directes générées par l'organisation, notamment par ses installations ou ses équipements qu'elle contrôle. Par exemple, les combustions de carburant dans les véhicules de l'entreprise ou les émissions des chaudières ou générateurs utilisés sur site.

Scope 2 : émissions indirectes liées à l'énergie couvrant les émissions liées à la production d'électricité, de chaleur ou de vapeur consommées par l'organisation. Bien que ces émissions soient produites hors site (chez les fournisseurs d'énergie), elles résultent directement de l'utilisation d'énergie par l'entreprise. Par exemple, les émissions associées à l'électricité achetée pour alimenter les bureaux ou les usines, ou la production de chaleur ou de froid consommée par l'entreprise.

Scope 3 : autres émissions indirectes provenant de l'ensemble de la chaîne de valeur de l'entreprise, en amont et en aval. Elles incluent des activités qui ne sont pas directement sous le contrôle de l'organisation. Ce scope est souvent le plus vaste et difficile à mesurer. En amont, l'extraction de matières premières, la production des biens achetés, le transport des intrants, en aval, l'utilisation des produits vendus, la fin de vie des produits, le transport des marchandises jusqu'aux clients.

b. Faire du soutien à l'innovation dans les secteurs stratégiques et sensibles un enjeu de sécurité européenne

La Directive 2014/24/UE encadre les marchés publics en imposant aux États membres les principes de transparence, d'égalité de traitement et de non-discrimination, interdisant ainsi toute préférence pour les entreprises nationales ou européennes, contrairement à d'autres pays qui pratiquent des achats souverains. **L'ouverture des marchés publics européens repose toutefois sur la réciprocité, *via* l'accord plurilatéral de l'OMC ou des accords bilatéraux**. L'Instrument relatif aux marchés publics internationaux précise ainsi que les entreprises de pays tiers non couverts par un accord ne bénéficient pas d'un accès garanti et peuvent être exclues (considérant 6). Si cette réglementation favorise l'intégration du marché européen en permettant aux entreprises d'un État membre de répondre aux appels d'offres d'un autre, elle impose aussi aux pouvoirs adjudicateurs d'accorder aux entreprises de certains pays tiers un traitement équivalent à celui des acteurs européens en son article 25.

La directive prévoit des exclusions et aménagements pour les marchés comportant des aspects ayant trait à la défense ou à la sécurité ou encore ceux relatifs à la mise à disposition ou l'exploitation de réseaux publics de communications ou la fourniture au public d'un ou de plusieurs services de communications électroniques. En outre, ses dispositions s'appliquent aux pouvoirs adjudicateurs et non aux entreprises privées.

Mais, en l'absence de réciprocité effective sur les marchés publics, cela revient à exposer les entreprises européennes à une concurrence faussée, puisque dans les autres zones géographiques les autorités publiques favorisent leurs acteurs nationaux, par exemple les stratégies de dumping chinoises condamnées en Europe[199]. Cette situation a été jugée comme particulièrement préoccupante pour les

équipements critiques comme les routeurs, capteurs et objets connectés IoT, qui peuvent constituer des points d'entrée vulnérables pour les cybermenaces. Le cas de l'entreprise NucTech a été emblématique, où des équipements non européens déployés dans des infrastructures critiques (scanners pour ports et aéroports) ont soulevé des risques de cyberespionnage en raison de *backdoors* potentielles. Face à ce type de pratiques, plusieurs pays européens ont pris des mesures comme par exemple le gouvernement allemand qui a décidé en juillet dernier[200] de retirer progressivement les composants chinois, notamment ceux de Huawei et ZTE, de ses réseaux 5G d'ici 2029, invoquant des préoccupations de sécurité nationale. Déjà en juin 2023[201], la Commission européenne avait estimé que les fournisseurs chinois d'équipements télécoms Huawei et ZTE représentaient un risque pour la sécurité de l'Union et avait annoncé qu'elle n'utiliserait plus de services reposant sur leurs infrastructures.

C'est pourquoi de nombreux acteurs ont plaidé pour l'adoption d'un *Buy European Act* (BEBE) pour les équipements d'infrastructures numériques fabriqués en Europe, adapté aux spécificités du marché européen. D'autres pays à la pointe dans le secteur des infrastructures numériques ont mis en place ce type d'initiative. Le *Buy American Act* (BABA), voté en 1933, **impose aux agences fédérales américaines de privilégier l'achat de produits fabriqués aux États-Unis pour les marchés publics, à condition que leur coût ne dépasse pas de 25 % celui de produits étrangers équivalents**. Il a l'avantage de soutenir l'industrie nationale, de préserver des emplois locaux et de renforcer l'autonomie économique des États-Unis. Dans le cadre du plan de relance économique de 2009, le *Buy American Act* a permis de privilégier

[199] *Pratique consistant à vendre des produits sur un marché étranger à des prix inférieurs à leur coût de production ou au prix pratiqué sur le marché domestique, souvent grâce à des subventions étatiques.*

[200] *Financial Times, juillet 2024, "Germany orders ban on Chinese companies from its 5G network".*

[201] *Le Monde avec AFP, juin 2023, « Huawei, ZTE : la Commission européenne accuse les fournisseurs chinois de représenter un risque pour la sécurité de l'Union ».*

les entreprises américaines dans les projets d'infrastructures financés par le gouvernement, comme la construction de ponts et de routes. En Chine, l'équivalent est la politique dite « *Indigenous Innovation* » (自主创新) (ou « Zizhu Chuangxin »), mise en place dès 2006 dans le cadre du programme national de développement technologique. **Elle impose une préférence nationale pour les technologies chinoises dans les achats publics et favorise le développement de champions industriels locaux**. Elle a l'avantage de stimuler l'innovation nationale, de réduire la dépendance aux technologies étrangères et de protéger les industries stratégiques. Grâce à cette politique, la Chine produit aujourd'hui plus de 70 % des panneaux solaires utilisés dans le monde, consolidant sa domination dans les technologies vertes. Une tribune de juillet 2022 pour un *Buy European Tech Act* (BETA) avait déjà été signée par des acteurs des nouvelles technologies en ce sens pour appeler l'UE à réserver une part de ses marchés publics aux technologies européennes pour encourager à l'achat de solutions européennes.

L'adoption d'un *Buy European Tech Act* (BETA) soulève toutefois des défis juridiques majeurs, notamment en raison des engagements internationaux de l'UE. Pour exclure les services numériques de l'Accord sur les marchés publics (AMP) de l'OMC, dont l'UE est signataire, il faudrait démontrer que ces services relèvent de la « sécurité nationale », une notion strictement encadrée. **Or, si les partisans d'un BETA soutiennent que l'innovation dans les secteurs stratégiques contribue indirectement à la sécurité nationale, cette justification ne peut être invoquée que dans des cas très spécifiques définis par la directive 2009/81/CE, qui concerne principalement la défense et la sécurité. À ce jour, aucune exception juridique ne permet d'exclure les services numériques sur la base d'un soutien à l'innovation stratégique. Une alternative consisterait à renégocier l'AMP pour retirer les services numériques de son champ d'application, mais une telle démarche serait complexe et politiquement sensible.** En outre, un BETA permettrait-il réellement de résoudre les faiblesses structurelles du marché numérique européen ? Le problème de la taille des acteurs

européens ne découle pas tant de l'absence d'un BETA que du retard dans l'intégration et la consolidation du marché, faute d'une politique industrielle ambitieuse et du conditionnement de l'attribution de marchés publics à des critères techniques plus exigeants.

Les nouvelles obligations introduites par le règlement européen pour une industrie « zéro-net » dit « NZIA » témoignent d'une approche plus stratégique des marchés publics européens. **Plutôt que d'instaurer une préférence explicite pour les entreprises européennes, ces mesures imposent des critères de cybersécurité et des limitations de dépendance à des fournisseurs tiers, ce qui revient à favoriser indirectement les acteurs industriels européens**. L'exigence de conformité avec le futur règlement sur la cyber-résilience crée un avantage compétitif pour les entreprises européennes, généralement plus alignées avec les standards de sécurité de l'UE. Cette contrainte impose aux fournisseurs étrangers des coûts d'adaptation et de mise en conformité supplémentaires, réduisant ainsi leur compétitivité sur certains segments stratégiques. De plus, la limitation à 50 % de l'approvisionnement en technologies « zéro net » ou en composants issus d'un seul pays tiers réduit la domination des fournisseurs extra-européens sur le marché. Dans un secteur où la production repose souvent sur quelques acteurs dominants, notamment asiatiques, cette règle encourage une diversification des sources d'approvisionnement et favorise mécaniquement les industriels européens capables d'offrir une alternative crédible. **En combinant exigences de cybersécurité et diversification des chaînes d'approvisionnement, l'UE adopte ainsi une forme de préférence industrielle sans entrer en contradiction avec ses engagements commerciaux internationaux**. Autrement dit, plutôt que d'imposer un *« Buy European Tech Act »*, elle encadre les conditions de marché pour que les entreprises européennes disposent de meilleures chances face à la concurrence mondiale.

4.2. CONTRE LA CYBERMENACE, RENFORCER LE CADRE JURIDIQUE PROTÉGEANT LES COMPOSANTS ESSENTIELS DES INFRASTRUCTURES NUMÉRIQUES

La sécurité est au cœur de la résilience des infrastructures numériques, devenues essentielles pour tous les secteurs d'activité. Les questions de cybersécurité se posent dès qu'une donnée transite sur un réseau ouvert, exposant le système à des vulnérabilités. Pour garantir la continuité des services essentiels, ces infrastructures doivent atteindre des taux de disponibilité toujours plus élevés. Cela nécessite une redondance rigoureuse à chaque maillon du réseau, afin de prévenir les interruptions dues à des pannes ou des cyberattaques. **Aujourd'hui, la sécurité est intégrée de manière systématique à chaque étape du cycle de vie des infrastructures numériques : depuis la conception et la fabrication des composants, jusqu'aux phases d'exploitation et de maintenance**.

La cybersécurité remonte aujourd'hui dans la chaîne de valeur pour se positionner à tous les niveaux des infrastructures numériques : depuis la sécurisation et la gestion de la localisation des équipements de base, comme les centres de données et les réseaux, jusqu'à la protection des services immatériels qui en dépendent. Cette évolution implique que chaque composant de l'infrastructure doit être conçu avec une résilience intrinsèque, incluant des technologies comme le chiffrement de bout en bout, la segmentation réseau et le monitoring automatisé des menaces. Par conséquent, cela nécessite **une approche de type « *zero trust* », où chaque accès, utilisateur ou système est constamment vérifié et authentifié, pour garantir une sécurité renforcée dans un environnement de plus en plus distribué**.

a. Une sophistication croissante des cyberattaques sur les infrastructures numériques

L'intensification des cyberattaques, tant en volume qu'en sophistication, nécessite une protection renforcée à tous les niveaux : matériel, logiciel et données en transit. La cybersécurité, désormais un enjeu stratégique majeur, doit être intégrée dès la phase de conception des infrastructures numériques, de manière dite « *secure by design* ». Cette approche proactive réduit les vulnérabilités qui surviennent fréquemment lorsque la sécurité est ajoutée a posteriori. La montée en puissance des menaces, exacerbée par la prolifération de l'IoT, complique davantage la détection et la réponse aux attaques, soulignant l'urgence d'une cyber résilience pensée dès l'assemblage des composants et équipements. **La cybersécurité ne doit pas être perçue comme un simple support, mais comme une composante essentielle, intégrée à chaque étape des processus numériques, de la conception à l'exploitation opérationnelle**. C'est pour pallier ce problème que le *Cyber Resilience Act*[202], entré en vigueur en décembre 2024, prévoit de mettre en place des règles de sécurité dès la conception des équipements numériques.

Encadré n° 24 • Focus sur le *Cyber Resilience Act* (CRA)

Le *Cyber Resilience Act* (CRA), adopté en octobre 2024, vise à renforcer la sécurité des produits numériques commercialisés dans l'Union européenne. Ce règlement impose aux fabricants et fournisseurs de garantir la cybersécurité de leurs produits tout au long de leur cycle de vie, notamment en intégrant des mécanismes de sécurité dès la conception et en fournissant des mises à jour régulières pour corriger les vulnérabilités.

[202] *Commission européenne, Cyber Resilience Act (CRA), décembre 2024.*

Le CRA s'applique à une large gamme de produits comportant des éléments numériques, tels que les objets connectés (IoT), les logiciels embarqués et les équipements informatiques. Il introduit des exigences strictes pour protéger les consommateurs et les entreprises contre les cybermenaces, tout en favorisant la confiance dans l'écosystème numérique européen. Les produits doivent, par exemple, être livrés avec des paramètres sécurisés par défaut et respecter des normes élevées de gestion des risques. En cas de non-conformité, des sanctions significatives sont prévues, pouvant atteindre 15 M€ ou 2,5 % du chiffre d'affaires annuel mondial.

Les cyberattaques représentent une activité très lucrative, avec un coût mondial estimé à 8 000 Mds $ en 2023[203]. En Allemagne, les pertes dues à la cybercriminalité ont atteint 267 milliards d'euros en 2024, marquant une hausse de près de 30 %[204] par rapport à l'année précédente. Les attaques dites de chaîne d'approvisionnement *(supply chain attacks)* illustrent particulièrement bien cette rentabilité. Par exemple, l'attaque SolarWinds révélée en 2020[205] a démontré l'ampleur des gains potentiels. En infiltrant le logiciel Orion, une plateforme de gestion informatique, *via* des mises à jour infectées, les attaquants ont compromis les systèmes de milliers d'utilisateurs, y compris des agences gouvernementales et des multinationales. Cette intrusion discrète a permis de monétiser des données stratégiques, telles que des informations confidentielles ou des secrets commerciaux, par leur revente sur des marchés noirs, leur exploitation dans des attaques ciblées, ou encore par le biais de ransomwares.

[203] *Statista, Cybersecurity Worldwide.*

[204] *Reuters, août 2024,"Cybercrime and sabotage cost German firms $300 bln in past year".*

[205] *Techtarget , novembre 2023, "SolarWinds hack explained: Everything you need to know".*

En 2023, le panorama de la cybermenace[206] publié par l'ANSSI a mis en évidence une recrudescence des attaques, notamment contre les téléphones portables, avec des objectifs d'espionnage, d'extorsion et de déstabilisation politique, dans un contexte où près de 50 % de la population doit voter en 2024. L'ANSSI a également noté une sophistication croissante des attaquants, notamment à travers l'utilisation de failles de type « *zero-day* » (cf. infra).

Encadré n° 25 • Focus sur le cadre réglementaire en vigueur

- En France, la régulation en cybersécurité a d'abord ciblé la protection des données personnelles avant de s'étendre aux systèmes critiques. Créée en 2009, l'ANSSI joue un rôle central en sécurisant les opérateurs d'importance vitale (OIV), répartis en douze secteurs essentiels : alimentation, santé, énergie, transports, communications électroniques, etc. Ces OIV sont protégés sur deux dimensions : physique, *via* l'identification de sites vitaux, et numérique, *via* les systèmes d'information d'importance vitale (SIIV). Par exemple, dans le domaine du trading, la fiabilité des prestataires est cruciale, justifiant des labels basés sur leur solidité financière et leur capacité à servir les infrastructures critiques.

- La Loi de Programmation Militaire (LPM) de 2013 a introduit des exigences strictes pour protéger les systèmes critiques français. À l'échelle européenne, la directive NIS adoptée en 2016 visait à harmoniser la sécurité des infrastructures critiques et des opérateurs de services essentiels (OSE). Face à

[206] *ANSSI, Panorama de la cybermenace 2023.*

l'évolution des cybermenaces, elle a été remplacée par NIS 2 en 2022, qui élargit son champ d'application et impose des normes plus rigoureuses pour les opérateurs critiques. Ces initiatives illustrent une volonté croissante de renforcer la résilience collective des infrastructures critiques en Europe.

→ La directive NIS 2 intègre un plus grand nombre de secteurs, y compris les entreprises stratégiques de taille moyenne, et impose des obligations renforcées en cybersécurité. Elle exige l'usage de technologies avancées (TDR, SIEM), une notification rapide des incidents sous 24 heures et identifie des secteurs critiques tels que l'énergie, la santé, les *data centers* ou encore les infrastructures *cloud*. En élargissant sa portée à des secteurs comme les banques, les réseaux sociaux ou les hôpitaux, NIS 2 vise à créer un écosystème de cybersécurité européen robuste et souverain, favorisant une coopération internationale accrue pour répondre aux menaces émergentes.

Dans son livre blanc[207] publié le 21 février 2024, la Commission européenne a aussi mis en avant la vulnérabilité particulière des câbles sous-marins en matière de cybersécurité. Ce document souligne que les activités de recherche et d'innovation (R&I) concernant les câbles sous-marins et terrestres sont cruciales pour renforcer la sécurité et la résilience des infrastructures numériques. **Les attaques visant à écouter ou capturer des données, qui constituent les principales menaces actuelles, combinent souvent des opérations numériques et physiques, en s'appuyant sur les infrastructures numériques telles que les réseaux, les satellites et les câbles sous-marins**.

[207] *Commission européenne, février 2024, Livre blanc – Comment maîtriser les besoins en infrastructures numériques de l'Europe ?*

En parallèle, les autorités françaises ont mis en place des dispositifs pour sécuriser les réseaux de télécommunication les plus critiques face à ces menaces. À ce titre, l'Agence des Communications Mobiles Opérationnelles de Sécurité et de Secours (ACMOSS) est un établissement public français créé en mars 2023. Elle est chargée de concevoir, de déployer, de maintenir et d'exploiter des services de communication mobile critique très haut débit pour les missions de sécurité, de secours, de protection de la population et de gestion des crises. C'est dans ce cadre qu'est actuellement conçu le réseau radio du futur (RRF), qui vise à remplacer les réseaux de télécommunications utilisés par les services publics tels que la police, la gendarmerie, les pompiers et le SAMU. Conçu, développé et exploité par l'ACMOSS, il a vocation à offrir un système de communication commun, prioritaire, sécurisé et haut débit, bénéficiant des meilleures technologies numériques et d'un haut niveau de résilience en cas de crise.

b. L'immixtion du logiciel dans les équipements physiques *(hardware)* : un facteur d'aggravation des risques en cybersécurité dès les premières étapes de la chaîne de valeur

Les vulnérabilités en cybersécurité se manifestent dès les premières étapes de la chaîne de valeur des infrastructures numériques, c'est-à-dire dès la conception et l'assemblage de leurs composants. **Cette intrusion précoce est accentuée par l'intégration croissante des logiciels dans ces infrastructures, faisant du logiciel un vecteur stratégique pour les cyber attaquants qui peuvent pré-positionner des cibles au sein des systèmes numériques**. Les attaquants peuvent exploiter ces failles pour insérer des portes dérobées au moment de l'installation ou des mises à jour des composants réseau. Les opérations préparatoires de ces attaques peuvent s'étendre sur plusieurs années, permettant aux attaquants d'installer des portes dérobées dans les systèmes et de faciliter un accès ultérieur *via* des logiciels malveillants.

Certains composants logiciels essentiels, comme les BIOS *(Basic Input/Output System)*, qui permettent de démarrer un ordinateur, sont particulièrement vulnérables aux attaques. Par exemple, en 2018, un virus nommé Lojax a été découvert. Ce malware, développé par le groupe russe APT28, se cachait dans le processus de démarrage du système, permettant aux attaquants d'accéder au système de manière furtive et durable. **Lojax ciblait spécifiquement les technologies plus récentes, comme les firmwares UEFI *(Unified Extensible Firmware Interface)*, qui ont remplacé les anciens BIOS, soulignant ainsi la manière dont les logiciels peuvent être exploités dès les étapes fondamentales du fonctionnement des systèmes informatiques**. Cette attaque illustre les risques liés à l'intrusion logicielle dans des composants critiques, souvent difficiles à détecter et à neutraliser. Ces schémas d'attaques s'appuyant sur des composants matériels sont à la base d'une nouvelle famille de menaces ciblant les systèmes industriels : les ICS attacks *(Industrial Control Systems attacks)*. Par exemple, les campagnes d'attaque Berserk Bear, Dragonfly 2.0, Havex, Triton et Trisis ont consisté en l'infiltration informatique de centrales électriques et de raffineries dans divers pays.

Face à ces attaques touchant les composants fondamentaux des infrastructures numériques, l'homologue américain de l'ANSSI, la *Cybersecurity and Infrastructure Security Agency* (CISA) a été créée en 2018. Cette agence publie régulièrement des alertes à destination des opérateurs critiques. Dans son rapport du 7 février dernier[208], elle a indiqué que des groupes de cybercriminels se sont pré-positionnés sur des réseaux informatiques en vue de mener des cyberattaques perturbatrices ou destructrices contre des infrastructures critiques américaines, notamment dans les secteurs de l'eau et de l'énergie. La majorité des compromissions sont dues au vol d'informations d'identification, comme les mots de passe. Les cyberattaques sont également facilitées

[208] *CISA, 7 février 2024, "PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure".*

par des « courtiers d'accès initial » *(initial access brokers)*, spécialisés dans la vente d'accès initial à des systèmes informatiques compromis (informations de compte, accès directs obtenus à la suite de l'exploitation d'une vulnérabilité, etc.). **Avec l'augmentation de l'interopérabilité des infrastructures numériques, la menace cyber devrait s'intensifier, car les multiples passerelles entre produits et systèmes différents élargissent la surface d'attaque et complexifient les efforts de défense**.

Actuellement, le *Common Vulnerability Scoring System* (CVSS) est le principal cadre de référence pour évaluer les vulnérabilités des infrastructures numériques, sur une échelle de 0 à 10 en fonction de leur gravité et de leur impact potentiel sur la sécurité des systèmes. **Les vulnérabilités les plus critiques sont les vulnérabilités « *zero-click* » et « *zero-day* » résultant en la prise de contrôle à distance d'un système ou d'un équipement**. Les failles *zero-click* permettent à un attaquant de prendre le contrôle d'un appareil sans aucune interaction de l'utilisateur. Un exemple notable est le logiciel espion Pegasus, qui aurait infiltré des centaines de milliers de smartphones sous iOS ou Android sans laisser de traces visibles, accédant aux données des appareils ciblés. Les failles *zero-day* sont des vulnérabilités inconnues de l'éditeur du logiciel et pour lesquelles aucun correctif n'est encore disponible, ce qui permet aux attaquants de les exploiter sans être détectés. L'exemple emblématique est la faille *zero-day* affectant le logiciel Log4j en 2021. Présent sur de nombreuses plateformes comme iCloud, Amazon et X (anciennement Twitter), Log4j a permis aux attaquants d'accéder à près de 40 % des réseaux mondiaux en exploitant cette vulnérabilité. Il suffisait, par exemple, de renommer un iPhone avec une chaîne de caractères spécifique pour obtenir un accès aux serveurs d'Apple.

Toute la difficulté consiste à identifier tous les dispositifs utilisant des bouts de logiciel potentiellement vulnérables dans un contexte où les « zones grises » entre cyberattaquants et cyberdéfenseurs augmentent et renforcent la pression sur les équipes chargées de

la prévention et de la réaction aux cyberattaques. Les attaquants exploitent souvent le laps de temps entre la découverte d'une vulnérabilité et le déploiement des correctifs. Un cas marquant est l'utilisation par la NSA de failles logicielles pour créer EternalBlue, un outil capable de cibler des vulnérabilités du protocole de communication réseau de Windows. Malgré la publication par Microsoft du correctif MS17-010, de nombreux utilisateurs et administrateurs n'ont pas installé la mise à jour à temps, permettant aux hackers de lancer l'attaque massive WannaCry. Ce ransomware, combinant EternalBlue avec une capacité d'auto-réplication, s'est propagé à une vitesse alarmante, infectant en 24 heures 230 000 machines Windows dans 150 pays. **Ce déplacement des menaces cyber vers les composants fondamentaux des infrastructures numériques souligne la nécessité critique de mesures proactives dès les premières étapes de la chaîne de valeur, en mettant l'accent sur une conception sécurisée, des tests rigoureux des composants et un déploiement rapide des correctifs pour atténuer les risques**.

c. Des nouvelles menaces cyber liées à l'IA et au quantique à intégrer dans la définition de nos outils de sécurité

La diffusion de l'IA a entraîné l'émergence de nouvelles formes de cyberattaques, telles que l'infiltration des modèles d'IA, la génération de contenus dangereux et l'extraction de données sensibles. Par exemple, la manipulation des modèles de langage (LLM) par des séquences de questions ciblées peut modifier leur comportement de façon malveillante. Bien que des entreprises comme Haize Labs collaborent avec Anthropic, une société d'IA, pour développer des algorithmes de détection et de correction de ces failles, les vulnérabilités restent nombreuses. L'IA elle-même est exposée à des attaques telles que l'empoisonnement de données, où des informations fausses sont insérées durant l'apprentissage, ou l'évasion, où des leurres numériques

trompent les réponses du modèle. L'inférence, utilisée pour contourner les systèmes de détection anti-fraude, représente également un danger. De plus, des techniques comme la sténographie permettent aux algorithmes de communiquer discrètement, dissimulant des messages dans des images diffusées sur des réseaux sociaux, augmentant ainsi la complexité de la détection des menaces.

Pour lutter contre ces menaces, des outils tels que les plateformes de détection d'attaques adverses *(Adversarial Attack Detection*[209]*)* existent et permettent d'identifier et de neutraliser les modifications malveillantes dans les modèles d'apprentissage automatique. Pendant la phase d'entraînement, cela consiste à exposer le modèle à des exemples adversariaux générés en temps réel, comme exemple des perturbations infimes des données d'entrées. Des pénalités peuvent être introduites d'emblée pour limiter la sensibilité du modèle à de petites perturbations dans les données d'entrée. Par ailleurs, des techniques comme la « robustesse différentiable » améliorent la résilience des algorithmes face aux attaques en renforçant les données et les modèles. Cette méthode ajuste les paramètres du modèle pendant l'entraînement pour garantir que les prédictions ne changent pas significativement face à des perturbations mineures. **Des frameworks comme ART *(Adversarial Robustness Toolbox)* de la fondation Linux offrent des solutions open source pour évaluer et améliorer la sécurité des modèles d'IA**.

En outre, la montée en puissance des technologies post-quantiques représente un défi majeur pour la sécurité des infrastructures numériques. Les ordinateurs quantiques, encore en développement, pourraient remettre en cause la robustesse des systèmes de cryptographie actuels, qui sous-tendent la plupart des communications sécurisées. Par exemple, l'algorithme de Shor permet de factoriser efficacement

[209] *M. Ciolino, J. Kalin, D. Noever, juin 2021, Fortify Machine Learning Production Systems: Detect and Classify Adversarial Attacks.*

de grands nombres, compromettant les systèmes à clés asymétriques comme RSA (Rivest-Shamir-Adleman) et ECC (Elliptic Curve Cryptography). Ces systèmes reposent sur un couple de clés – une publique pour le chiffrement, l'autre privée pour le déchiffrement – dont la sécurité dépend de la difficulté de factorisation, un problème que les ordinateurs quantiques pourraient résoudre rapidement. De plus, l'algorithme de Grover, qui optimise la recherche dans des bases de données non structurées, pourrait affaiblir les systèmes de chiffrement symétrique en réduisant le temps nécessaire pour deviner une clé de chiffrement. Cela engendre deux risques principaux : d'une part, la compromission des algorithmes de cryptographie asymétrique largement utilisés aujourd'hui, et d'autre part, un risque de collecte anticipée de données chiffrées qui pourraient être déchiffrées ultérieurement par des ordinateurs quantiques. **Face à ces menaces, les acteurs français se trouvent confrontés au défi d'intégrer rapidement des solutions de chiffrement « quantum-safe », telles que la cryptographie post-quantique (PQC) lorsqu'elle sera largement disponible, la distribution de clés quantiques (QKD), ou le chiffrement symétrique dans les infrastructures réseau**. Ces technologies visent à garantir la résilience des systèmes critiques tout en tenant compte des contraintes budgétaires et technologiques. Leur adoption ne doit pas seulement protéger les communications actuelles, mais aussi anticiper une reconfiguration structurelle des réseaux, en inventant des méthodes encore plus robustes capables de faire face aux capacités croissantes des ordinateurs quantiques.

Encadré n° 26 • Les solutions Quantum Safe, un exemple de défense active

Les ordinateurs quantiques, grâce à leur puissance de calcul en parallèle, pourraient rendre obsolètes les systèmes de chiffrement traditionnels, comme RSA, en exploitant des algorithmes tels que celui de Shor pour casser rapidement des clés de sécurité. **Pour contrer cette menace, des solutions dites « Quantum Safe », développées par des acteurs comme Thalès, CryptoNext, ou Nokia, intègrent des méthodes de chiffrement capables de résister aux capacités des ordinateurs quantiques**. Ces technologies permettent une distribution sécurisée des clés et l'intégration de protocoles robustes, déjà opérationnels sur les réseaux IP et la couche Optique/WDM. Ces innovations soutiennent la mise en conformité avec les directives NIS 2 et DORA, garantissant la protection des données critiques face aux menaces futures.

Cependant, de nombreuses organisations n'ont pas encore déployé des solutions de prévention et de sensibilisation à la hauteur de l'enjeu. En Allemagne, par exemple, bien que la cybercriminalité et le sabotage aient coûté environ 267 Mds € aux entreprises au cours de l'année écoulée, seulement 37 % des entreprises disposent de plans d'urgence pour les incidents dans la chaîne d'approvisionnement. Une étude de Thalès[210] publiée en juin 2024 révèle que les ressources *cloud* sont désormais les principales cibles des cyberattaques, avec moins de 10 % des entreprises chiffrant leurs données *cloud* les plus sensibles.

[210] *Thales, 25 juin 2024, "Cloud Resources Have Become Biggest Targets For Cyberattacks, Finds Thales".*

Recommandation 9

Adapter et simplifier le cadre normatif européen pour renforcer notre compétitivité et favoriser les consolidations d'acteurs à l'échelle mondiale.

Recommandation 9.1 : simplifier les règles administratives européennes en tirant parti de l'instauration d'un 28e régime européen pour alléger les contraintes administratives pesant sur les entreprises européennes. En particulier, tirer les conséquences des enseignements de la *Corporate Sustainability Reporting Directive* (CSRD) pour s'assurer que nos engagements en matière de développement durable et de démocratie se traduisent en avantages concurrentiels sur la scène mondiale, plutôt qu'en obstacles supplémentaires pour nos entreprises.

Recommandation 9.2 : le plus rapidement possible introduire des critères stratégiques dans les achats publics pour favoriser les entreprises européennes. Pour cela, mettre en place des critères de cybersécurité et de limitation des dépendances à des fournisseurs tiers, sur le modèle du règlement *Net Zero Industry Act* (NZIA). Ces critères doivent imposer aux acteurs étrangers des coûts de mise en conformité proportionnels aux risques de dumping ou d'ingérence économique, tout en incitant à une diversification des chaînes d'approvisionnement. Ils pourraient par exemple inclure des

exigences d'approvisionnement minimum en composants d'infrastructures numériques européens à haute efficacité énergétique (dits frugaux). Cette approche permettrait d'atteindre les objectifs d'un *Buy European Tech Act* sans remettre en cause nos engagements internationaux.

Recommandation 9.3 : renforcer la sécurité et la résilience des infrastructures critiques grâce aux appels d'offres européens. Pour y parvenir, accélérer l'intégration de critères favorisant les acteurs et composants européens dans les appels d'offres, en mettant l'accent sur la sécurité avancée et la transparence des infrastructures numériques critiques. Pour garantir l'effectivité du *Cyber Resilience Act*, les marchés publics européens doivent mieux valoriser les outils d'évaluation des vulnérabilités dès la conception des logiciels et équipements. En complément, les labels attribués aux fournisseurs de services pour Opérateurs d'Importance Vitale (OIV) doivent inclure des critères de solidité financière et de viabilité à long terme, afin d'assurer leur capacité à répondre durablement aux exigences des infrastructures critiques.

Ce rapport intervient à un moment clé, où la convergence entre les technologies de traitement et de réseau révèle un immense potentiel de valeur. L'enjeu est de définir une stratégie pour nos infrastructures numériques qui, tout à la fois, capitalise sur les atouts existants et nous permet d'aborder les besoins futurs. Le point de départ de cette réflexion doit être l'ambition portée sur les usages et leur évolution constante, à l'aube d'une révolution scientifique, économique, industrielle et sociétale.

Objectif n° 1

Pour les 10 prochaines années, convenir d'une stratégie étatique d'infrastructures numériques, indispensable à la compétitivité et à la souveraineté de notre pays

Les moyens financiers issus des partenariats publics privés (PPP) signés à l'occasion du Sommet de l'Action de l'IA devront être fléchés dans la durée, et s'inscrire dans une politique cohérente qui dépasse les cycles électoraux. Dans un contexte marqué par un fort taux d'endettement et une instabilité politique en France, il est impératif de libérer les initiatives privées. Trop souvent, les grands changements industriels ou technologiques sont envisagés comme dépendants de l'impulsion publique, sans implication suffisante des entreprises privées ni prise en compte réaliste de l'état du marché, de la concurrence et des besoins.

Recommandation 1

Construire une offre souveraine cloud-réseau-edge-IoT « bout-en-bout » au niveau français et européen pour des usages aux dépendances maîtrisées.

Que souhaitons-nous réellement maîtriser et pourquoi ? Les réseaux, les infrastructures, les applications, les données ou les usages ? Cette réflexion implique de déterminer le périmètre géographique pertinent, et en particulier la taille du marché adressable, et les éléments de la « *stack* » technologique pour lesquels il faut faire émerger des acteurs souverains. La réponse réside sans doute dans une combinaison de ces dimensions. Dans un contexte où le paysage des infrastructures numériques évolue rapidement, nos opérateurs télécoms risquent d'être relégués aux segments à plus faible valeur ajoutée de la chaîne de valeur alors qu'ils disposent d'avantages comparatifs technologiques majeurs. Cela exige donc une anticipation des technologies et des usages souverains afin de planifier les options de mitigation des risques.

Pour relever ces défis, notre réponse repose sur trois axes essentiels : lutter contre l'ingérence technologique, renforcer la sécurité des systèmes numériques et mener la bataille du B2B industriel grâce à une offre compétitive et souveraine. Ces priorités sont indispensables pour préserver notre autonomie technologique tout en restant compétitifs à l'échelle mondiale.

Au vu des capitaux nécessaires pour développer en propre des plateformes numériques et compte tenu de nos dépendances aux chaînes d'approvisionnement de composants, matériels et intergiciels reliant les matériels aux applications, **il est impératif de comprendre les possibilités technologiques qui nous sont encore accessibles pour poser nos arbitrages. Les infrastructures de traitement de données**

de proximité, dites « edge », sont un élément de réponse adapté à des usages spécifiques car cette technologie est particulièrement appropriée à la cartographie et à la nature des données industrielles de nos ETI et PME.

Recommandation 1.1 : mettre en place une stratégie d'intégration verticale du continuum cloud-réseau-edge-IoT dans les secteurs suivants : la santé, les systèmes critiques (mission critical systems), l'aviation, la mobilité, les lanceurs spatiaux et la finance (cf. graphique n° 9).

Favoriser la création d'un acteur d'envergure couvrant l'ensemble de la chaîne de valeur, exclusivement dédié à ces secteurs stratégiques. Cette initiative doit s'appuyer sur les industriels français proposant déjà des services multiplateformes (Atos, Orange, Dassault Systèmes, Bouygues, Illiad, Docaposte, Eclairion, Sesterce, etc.), tout en évitant de disperser les efforts dans des secteurs où la loi du marché peut prévaloir.

Développer des passerelles IoT, des serveurs edge, des micro *data centers* et des solutions de 5G privée pour garantir la disponibilité et la sécurisation des données sur les territoires concernés. **Ces infrastructures devraient être déployées en priorité sur des sites sensibles, permettant d'expérimenter l'automatisation et la sécurisation des processus industriels critiques. À terme, elles serviraient de socle à un projet ambitieux de « territoires d'avenir »**, où l'Europe aurait un contrôle total sur les données sensibles, qu'il s'agisse de la gestion des infrastructures (routes intelligentes, gestion énergétique) ou de l'offre de services aux populations et aux entreprises.

Promouvoir une architecture cloud-edge-IoT combinant des connexions fibre haut débit reliant les bassins de population, économiques et industriels, et des solutions sans fil pour la périphérie. Les boxes 5G proposées

par Orange, Bouygues, Free ou SFR, offrant des débits proches de la fibre à un coût d'infrastructure réduit (une antenne étant bien moins onéreuse que le câblage intégral d'un quartier), illustrent le potentiel de ce modèle.

Explorer une approche complémentaire reposant sur des solutions satellitaires, telles que celles proposées par OneWeb, adaptées à des besoins spécifiques comme le *backup* dans les territoires ultra-marins, la couverture des zones blanches ou les régions difficiles d'accès. Ces infrastructures, encore coûteuses, nécessitent un financement à l'échelle européenne et restent pour l'instant majoritairement orientées vers des usages de niche en expansion.

Recommandation 1.2 : permettre une meilleure accessibilité de la donnée au plus proche de l'équipement de l'utilisateur pour permettre une offre souveraine fonctionnelle par le biais d'incitations à l'égard des créateurs et détenteurs de données.

Ces incitations peuvent être imposées aux collectivités locales et aux acteurs publics à des fins de participation à l'innovation nationale et d'exemplarité.

Elles peuvent être financières à l'égard d'acteurs privés (en santé, en formation, en mobilité…) afin de favoriser un partage optimal permettant ainsi la standardisation des formats de données industrielles, et leur structuration en temps réel.

Recommandation 1.3 : sécuriser l'offre souveraine qui en découlerait en constituant des *data spaces* (infrastructures de partage sécurisé de données) français dédiés.

Ce dispositif devra faire l'objet d'une redondance systématique sur les couches logicielles et s'assurer de la solidité de son approvisionnement en composants critiques. Il sera sans doute nécessaire de renforcer les prérogatives (autorités et moyens) du Délégué interministériel aux approvisionnements en minerais et métaux stratégiques et des instances impliquées (DGE et SISSE).

La question de l'investissement et de la participation européenne à un projet souverain reste aujourd'hui sans réponse véritable.

Objectif n° 2

Mieux dimensionner nos infrastructures numériques de traitement de données pour faire face aux besoins

L'enjeu principal consiste à structurer une commande publique d'infrastructures numériques de traitement capable de justifier et d'amorcer une demande privée, afin d'exporter l'offre française d'infrastructures numériques de traitement. Cette dynamique est d'autant plus urgente que la compétition internationale s'intensifie, avec des acteurs comme Nvidia qui ont récemment annoncé, lors du CES de Las Vegas, des machines edge commercialisées à 3 000 $, contre 15 000 à 40 000 € pour leurs équivalents français.

Dans ce contexte, une approche cloisonnée doit être évitée, au profit de collaborations internationales accrues, car elle risquerait d'affaiblir la capacité de la France à répondre aux besoins des marchés internationaux.

Actuellement, la stratégie sur les supercalculateurs repose sur l'achat de matériel américain, une dépendance acceptée, mais partiellement compensée par l'intégration de briques logicielles françaises pour limiter les risques liés à la souveraineté. **Cependant, la France dispose d'acteurs capables d'émerger sur ce marché et de développer des solutions compétitives à l'export**. Si la puissance publique ne peut financer seule l'ensemble des investissements, elle doit néanmoins définir un niveau de commande publique suffisant pour stimuler la demande privée et garantir un effet d'entraînement durable.

Le défi est notamment de prendre le relais des initiatives comme les machines exascale de type GENCI, en adaptant ces technologies aux cas d'usage industriels et commerciaux du supercalcul. Bien que les besoins précis en infrastructures ne seront clairs que dans un à deux ans, le temps que les cas d'usage se précisent, il est impératif d'évaluer dès à présent le potentiel de création de valeur qui justifiera ces investissements.

Pour rentabiliser ces efforts, les entreprises françaises doivent impérativement se positionner sur le marché mondial, qui recherche de plus en plus des solutions alternatives aux offres américaines ou chinoises. Ce contexte représente une opportunité stratégique unique pour la France. **La réussite dépendra de la capacité des pouvoirs publics à réellement soutenir les entreprises françaises dans leur effort d'exportation, conformément aux ambitions affichées dans la déclaration de politique générale du Premier ministre**.

Recommandation 2

Entamer, dès aujourd'hui, *a minima*, en France, la construction de 6 supercalculateurs exaflopiques additionnels afin de proposer à l'Europe une capacité de calcul de 9 exaflop.

Le besoin en puissance de calcul, recherche et industries incluses est estimé à 15 exaflops d'ici à 2030, soit une part de marché mondiale de 25 % pour l'Europe. Cela correspond à une base installée de 15 supercalculateurs exaflopiques nécessitant donc la construction de 12 supercalculateurs additionnels européens.

Nous préconisons le strict minimum pour rester dans la course mondiale, compte tenu de nos marges de manœuvre limitées, soit 6 supercalculateurs exaflopiques additionnels. Cela permettrait à l'Europe de sécuriser 15 % de la puissance de calcul mondiale.

Les avis divergent quant à la nécessité de confier cette ambition à l'Europe ou à la France en particulier.

L'estimation budgétaire nécessaire pour construire ces infrastructures repose sur une analyse comparative des projets existants. **Le supercalculateur Alice Recoque, installé en 2026 sur le site TGCC du CEA, a été cofinancé par EuroHPC à hauteur de 540 M€ sur cinq ans dans le cadre d'un partenariat avec la France (GENCI) et les Pays-Bas**. Sur cette base, et en prenant en compte l'évolution des coûts liés à l'approvisionnement en GPU et à l'échelle des infrastructures, le financement nécessaire est estimé comme suit : **3 Mds € correspondant à 6 supercalculateurs exaflopiques et 6 Mds € correspondant à 12 supercalculateurs exaflopiques**. Ces estimations reposent sur une répartition équilibrée des financements : 50 % issus des crédits EuroHPC et 50 % provenant de fonds publics et privés, notamment des entreprises européennes du *cloud* prêtes à investir dans des infrastructures stratégiques.

Recommandation 2.1 : confier à la France la construction de 6 supercalculateurs sur le territoire national, le reste pouvant être réparti sur le territoire européen, en capitalisant sur nos capacités existantes (composants, compétences et énergie).

La répartition concertée des responsabilités entre le public et le privé sera une condition de succès.

Les infrastructures publiques doivent conserver leur rôle central dans l'entraînement de modèles de grande échelle, en s'appuyant sur des initiatives européennes.

Les acteurs du *cloud* privé ne doivent pas limiter leurs investissements à l'entraînement de modèles d'IA, une tâche déjà prise en charge par des infrastructures publiques comme celles d'Euro HPC ou du GENCI en France. **Ils doivent se concentrer sur l'industrialisation et l'inférence à grande échelle des modèles open-source ou propriétaires, en les adaptant à des usages spécifiques.** Leur rôle est également crucial pour fournir des infrastructures d'inférence dédiées aux besoins métiers de communautés variées, tout en répondant à des exigences de performance et de confidentialité. **Une attention particulière pourrait être portée aux petits modèles de langages distillés dits « SLM », qui sont plus légers et optimisés que les grands modèles, et permettent de gagner en efficacité sur l'inférence. Sur ce type de modèles, ce sont les modalités de sélection des données de base et leur préparation qui feront toute la différence**.

Recommandation 2.2 : mettre en place une gouvernance efficace qui incite des acteurs comme Outscale (Dassault Systèmes), Scaleway (Iliad) ou OVH Cloud à investir dans des machines exascale dédiées aux usages commerciaux, prenant le relais des infrastructures publiques axées sur la recherche. L'objectif central est de faire émerger des acteurs français capables d'exporter une capacité de calcul suffisante pour être compétitif. Les comités stratégiques de filière (CSF) sont des outils intéressants et ayant une vocation éminemment stratégique, mais ils peinent à se déployer et souffrent d'une lenteur d'exécution pénalisante faute de bénéficier de l'intérêt politique nécessaire.

Recommandation 3

Construire une réelle planification étatique en matière d'approvisionnement électrique pour mailler le territoire français en *data centers* de grande capacité en anticipant les usages futurs.

Aujourd'hui, les projets de construction de *data centers* relèvent d'une dynamique propre aux acteurs privés et de raccordements effectués dans l'ordre des demandes, sans hiérarchisation (premier arrivé, premier servi). **Une planification souple est nécessaire qui s'intéresse à la fois à la quantité de data centers nécessaires sur la base des besoins futurs et à l'ordonnancement des projets selon leur criticité et leur localisation**. Un des problèmes stratégiques majeurs est la mise en concurrence indistincte des projets hors de toute analyse des besoins territoriaux et des avantages concurrentiels. L'enjeu des raccordements au réseau électrique en est une manifestation criante. Il se double d'une

crainte liée à la rareté de l'électricité disponible, aux conflits d'accès et à la concurrence indue entre réindustrialisation, décarbonation, aménagement du territoire et transition numérique.

Recommandation 3.1 : communiquer au plus haut niveau de l'État pour démontrer la compatibilité entre disponibilité énergétique et infrastructures numériques. L'État doit porter un discours clair et stratégique sur l'adéquation entre les besoins numériques, industriels et de décarbonation, faisant valoir l'excédent de production électrique français. **Une initiative conjointe entre acteurs publics (RTE, CNDP) et privés (opérateurs de *data centers*, industriels) pourrait inclure une campagne de communication mettant en avant les preuves terrain (exemples de sites opérationnels redynamisant l'économie dans le respect des engagements de trajectoire climat)**.

Recommandation 3.2 : planifier et piloter le raccordement électrique des *data centers* – ceux de grande capacité comme ceux de proximité – à l'échelle nationale. L'État doit engager une planification systémique, intégrant toutes les parties prenantes (RTE, collectivités, acteurs économiques), pour identifier les besoins réels de la population. **Cela passe par une cartographie nationale co-construite avec RTE, l'État et les collectivités, localisant des « zones stratégiques » pour des infrastructures adaptées à des usages à latence faible ou forte**. Cette cartographie devra inclure les infrastructures modulaires (« sans murs ») pour les usages IA-HPC et se fonder sur quatre critères prioritaires :

le différentiel de coût avec de l'infrastructure fixe (fonction du niveau de flexibilité attendu) la pertinence énergétique (proximité des sources), le potentiel économique (emplois locaux), et l'impact social (développement des territoires).

À plus long terme, une fois l'infrastructure opérante, une réflexion devra être initiée

- **Sur la nature des offres commerciales et la priorisation des projets et des demandes de raccordement afin d'éviter l'engorgement de projets « zombis » ou d'usages à très faible valeur ajoutée**. Cette réflexion doit aussi permettre de réduire les délais de fabrication des transformateurs de puissance qui s'étendent constamment dans le secteur des utilities.

- **Sur la manière de valoriser la capacité du parc nucléaire français à transférer efficacement la puissance électrique entre différentes zones**, un domaine où la France bénéficie d'un avantage compétitif grâce à son réseau haute tension maillé et à l'expertise de RTE dans l'optimisation des flux énergétiques à grande échelle.

Recommandation 3.3 : capitaliser sur l'avantage géographique de la France dans les négociations internationales et les partenariats stratégiques. La position géographique de la France, au carrefour des axes nord-sud en Europe et bénéficiant d'une connectivité privilégiée avec les États-Unis, constitue un atout majeur à valoriser. De plus, le territoire français offre des caractéristiques naturelles uniques pour le développement d'infrastructures numériques durables. Par exemple, les régions montagneuses, avec leur potentiel

hydroélectrique, permettent d'alimenter des *data centers* en énergie renouvelable. Implantés dans des vallées où les températures plus basses contribuent au refroidissement, ces centres pourraient réduire significativement leur consommation énergétique liée au refroidissement, qui représente une part importante de leur empreinte carbone.

Recommandation 4

Capitaliser sur le lancement des 35 sites clés en main pour raccourcir les délais de construction de *data centers* dont l'intérêt économique et social est démontré en simplifiant les procédures administratives.

Les délais administratifs liés aux permis de construire et à la sortie de terre des *data centers* de grande capacité sont la principale raison de notre manque d'attractivité et nous privent de près de 100 Mds € d'investissements effectués par défaut dans des pays limitrophes, alors que notre électricité décarbonée devrait être un atout recherché.

Recommandation 4.1 : le projet de loi Simplification de la Vie Économique et les dispositions de type *fast track* dans la loi Industrie Verte en cours d'examen doivent doter d'un statut de projet d'intérêt national majeur les *data centers* de grande capacité et de proximité. Pour assurer l'efficacité de ce statut, certaines dispositions administratives doivent être intégrées au projet de loi qui veilleront à

accélérer les procédures sans sacrifier la rigueur de sélection des projets les plus qualitatifs.

- **Regrouper les services environnement et urbanisme** au sein des DREAL pour accélérer le traitement des demandes administratives.

- L'un des critères retenus pour l'obtention du permis de construire d'un *data center* est aujourd'hui celui de la chaleur fatale dissipée dans l'environnement sans réutilisation. **Ce critère n'est pas toujours pertinent, bien qu'il soit important dans certains cas, et doit être remplacé par le ratio PUE (Power Usage Effectiveness) / WUE (Water Usage Effectiveness) afin de prendre en compte l'efficacité des techniques de refroidissement comme critère d'impact environnemental**.

- **Confier au Conseil d'État le traitement en premier et dernier ressort des recours contre l'implantation de *data centers* dans un territoire après l'octroi des permis, sur le modèle de l'éolien offshore**[211]. En effet, les procédures de recours auprès des différents tribunaux peuvent prolonger de 7 ans les délais de construction des *data centers* une fois le permis de construire obtenu, ce qui représente une incertitude économique et industrielle insupportable et dissuade les investisseurs.

Recommandation 4.2 : renforcer l'acceptabilité sociale des projets de *data centers* situés dans des agglomérations de taille moyenne grâce à un « récit territorial » puissant.

[211] *Rapport du Conseil Général de l'Environnement et du Développement Durable, « La simplification des procédures d'autorisation applicables aux éoliennes en mer », juin 2021.*

Pour cela, inciter fiscalement par le biais de subvention à la construction de « centres d'excellence technologiques » à proximité des nouveaux *data centers*. Pour les *data centers* assortis à ce type de centres, expérimenter une dérogation ciblée au principe de zéro artificialisation nette (ZAN) dans les communes peu urbanisées qui souhaiteraient accueillir ce type de projet pour favoriser leur développement économique.

Recommandation 5

Lancer un projet « commando » pour développer des formations continues rapprochant les métiers de l'infrastructure numérique de réseaux de ceux de l'infrastructure numérique de traitement de données.

Recommandation 5.1 : créer des centres pluridisciplinaires inspirés du modèle IMT pour rapprocher les formations aux métiers du réseau de celles dédiées au traitement des données, en favorisant l'émergence de profils d'« ingénieurs maisons ». Ces centres permettraient de créer des passerelles entre formation initiale et continue, stimulant ainsi la mobilité entre filières et orientant les talents vers des projets industriels concrets.

Par exemple, il faut repenser la formation des équipes des centres de calcul qui accompagnent les entreprises utilisatrices. Aujourd'hui majoritairement axées sur des dimensions techniques ou généralistes, ces

formations doivent intégrer une approche métier et usage, ou élargir les recrutements à des profils plus sectoriels.

Recommandation 5.2 : Valoriser les engagements préalables entre l'entreprise et le salarié pourrait être envisagé pour assurer un minimum de fidélité. En effet, cela pose le problème de l'investissement privé pour la formation de ces ingénieurs qui seront par la suite fortement sollicités par les entreprises étrangères avec des moyens d'attractivité salariale nettement supérieurs aux nôtres.

Objectif n° 3

Maintenir et exporter l'excellence des infrastructures numériques de réseau françaises

Recommandation 6

Accélérer le déploiement de la 5G en milieu industriel, au moins sur les projets *greenfield*, en ciblant résolument les besoins des entreprises utilisatrices (TPE-PME-ETI).

La mise à l'échelle des infrastructures numériques est une réalité, avec la 5G appelée à remplacer progressivement la 4G. Si cette transition semble assurée à court et moyen terme, elle reste incertaine sur le long terme. En effet, les entreprises seront appelées à faire des choix et des investissements sur des déploiements encore coûteux mais dont la

valeur reste à démontrer : 5G versus Wifi, 6G et OpenRAN par exemple. Cet inconfort ne doit pas empêcher la préparation du tissu économique à l'adoption généralisée de la 5G. Très classiquement, le déploiement privé suit le déploiement public et nous y sommes pour les usages quotidiens (JO de Paris, transport, vidéosurveillance…). **Il est essentiel de démocratiser l'accès à la 5G pour toutes les entreprises et zones géographiques (stades, usines, ports, zones rurales), tout en se préparant à la 6G, qui mobilise dores et déjà la Californie et nous invite à une réflexion de suiveur agile ou de parieur technologique**.

Recommandation 6.1 : recentrer les 735 M€ alloués à la mission 5G Industrielle dans le cadre de France 2030 autour de deux objectifs clés.

D'une part, l'accélération de la commercialisation de fonctionnalités 5G avancées – comme le *slicing* – à moindre coût. Cela nécessite de mettre en œuvre un cadre structuré de subventions publiques pour soutenir les entreprises devant adapter leurs outils de production à ces nouvelles technologies, et ainsi réduire leurs coûts d'adoption. Ces aides financières doivent cibler à la fois les grandes entreprises, qui peuvent servir de locomotives, et les ETI-PME innovantes, qui jouent un rôle clé dans les chaînes de valeur industrielles. Elles doivent être fournies en impliquant les collectivités locales en leur donnant pour instruction de privilégier la commande publique de fonctionnalités 5G avancées afin de renforcer la cohérence du dispositif. Pour cela, il serait opportun de développer des projets pilotes en 5G industrielle, en collaboration avec des industriels régionaux, pour démontrer la valeur ajoutée de la 5G dans des environnements à forte valeur ajoutée (usines, ports, agriculture de précision).

D'autre part, l'utilisation de réseaux privés comme *enablers* des usages identifiés dans le cadre du continuum souverain

cloud-edge-réseaux IoT. Les réseaux privés 5G sont aujourd'hui la seule solution technologique capable d'assurer une transmission sécurisée et adaptée à la criticité des données, tout en garantissant une intégration fluide et cohérente entre les différents canaux du continuum.

Recommandation 6.2 : mettre en place des partenariats publics privés à l'échelle européenne avec les pays à la pointe en matière de connectivité réseau, tels que la Corée du Sud, le Japon, l'Inde et Singapour pour faire émerger des offres compétitives de solution de type PaaS et les intégrer aux initiatives Open-RAN. Cela permettrait de soutenir l'émergence d'un écosystème européen capable de capter la valeur effective de la 5G et des générations suivantes et de ne pas se cantonner à la fabrication d'antennes. Cela favoriserait aussi la diffusion de solutions *open source* de qualité pour encourager le développement de sociétés européennes spécialisées dans l'*open source* communautaire.

Recommandation 6.3 : intégrer à l'agenda de la prochaine Commission européenne une consultation approfondie des industriels et des experts sur les impacts du nouveau système de redevances sur les brevets essentiels aux normes (BEN), afin d'éviter des freins à l'innovation et de la calquer sur les besoins du terrain.

Recommandation 7

Sécuriser les nœuds critiques de distribution des câbles terrestres par une politique d'enfouissement raisonnée des câbles terrestres et aériens.

Les câbles sont devenus des équipements vitaux et vulnérables, de plus en plus soumis aux menaces de sabotages et aux aléas climatiques. Il est donc nécessaire de les sécuriser afin de sécuriser le réseau. Le coût estimé d'une telle sécurisation se situe entre 6 à 17 Mds € ce qui représente un montant difficilement accessible compte tenu de nos marges de manœuvre budgétaires. Le choix est donc de proposer une sécurisation prioritaire des réseaux de collecte de données les plus critiques dits NRO.

Pour y parvenir, intégrer dans le plan national d'adaptation au changement climatique (PNACC) une politique d'enfouissement raisonnée à partir d'une identification précise des réseaux de collecte des données et des risques sur les nœuds critiques de distribution.

Recommandation 7.1. : concentrer les efforts financiers sur le réseau de collecte des données (NRO) pour enfouir les points névralgiques des réseaux de communications électroniques. Face aux contraintes budgétaires, il serait inefficace de tenter simultanément l'enfouissement du réseau de distribution, qui dessert un nombre plus restreint d'utilisateurs et est plus rapide à réparer. En priorisant le réseau de collecte, notamment dans les zones vulnérables identifiées (forêts, zones à risque d'inondations), il est possible d'obtenir un impact immédiat et significatif.

Recommandation 7.2 : inclure les menaces liées aux vents violents et aux incendies dans la hiérarchie des risques climatiques identifiée dans le PNACC (en mesure 32). Une approche annuelle est indispensable, avec des expérimentations dès les périodes estivales, où les infrastructures numériques sont exposées aux incendies et à la chaleur extrême. Aligner cette mesure avec la mesure 7 du PNACC dédiée aux incendies permettrait une gestion des risques plus cohérente.

Recommandation 7.3 : intégrer dans le PNACC l'utilisation des technologies avancées de surveillance pour anticiper les nouveaux risques climatiques (mesure 19). Ce type d'outils est susceptible de contribuer à l'adoption du *edge computing* sur le territoire français (conformément à la recommandation 1).

Veiller à associer les collectivités territoriales et autorités déconcentrées compétentes, ainsi que les parties prenantes du génie civil pour garantir la faisabilité de ces mesures, et à bien coordonner ces mesures avec les travaux publics.

Recommandation 8

Valoriser l'atout stratégique que représentent les câbles sous-marins français via une stratégie intégrée combinant surveillance renforcée, investissements ciblés en Outre-mer et influence accrue dans les instances internationales.

La résilience des câbles sous-marins français est menacée par des actes de sabotage terroriste et militaire de plus en plus nombreux et sophistiqués. La France a un avantage compétitif majeur dans le secteur grâce à ses entreprises (ASN, Nexans et Orange Submarine Cable) qui fabriquent des câbles haute performance et nous assurent une maîtrise de bout en bout de notre infrastructure sous-marine (conception, fabrication, pose et maintenance). Pour maintenir cet avantage, dans le contexte géopolitique tendu que nous connaissons, deux priorités :

Recommandation 8.1 : investir, *via* France 2030, pour renforcer la résilience des câbles sous-marins en se concentrant sur les deux éléments ci-dessous.

1. **Les technologies de surveillance et de gestion innovantes, comme le *fiber sensing*, qui détecte les variations de tension et de température pour une maintenance préventive efficace**. Compléter ces solutions par des systèmes d'IA capables de différencier anomalies et menaces réelles, afin de détecter de manière proactive les risques de rupture ou de sabotage et d'optimiser les interventions sous-marines.

2. **Passer commande à des sociétés de satellites géostationnaires européennes, de type Eutelsat, pour sécuriser à moindre coût les câbles situés dans les territoires ultra-marins, afin de garantir redondance et fiabilité du système dans son ensemble**. Ces « back-ups spatiaux » auraient pour objectif de pallier les interruptions des câbles sous-marins, en cas de sabotage ou de défaillance des réseaux terrestres, pour renforcer leur sécurité physique.

Recommandation 8.2 : renforcer la participation des autorités publiques et des industriels français au sein des organismes stratégiques tels que l'UIT (Union Internationale des Télécommunications), l'ICPC *(International Cable Protection Committee)*, les groupes de travail de l'OCDE et la commission CF2 de la Commission européenne. À l'instar du modèle portugais, tirer parti de cette présence pour développer des partenariats avec les consortiums internationaux gestionnaires de câbles pour positionner la France comme un acteur central et incontournable dans l'installation, la gestion et la sécurisation des câbles sous-marins.

Objectif n° 4

Se doter des moyens réglementaires nécessaires à une compétition non biaisée au niveau européen face à l'agressivité mondiale

Recommandation 9

Adapter et simplifier le cadre normatif européen pour renforcer notre compétitivité et favoriser les consolidations d'acteurs à l'échelle mondiale.

Recommandation 9.1 : simplifier les règles administratives européennes en tirant parti de l'instauration d'un 28e régime européen pour alléger les contraintes administratives pesant sur les entreprises européennes. En particulier, tirer les conséquences des enseignements de la *Corporate Sustainability Reporting Directive* (CSRD) pour s'assurer que nos engagements en matière de développement durable et de démocratie se traduisent en avantages concurrentiels sur la scène mondiale, plutôt qu'en obstacles supplémentaires pour nos entreprises.

Recommandation 9.2 : le plus rapidement possible introduire des critères stratégiques dans les achats publics pour favoriser les entreprises européennes. Pour cela, mettre en place des critères de cybersécurité et de limitation des dépendances à des fournisseurs tiers, sur le modèle du règlement *Net Zero Industry Act* (NZIA). Ces critères doivent imposer aux acteurs étrangers des coûts de mise en conformité proportionnels aux risques de *dumping* ou d'ingérence économique, tout en incitant à une diversification des chaînes d'approvisionnement. Ils pourraient par exemple inclure des exigences d'approvisionnement minimum en composants d'infrastructures numériques européens à haute efficacité énergétique (dits frugaux). Cette approche permettrait d'atteindre les objectifs d'un *Buy European Tech Act* sans remettre en cause nos engagements internationaux.

Recommandation 9.3 : renforcer la sécurité et la résilience des infrastructures critiques grâce aux appels d'offres européens. Pour y parvenir, accélérer l'intégration de critères favorisant les acteurs et composants européens dans les appels d'offres, en mettant l'accent sur la sécurité avancée et la transparence des infrastructures numériques critiques. Pour garantir l'effectivité du *Cyber Resilience Act*, les marchés publics européens doivent mieux valoriser les outils d'évaluation des vulnérabilités dès la conception des logiciels et équipements. En complément, les labels attribués aux fournisseurs de services pour Opérateurs d'Importance Vitale (OIV) doivent inclure des critères de solidité financière et de viabilité à long terme, afin d'assurer leur capacité à répondre durablement aux exigences des infrastructures critiques.

Transport des données

LES RÉSEAUX FILAIRES LOCAUX

- **Cuivre :** réseau déployé dès les années 1970 pour la téléphonie fixe, utilisé pour l'accès internet (ADSL, SDSL, VDSL) offrant un débit inférieur à 30 Mb/s. Fin du réseau prévue pour 2030.

- **Fibre optique :** fil en verre ou plastique conduisant la lumière pour transmettre des données avec un débit moyen de 1 Gb/s, très résilient sur de longues distances.

- ***Content Delivery Network*** **(CDN) :** réseau de serveurs distribués stockant temporairement des contenus pour réduire les temps de chargement grâce à la mise en cache.

- **Serveur :** ordinateur ou système partageant des ressources ou services avec des machines clientes (web, messagerie, virtuels).

- **Réseaux de communication quantique :** nœuds utilisant l'intrication pour transmettre des bits quantiques et assurer la sécurité des données.

- **Intrication quantique :** phénomène reliant des particules partageant les mêmes états, utilisé pour sécuriser les réseaux quantiques.

- **Bits quantiques :** unités d'information pouvant exister dans plusieurs états simultanément, contrairement aux bits classiques.

- **Débit descendant :** vitesse de téléchargement des données reçues par l'utilisateur.

- **Débit vs latence vs bande de fréquence :** débit mesure la vitesse de connexion, latence indique le délai de transfert, et bande concerne les fréquences utilisées.

- **Technologie FTTH :** fibre optique directement connectée au domicile de l'abonné.

- ***New Deal Mobile* :** plan de 2018 pour généraliser la 4G, couvrant 99 % de la population en 2023.

- **5G Standalone :** prévue dès 2024, nécessitant de nouveaux équipements, avec investissements conséquents (ex. vente de pylônes Iliad en 2019).

- ***Network slicing* :** division d'un réseau en segments virtuels pour des usages variés sur une même infrastructure.

LES RÉSEAUX DE TRANSIT INTERNATIONAUX

- **Câbles sous-marins :** installés au fond des océans, ils assurent la majorité des télécommunications mondiales, avec plus de 400 câbles en 2020.

- ***Internet Exchange Point* (IXP) :** infrastructure permettant aux fournisseurs d'échanger du trafic Internet *via* des accords de peering.

LES RÉSEAUX SANS FIL

- **WiFi 6/7 :** standard offrant un débit maximal de 10 Gb/s (+40 % par rapport au WiFi 5), meilleure portée et traversée des obstacles.

- **Réseau mobile :** antennes-relais transmettant les communications *via* ondes radio sur des cellules géographiques.

- **Ondes radio :** ondes électromagnétiques (<300 GHz) pour la transmission de signaux (voix, image, données).

- **Réseaux mobiles 3G à 6G :**
  - 3G : Premiers accès mobiles à Internet (2-42 Mb/s).
  - 4G : Déploiement dès 2016, débits moyens de 100 Mb/s.
  - 5G : Débits 10x supérieurs à la 4G, faible latence, supportant de nouveaux usages comme le *slicing*.

- **Réseaux LoRaWAN :** réseau longue portée et bas débit pour l'IoT, traversant facilement obstacles et zones confinées.

- **Réseaux NB-IoT :** protocole basse consommation dédié à l'IoT, utile pour des applications comme la télérelève des compteurs.

LES RÉSEAUX SATELLITE

- **Satellites GEO :** ciblent une zone fixe avec une orbite stable, idéaux pour les communications locales.

- **Satellites LEO :** faible latence et haute résolution pour les télécommunications et l'observation (ex. Starlink, OneWeb).

- **Satellites MEO :** alternative entre les GEO coûteux et les LEO saturés.

Traitement des données

MATÉRIEL

- ***Data center* :** bâtiment regroupant les infrastructures pour stocker et traiter des données, combinant composants informatiques, énergétiques et structurels pour un service optimal.

- **Semi-conducteurs :** composants essentiels des objets numériques (ordinateurs, smartphones, véhicules) et des chaînes industrielles (automobile, défense…). Ils incluent des CPU (calcul général), GPU (calcul parallèle), TPU (IA), LPU (analyses temps réel), et DPU (stockage). Fabriqués par des acteurs comme TSMC, Intel et Nvidia.

- **Puissance de calcul :** basée sur la logique, la mémoire et l'interconnexion entre les deux.

- **Puces :** mesurées en FLOP, elles incluent des CPU (50 % du coût pour la mémoire), GPU (calcul IA en parallèle), TPU *(machine learning)*, DPU (stockage), et LPU (langage). Fabricants : TSMC, Samsung ; designers : Nvidia, AMD, GAFAM.

- **FPGA :** circuits logiques programmables pour des applications comme la reconnaissance vocale ou le traitement de langage naturel.

- **ASICs :** puces optimisées pour des tâches spécifiques.

IMMATÉRIEL

- **Technologies d'infrastructure :** organisation en couches, depuis les composants physiques jusqu'aux applications, avec des services *cloud* (IaaS, PaaS, SaaS).

- **Logiciels *open-source* :** code accessible et modifiable (ex. Linux). Les logiciels libres ajoutent des garanties sur l'utilisation et la redistribution. Les logiciels propriétaires, comme Windows, restreignent l'accès et la personnalisation.

- **IaaS *(Infrastructure-as-a-Service)* :** externalisation de l'infrastructure physique, nécessitant des compétences pour gérer les systèmes et applications.

- **PaaS *(Platform-as-a-Service)* :** fournit une infrastructure pré-configurée et évolutive pour le développement d'applications.

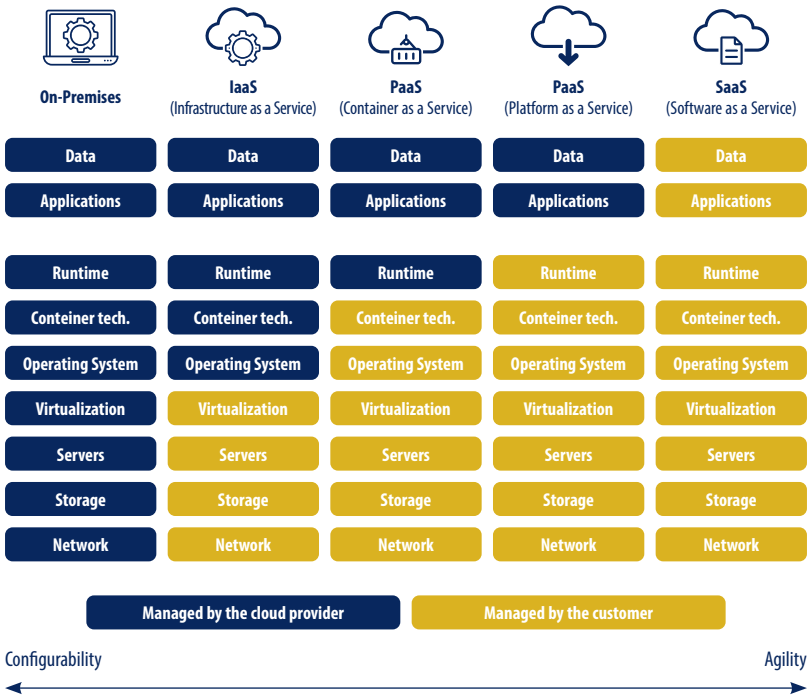
- **SaaS *(Software-as-a-Service)* :** solution logicielle clé en main accessible *via* abonnement, sans besoin de compétences techniques.

- **Conteneurs (CaaS) :** virtualisation des systèmes *via* Docker ou Kubernetes, rendant les applications portables et réduisant la dépendance aux fournisseurs *cloud*.

- **Systèmes d'exploitation (OS) :** logiciels gérant les ressources d'un ordinateur (Windows, MacOS, Linux).

- ***Lock-in Cloud* :** dépendance à un fournisseur *cloud* due aux briques propriétaires limitant la portabilité des systèmes, augmentant les coûts et la complexité.

• **Virtualisation :** technique permettant à des machines virtuelles d'exister sur une infrastructure partagée tout en restant isolées, réduisant les coûts et améliorant la confidentialité des données.



Accès aux données

• **Infrastructures d'accès :** leur robustesse et leur intégration avec les infrastructures de transport et traitement des données permettent une collecte en temps réel, offrant une connaissance fine de l'environnement.

- **Terminaux utilisateurs :** incluent ordinateurs, smartphones, tablettes, montres et objets connectés, casques, lunettes de réalité augmentée ou virtuelle, téléviseurs et robots.

- **Terminaux industriels :** automates (PLC), capteurs, objets connectés et robots, adaptés aux environnements de production.

- **Embarqué :** initialement développé pour des environnements hostiles (ex. nucléaire), l'embarqué s'est étendu à divers secteurs (automobile, santé, agriculture), facilitant l'automatisation et réduisant la pénibilité. Ex. : robotique chirurgicale en plein essor.

- ***Internet of Things*** **(IoT) :** connecte des équipements aux réseaux sans fil, permettant de développer de nouveaux services ou d'intégrer l'IoT dans la production. L'IoT a évolué depuis des systèmes en vase clos (comme les calculateurs automobiles) vers des réseaux ouverts, transformant les interactions humain-machine. Son omniprésence soulève des enjeux pour la vie privée et l'organisation du travail.

Annexe 1 : Besoins sous-tendus par nos infrastructures de réseaux

BESOIN 1 : SECTEUR DE L'AGRICULTURE

La compétitivité du secteur tient à l'exploitation en temps réel des données récoltées dans des zones difficiles à couvrir et au traçage et à l'analyse de ces données (niveau de pollution, bruit, température, humidité, etc.). Cela implique d'être en capacité de fournir des zones difficiles d'accès (IoT, satellite, GNSS), de compléter cette couverture par des outils en temps réel comme de la télédétection (drones et satellites) et d'optimiser les intrants dans le sol.

Exemple : mise à disposition d'infrastructures réseau (outils télématiques, récepteurs réseau mobile) pour faire de l'agriculture de précision[212] par John Deere.

BESOIN 2 : SECTEUR DE LA SANTÉ

La crise du covid a été un puissant déclencheur pour accélérer la numérisation du secteur, en particulier par la médecine de précision. L'efficacité de notre système de santé tient à sa capacité à prendre en charge rapidement les patients, et à les orienter rapidement vers les

[212] *Principe de gestion des terres agricoles qui vise l'optimisation des rendements et des investissements en tenant mieux compte de la variabilité des milieux et des conditions entre parcelles différentes à une échelle intra-parcellaire.*

bons professionnels. Pour répondre à cet objectif, une connectivité haut débit est nécessaire pour mieux connecter le bloc opératoire, fluidifier la gestion des urgences et développer la médecine à domicile pour désencombrer nos hôpitaux.

Exemple : mise à disposition de 5G dans le CHU de Tours pour suivre à distance les paramètres vitaux et la notification d'urgence, assurer un routage optimal de l'ambulance pour une prise en charge rapide, faire du télédiagnostic et s'assurer d'un acheminement optimal des ambulances.

BESOIN 3 :
SECTEUR AUTOMOBILE

La compétitivité du secteur tient à la capacité d'optimisation de la cadence pour produire efficacement de grandes séries de véhicules. Pour atteindre cet objectif, les acteurs doivent autonomiser de nombreuses tâches, dans des environnements métallisés, ce qui nécessite du très haut débit que le wifi ne peut pas fournir. Dans ce secteur, la 5G est un *enabler* pour transmettre des données à haut débit et en temps réel de manière sécurisée dans des environnements industriels métallisés. Et ceux, particulièrement dans le contexte de l'arrivée de flottes de véhicules autonomes.

Exemple : Usine intelligente 5G d'Ericsson au Texas ayant entraîné une amélioration de la productivité de 120 %. NB que ce besoin concerne aussi les accès, traités plus loin.

BESOIN 4 :
SECTEUR DE L'AÉROSPATIAL

Le contexte géopolitique de la guerre russo-ukrainienne a rappelé l'importance de l'accès au réseau pour les populations civiles ainsi que la forte dépendance de l'armée ukrainienne vis-à-vis des fournisseurs de réseau. Les fonctionnalités de communication par satellite permettent en effet d'envoyer des SOS d'urgence, d'accéder à de l'assistance routière dans des zones non couvertes par le réseau wifi/mobile, de partager sa position en cas d'urgence et ainsi de planifier des opérations militaires.

Exemple : les 6 000 satellites en orbite basse (LEO) de Starlink pour fournir un service internet à tout le monde qui reposent sur la capacité de réemploi de lanceurs spatiaux.

BESOIN 5 :
SECTEUR DES MÉDIAS ET DU DIVERTISSEMENT

Le *streaming* et la vidéo concentrent à eux seuls 70 %[213] de la bande passante, dont la proportion ne cesse d'augmenter (écrans plus grands, meilleure qualité, multi-écran…). Les jeux vidéo (62 %[214] des français y jouent) soulèvent quant à eux de forts enjeux de réduction de la latence, élément clef dans l'expérience de jeu, encore plus en VR.

Exemple : utilisation quasi-quotidienne de Netflix par une majorité de la population.

[213] *ARCEP, Baromètre de l'interconnexion de données en France, juillet 2023.*

[214] *Syndicat des Éditeurs de Logiciel et de Loisirs, l'essentiel du jeu vidéo, octobre 2023.*

BESOIN 6 :
SECTEUR DE L'ÉNERGIE

L'efficacité du secteur est désormais conditionnée à la capacité de transmission d'un nombre croissant de données acquises dans des milieux fortement normés et à haut niveau de sécurité. Cela nécessite de déployer des réseaux 5G -NR à fort débit et faible latence en milieu.

Exemple : déploiement de réseaux 5G NR et Lora 2,4 GHz chez CEA et Orano[215], pour prévenir d'éventuelles failles et des défaillances.
Exemple : jumeaux numériques de centrales nucléaires développés par EDF et Total Energies.

BESOIN 7 :
SECTEUR BANCAIRE

Les régulateurs européens, comme l'Autorité Bancaire Européenne (EBA) ou la Banque Centrale Européenne (BCE) imposent des contraintes strictes aux banques commerciales en matière de sécurité, de transparence et de résilience des systèmes financiers. Par exemple, la réglementation *Digital Operational Resilience Act* (DORA) en vigueur depuis janvier 2025 impose aux banques et assurances de renforcer leur résilience opérationnelle numérique, particulièrement en assurant la continuité de l'activité en cas de perturbations. Cela implique la mise en place d'infrastructures de réseau robustes et redondantes, capables de fonctionner malgré des incidents majeurs.

Exemple : BNP Paribas a développé une infrastructure de réseau spécifique avec des centres de données redondants et des systèmes de sauvegarde automatisés pour garantir la continuité de ses services financiers, conformément aux exigences de DORA.

[215] *Le Magit, avril 2018, « 2IDO et 5G NR : ces nouveaux réseaux IoT en approche », https://www.lemagit.fr/.*

Annexe 2 :
Besoins sous-tendus par nos infrastructures de traitement

BESOIN 1 :
SECTEUR DE LA SANTÉ

La compétitivité du secteur dépend de la capacité des acteurs à intégrer de manière fluide des innovations de pointe dans les systèmes d'information existants, à l'instar d'algorithmes d'IA ou de technologies quantiques et de stockage ADN pour démultiplier les calculs en parallèle et trouver de nouvelles molécules. Pour utiliser ces technologies, les établissements de santé sont toutefois souvent confrontés à des pics de charge inattendus et à des besoins de calcul supérieurs à la capacité des sites existants. Par sa flexibilité et sa *« scalabilité »*, le *cloud* apporte une réponse à ce besoin, et le *edge* semble prometteur. Le secteur est également fortement soumis à la réglementation sur les données personnelles[216].

Exemple : collaboration entre l'Institut Gustave Roussy, l'hôpital Necker, et l'Institut Curie pour améliorer la détection de tumeurs cérébrales rares avec une approche combinée de détection d'objets et de segmentation par du deep learning via le réseau U-Net. L'université Polytechnique de Montréal est également très active dans le domaine.

[216] *Décret 2018-137 du 26 février 2018 selon lequel toute entité opérant un système d'information de santé pour le compte d'un tiers doit être en conformité avec le référentiel de certification « Hébergeur de données de santé » (HDS).*

BESOIN 2 :
SECTEUR DE L'AÉROSPATIAL

La performance de nos aéronefs (avions, hélicoptères, drones) et astronefs (véhicules spatiaux, lanceurs, satellites), que ce soit pour des usages civils ou militaires, dépend de la capacité des acteurs à mettre en réseau différents territoires à différentes échelles (autorités, compagnies, détaillants, traiteurs, services de maintenance, organismes publics, etc.). Pour cela, le secteur a besoin d'analytique *cloud* intelligente, d'IA et de *machine learning* pour faire de nombreux calculs en parallèle, et d'une sécurité renforcée.

Exemple 1 industriel : utilisation par Airbus du cloud Azure pour affiner ses processus de conception, de construction et d'exploitation de produits complexes.
Exemple 2 de recherche fondamentale : simulation de la réionisation de l'univers[217] en exploitant 16 000 cartes GPU en parallèle grâce à un accès privilégié au GENCI français[218] et au PRACE européen[219].

BESOIN 3 :
SECTEUR DU RETAIL ET DU E-COMMERCE

Pour répondre aux préférences de consommation de la population, nos *retailers* et commerçants ont largement adopté l'omnicanal, soit le fait de proposer à leurs clients d'effectuer leurs achats sur différents canaux (boutique physique, site en ligne, voir métavers). Cela nécessite d'être capable de traiter efficacement de gros volumes de données produit et

[217] *Ministère de l'enseignement supérieur et de la recherche et de l'innovation, 2021, Stratégie nationale des infrastructures de recherche.*

[218] *Offre de HPC grâce au CNRS, au CEA et à l'enseignement supérieur (IDRIS, TGCC, CINES), avec une offre élargie au stockage de données computationnelles.*

[219] *Pyramide de calcul européenne regroupant 26 États membres, dont le GENCI avec 7 calculateurs de niveau tier 0, remplacée maintenant par EuroHPC.*

client, et donc des infrastructures de traitement de proximité *(data center/Edge)* et en réseaux avec une faible latence (chargement des images, calcul du panier et modes de transport…).

Exemple : unification des univers de vente et de marketplace du groupe Leclerc[220] en un site unique en utilisant Google Cloud Platform (GCP) pour faire face à des pics importants de consommation sans transiger sur la sécurité et la performance des systèmes.

BESOIN 4 : SECTEUR DE L'ÉNERGIE

Dans un contexte de limitation croissante de nos ressources énergétiques, les entreprises les plus compétitives dans le secteur sont aussi celles qui utilisent le mieux leur électricité. Cela nécessite des capacités importantes de simulation, de modélisation et de stockage sécurisé des données directement utiles pour mieux appréhender le système électrique compte tenu des moyens de production disponibles et des changements climatiques à venir.

Exemple : Supercalculateur d'EDF[221] utilisé à l'aval du compteur électrique pour faire des smartgrids qui ajustent en temps réel les flux d'électricité disponibles et nécessaires.

BESOIN 5 : SECTEUR FINANCIER

Près de 69 % des entreprises françaises déclarent avoir subi au moins une tentative de fraude en 2022[222]. Dans ce contexte, l'intelligence

[220] *Hunik groupe, Groupe Leclerc.*

[221] *EDF, les Supercalculateurs.*

[222] *Allianz Trade, Etude fraude 2022.*

artificielle permet de capturer plus finement des informations en temps réel pour détecter des fraudes, optimiser des rendements, prédire des résultats ou améliorer l'expérience client.

Exemple : des entreprises comme Mastercard, Stripe[223] ou encore PayPal[224] utilisent le machine learning, soit le sous-domaine de l'IA qui confère la capacité aux ordinateurs d'apprendre à partir de jeux de données structurés ou non, pour détecter et prévenir la fraude en ligne. L'apprentissage supervisé permet de détecter des modèles de fraude établis (anomalies, cotation de risques, analyses réseaux, texte, identité), tandis que l'apprentissage non supervisé permet de détecter de nouvelles techniques de fraude en temps réel.

BESOIN 6 :
SECTEUR DE L'AUTOMOBILE

Dans ce secteur, des besoins croissants de gestion de l'urbanisation croissante (surcharge des infrastructures, congestion du trafic, demande accrue en services et problèmes environnementaux) apparaissent.

Exemple : Les États-Unis[225] ont développé des systèmes de transport intelligent (STI) qui intègrent la communication véhicule-tout (V2X) pour optimiser la gestion du trafic en temps réel. Cette approche, basée sur le edge computing, vise à réduire les latences et protéger les données tout en améliorant l'efficacité des transports urbains.

Parmi ces besoins, ceux qui font intervenir de la résolution de problèmes complexes (dits non-polynomiaux[226]) ou dont la complexité

[223] *Stripe, juin 2023, « Exploitation du machine learning dans le cadre de la détection et de la prévention de la fraude aux paiements ».*

[224] *Intel, « Déployer l'IA au sein des services financiers ».*

[225] *Cambridge Consultants & ITS America, 2023, Digital Infrastructure Strategy Report Shaping the future of transportation in the United States.*

croît de manière exponentielle (dit non-polynomiaux complets[227]), peuvent être enrichis par de la mécanique quantique, qui permet de traiter simultanément toutes les réponses possibles à un problème donné.

Annexe 3 : Besoins sous-tendus par les couches logicielles

1. LES BESOINS

BESOIN 1 : FACILITER LE DÉVELOPPEMENT ET LE TEST D'APPLICATIONS

La technologie de conteneur informatique permet une virtualisation des systèmes d'exécution, et offre un espace dédié à une application ou un logiciel. Il intègre le code et les fichiers nécessaires au bon fonctionnement de l'application. Cette technologie répond aux besoins des développeurs de créer des environnements de développement cohérents, reproductibles et jetables. Par exemple, la solution *Docker* permet de conteneuriser une application web avec toutes ses dépendances (serveur web, base de données, bibliothèques) et garantit qu'elle fonctionne de la même manière sur les machines des développeurs (conception), les serveurs de test, et une fois en phase d'utilisation.

[226] *Par exemple, le fait de vérifier qu'une séquence d'ADN se retrouve bien dans plusieurs gènes.*

[227] *Par exemple, le fait de remplir son sac ou son coffre de manière optimale pour respecter une contrainte de temps, sans avoir à tester toutes les solutions possibles.*

BESOIN 2 :
ADOPTER UNE APPROCHE MICROSERVICES

Les architectures de microservices ont recours à des logiciels de conteneurisation pour décomposer une application en services indépendants, chacun exécuté dans son propre conteneur. Par exemple, une application de commerce électronique peut être divisée en plusieurs microservices : gestion des utilisateurs, traitement des paiements, gestion des stocks, etc. Chaque microservice peut être développé, déployé et mis à jour indépendamment, ce qui améliore la flexibilité et la scalabilité de l'application.

BESOIN 3 :
OPTIMISER LES RESSOURCES EN UTILISANT DES LOGICIELS DE VIRTUALISATION

Les logiciels de virtualisation permettent de créer des versions virtuelles de ressources informatiques, telles que des serveurs, des systèmes d'exploitation ou des réseaux. Ils permettent d'exécuter plusieurs environnements isolés sur une seule machine physique. Cela optimise l'utilisation des ressources et améliore la flexibilité et la gestion des infrastructures informatiques. De telles solutions permettent ainsi de rationaliser le nombre d'équipements.

BESOIN 4 :
SUIVRE L'ÉTAT DE SANTÉ DE L'INFRASTRUCTURE INFORMATIQUE

L'état de santé du parc est aujourd'hui suivi par des logiciels dits de supervision, ou *monitoring* en anglais. Ils permettent de surveiller et gérer les performances, la disponibilité et la santé des systèmes informatiques, des réseaux et des applications, et ainsi d'en détecter les

dysfonctionnements puis d'alerter les personnes compétentes en vue d'activer les actions correctives nécessaires.

Dans le domaine de l'IoT, les outils de supervision sont cruciaux, car ils permettent de récolter l'intégralité des données des capteurs connectés, puis de les exploiter en vue d'optimiser les opérations et la résilience du site. De nombreux cas d'applications existent dans l'industrie, par exemple dans le secteur du nucléaire, et l'optimisation du suivi des déchets radioactifs et de paramètres (consommation, activité).

2. LES TECHNOLOGIES

TECHNOLOGIE 1 : SYSTÈMES D'EXPLOITATION (SE OU O/S)

Les SE permettent de gérer les ressources matérielles des serveurs et des terminaux, et d'exécuter les applications. Ils doivent garantir la stabilité, la sécurité et la compatibilité des environnements informatiques tout en supportant une multitude d'applications et de services. Aujourd'hui, cela nécessite d'intégrer des fonctionnalités avancées de sécurité, de gestion de données et de virtualisation. Plus que des gestionnaires de ressources matérielles, ce sont désormais des plateformes complexes, qui doivent s'intégrer de manière transparente avec les infrastructures *cloud*, gérer des conteneurs et des microservices, et des mécanismes de sécurité renforcée contre des menaces toujours plus sophistiquées.

Exemple : Windows Server 2022 pour les environnements d'entreprise, offrant des fonctionnalités avancées telles que la gestion des conteneurs avec Docker[228], la sécurité renforcée avec des protections contre les menaces de ransomware, et une intégration fluide avec les services de cloud hybride.

[228] *Approche de virtualisation qui permet d'exécuter des applications dans des environnements isolés appelés conteneurs.*

Encadré n° 27 • Focus sur la virtualisation

La virtualisation permet de créer des versions virtuelles de ressources physiques telles que des serveurs, des systèmes de stockage et des réseaux. Introduite au début des années 1970, elle a été popularisée dans les années 2000 avec des entreprises comme VMware et Microsoft. Elle repose sur un logiciel appelé hyperviseur[229], qui s'installe sur un serveur physique (appelé hôte) et permet de créer plusieurs environnements virtuels appelés machines virtuelles (VM).

Chaque VM fonctionne comme un serveur indépendant avec son propre système d'exploitation et ses propres applications, malgré le fait qu'elle partage les mêmes ressources matérielles que les autres VMs sur le même hôte. La virtualisation permet de maximiser l'utilisation des ressources matérielles en exécutant plusieurs machines virtuelles sur un même serveur physique, réduisant ainsi les coûts liés à l'achat de matériel. Elle facilite aussi la gestion des infrastructures en permettant la création, le déploiement, et la gestion des VM de manière dynamique. Les VM peuvent être déplacées entre différents hôtes physiques sans interruption de service, ce qui simplifie la maintenance et les mises à jour. Les VM sont isolées les unes des autres, ce qui améliore la sécurité en limitant l'impact des défaillances ou des attaques sur une VM sur les autres VM. La virtualisation a évolué pour répondre aux exigences croissantes de flexibilité et d'efficacité des ressources. Les premières solutions de virtualisation se sont concentrées sur l'optimisation des ressources matérielles, tandis que les besoins

[229] *Il existe deux types d'hyperviseurs : le type 1, ou bare-metal, qui s'exécute directement sur le matériel physique de l'hôte, offrant une performance optimisée et une isolation accrue. Exemple : VMware ESXi, Microsoft Hyper-V; et le type 2, ou hosted, qui s'exécute sur un système d'exploitation hôte, ce qui peut introduire une certaine surcharge en termes de performance. Exemple : VMware Workstation, Oracle VirtualBox.*

actuels incluent la gestion de ressources multi-*cloud*, la séparation des environnements de développement et de production, et la mise en place d'infrastructures virtualisées pour les applications modernes. Les solutions de virtualisation doivent maintenant supporter une gestion intégrée et une orchestration des ressources à travers des environnements complexes.

La création de nouvelles instances virtuelles peut se faire en quelques minutes, permettant une réponse rapide aux besoins fluctuants en ressources. La virtualisation continue d'évoluer avec l'émergence de technologies complémentaires comme la **virtualisation de conteneurs** (par exemple, Docker et Kubernetes). Ces technologies permettent de créer des environnements d'exécution isolés au niveau des applications plutôt qu'au niveau des systèmes d'exploitation pour davantage de flexibilité et de performance dans l'utilisation des ressources.

Cependant, la position dominante des leaders du secteur rend les clients captifs ce qui peut, en cas de changement de stratégie par l'entreprise, créer d'énormes contraintes pour les utilisateurs. Ainsi, VMware vSphere, leader de la virtualisation des serveurs, qui permet une gestion centralisée et une optimisation des ressources dans les environnements de *data centers* et *cloud* hybrides, a changé de stratégie suite à son rachat par Broadcom, pénalisant de nombreux clients en imposant par exemple la vente forcée de l'ensemble de leurs suites. Un tel scénario est envisageable pour les *hyperscalers*. Dans certains secteurs des obligations existent déjà, par exemple le secteur bancaire aux États-Unis, de ne pas être exclusif envers un fournisseur, mais d'être multi *cloud*, ce qui évidemment alourdit les complexités des architectures, la difficulté d'avoir des ressources formées et le coût global de la migration vers le *cloud* public.

TECHNOLOGIE 2 : SUPERVISION

La supervision des infrastructures informatiques est cruciale pour assurer le bon fonctionnement et la disponibilité des services. Un prérequis est de la doter de capacités de surveillance en temps réel pour réagir rapidement aux incidents techniques. Les outils de supervision doivent ainsi offrir une vue centralisée de l'état des systèmes. Au départ, la supervision se concentrait sur la détection des pannes matérielles et des erreurs logicielles. Aujourd'hui, les besoins en supervision incluent la surveillance en temps réel des performances des systèmes, la gestion proactive des incidents, et l'analyse prédictive pour anticiper les problèmes avant qu'ils n'affectent les opérations.

Exemple : Prometheus permet de surveiller les performances des applications en temps réel et de détecter les anomalies dans des environnements de cloud natifs[230], permettant une gestion proactive des incidents.

TECHNOLOGIE 3 : ORCHESTRATION

L'orchestration des ressources et des services informatiques est nécessaire pour automatiser les déploiements, la gestion des configurations et la mise à l'échelle des applications. Elle a évolué pour répondre aux défis d'environnements de plus en plus complexes et distribués. Alors que les premières solutions se concentrent sur l'automatisation des tâches manuelles, les besoins actuels en orchestration incluent la gestion des déploiements, la mise à l'échelle des applications, et l'intégration fluide

[230] *Approche de conception et de déploiement des applications qui est conçue spécifiquement pour fonctionner de manière optimale dans un environnement cloud, en utilisant les services et les fonctionnalités proposés par les fournisseurs de cloud, comme les architectures microservice, la conteneurisation ou l'orchestration.*

des services dans des architectures de microservices. Les outils d'orchestration modernes permettent de coordonner les déploiements de conteneurs et de gérer des clusters à grande échelle de manière automatisée et efficace.

Exemple : Kubernetes pour orchestrer les conteneurs dans des environnements de cloud, permettant de gérer automatiquement le déploiement, la mise à l'échelle et la gestion des services dans des environnements de production complexes.

TECHNOLOGIE 4 : CONFIGURATION

Les outils de configuration permettent d'automatiser la gestion des paramètres des systèmes, d'assurer la conformité aux politiques de sécurité et de faciliter le déploiement cohérent des mises à jour et des patches[231]. Auparavant, la configuration des systèmes était souvent manuelle et sujette à des erreurs. Aujourd'hui, les besoins se concentrent sur l'automatisation de la gestion des configurations, la garantie de la conformité aux politiques de sécurité, et l'efficacité du déploiement des mises à jour. Les outils de configuration modernes permettent une gestion cohérente et reproductible des systèmes à travers des environnements variés.

Exemple : Ansible pour automatiser la gestion des configurations et des déploiements dans des environnements multi-cloud, assurant une cohérence et une conformité accrue des configurations systèmes.

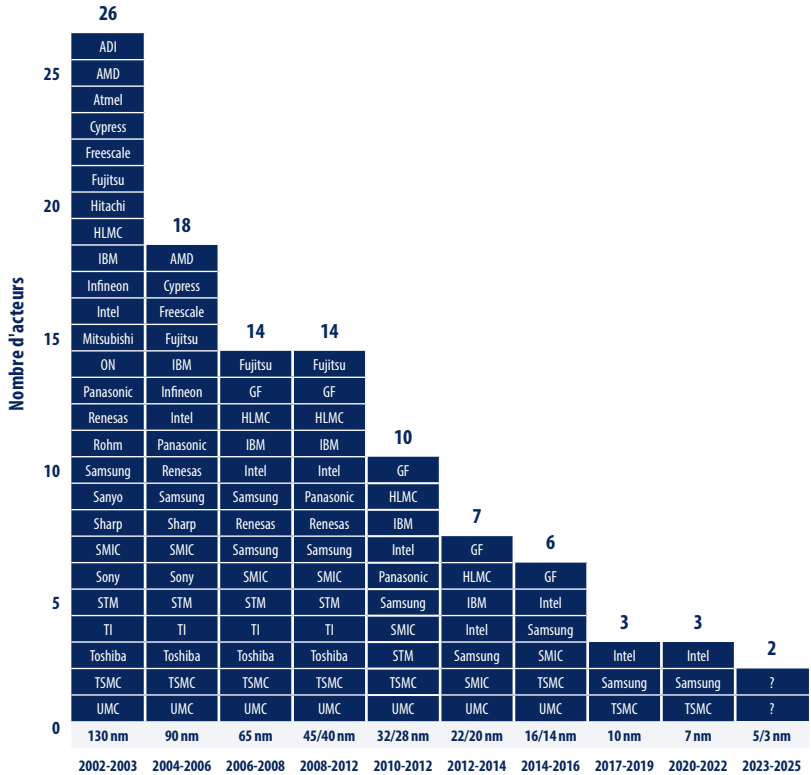
[231] *Section de code que l'on ajoute à un logiciel pour y apporter des modifications : correction d'un bug, traduction, crack.*

Annexe 4 :
Types de besoins par taille de puces

La taille du semi-conducteur, ou taille du nœud, indiquée en nanomètre (équivaut à un milliardième de mètre) est la grandeur caractéristique de la qualité des puces produites, permettant de réduire la place occupée et donc d'augmenter la densité et les performances des appareils équipés. Actuellement, les puces de 28 à 7 nm sont produites en masse, et certains acteurs leaders du marché peuvent produire des puces allant jusqu'à 3 nm, le saut vers les 2 nm étant prévu en 2025 pour TSMC, Samsung et Intel.

Graphique n° 45 • Évolution de l'industrie des semi-conducteurs

(Nombre d'acteurs disposant de capacités de production de pointe selon le nœud technologique et par année)



Note : il est à noter que les définitions d'un nœud sont variables d'un constructeur à l'autre, les puces de processeur de Samsung et TSMC définissant un nœud de 7 nm de la même manière qu'Intel définit son nœud de 10 nanomètres.

BESOIN 1 :
CALCULS EXIGEANTS ET IA – 2 NM

Pour équiper les derniers appareils mobiles de capacités de calculs capables de supporter l'IA en local, Apple serait prêt à exploiter les puces 2 nm de TSMC, le fondeur taïwanais, dès 2025[232].

BESOIN 2 :
RÉDUCTION DE LA CONSOMMATION D'ÉNERGIE – 3 À 5 NM

Le passage à une gravure de 3 nm permet de condenser les micro-processeurs en adoptant une architecture en millefeuille, et donc d'en décupler les performances. Les gains en termes de performances sont estimés à 45 %, et le passage aux 3 nm promet une consommation d'énergie divisée par quatre[233], soit quatre fois plus d'autonomie pour les appareils mobiles.

BESOIN 3 :
PERFORMANCES POUR LE DIVERTISSEMENT ET ÉQUIPEMENTS GRAND PUBLIC – 5 À 10 NM

Les appareils portables et les consoles sont massivement équipés de puces faisant partie des dernières générations. À titre d'exemple, la PS5 était équipée d'une puce AMD gravée en 7 nm gravée par TSMC jusqu'en 2022, les derniers modèles intégrant une puce de 6 nm.

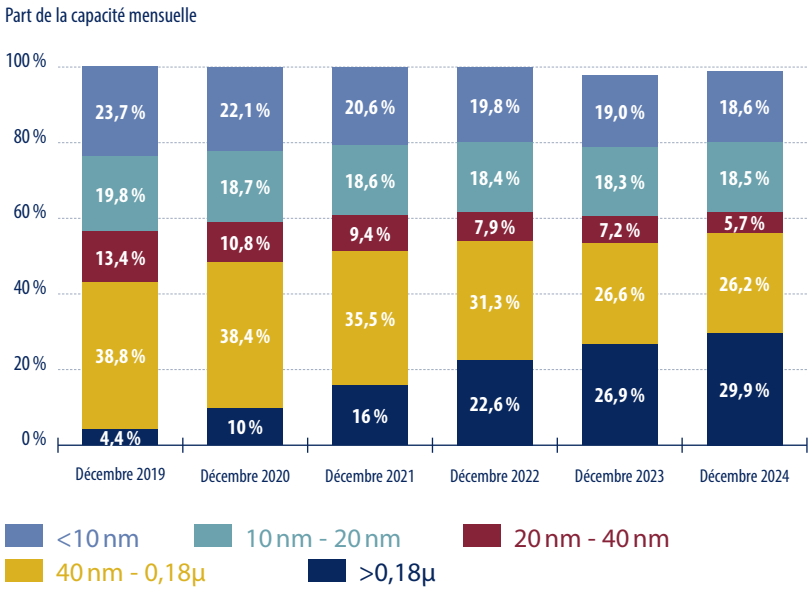
[232] *Les Numériques, « Avec presque deux ans d'avance, Apple s'affairerait déjà sur des puces 2 nm », mars 2024.*

[233] *Jean-Michel Sallese, « Les Nanotechnologies », Éditions Que sais-je, 2022.*

BESOIN 4 : COMMODITÉS SANS BESOINS DE PERFORMANCE PARTICULIERS – SUPÉRIEUR À 10 NM

Enfin, la majorité des cas d'usage ne nécessitent pas de puissance de calcul particulièrement élevée et peuvent avoir recours à des générations plus anciennes de puces. C'est notamment le cas de l'industrie et des capteurs utilisés dans l'IoT, et de la plupart des équipements domotiques d'anciennes générations.

Graphique n° 46 • Répartition mensuelle des capacités de production de plaques installées dans le monde
(de décembre 2019 à décembre 2024, par taille d'éléments)



Source : IC Insights.

Encadré n° 28 • Panorama des dernières innovations en matière de puces

- **Série Ryzen AI 300 de AMD**, conçue pour être intégrée aux nouveaux PC Copilot+ AI de Microsoft. Avec une puissance de 50 TOPS, ces puces auraient battu le M4 d'Apple et d'autres concurrents pour devenir les puces d'IA les plus puissantes pour les ordinateurs portables.

- **Puces Snapdragon X Elite de Qualcomm** pour les ordinateurs de bureau et les ordinateurs portables dotés d'un système d'IA. Il s'agit d'un changement important, car ses semi-conducteurs – qui reposent sur l'architecture de puce du concepteur britannique Arm – n'étaient jusqu'à présent utilisés que dans les appareils mobiles. Acer, Asus, Microsoft et d'autres sociétés prévoient de commercialiser prochainement des PC équipés de ces nouvelles puces.

- **Nouvelles versions d'Intel qui comprennent** les processeurs Xeon 6, qui rendent les centres de données plus rentables en divisant les charges de travail en deux catégories différentes – E-core pour l'efficacité, et P-core pour la performance, les accélérateurs d'IA gaudi 3, qui sont spécifiquement conçus pour exécuter des tâches plus importantes, comme l'entraînement de nouveaux modèles d'IA & les processeurs *Lunar Lake*, qui seront utilisés dans la prochaine génération de PC d'IA à travers plus de 80 modèles d'ordinateurs portables différents, offrant 48 TOPS de performance d'IA, soit environ trois fois la vitesse de la génération précédente.

- **Le géant du matériel informatique Huawei basé à Shenzhen a récemment affirmé que sa puce d'IA Ascend 910B est au**

moins aussi bonne, voire meilleure, que l'A100 de Nvidia : dans certains tests, la puce Ascend 910B bat de 20 % l'A100 de Nvidia, l'un des GPU d'IA les plus populaires jamais créés, a déclaré le cadre. Nvidia réduit les prix de ses puces les plus avancées en Chine pour concurrencer Huawei.

- **ByteDance, la société mère de TikTok, a aussi annoncé une collaboration avec la société Broadcom sur de nouvelles puces d'IA**, malgré les tentatives des États-Unis de couper les entreprises chinoises des semi-conducteurs de pointe.

Annexe 5
Dépendances françaises et européennes sur la chaîne de valeur

	Potentiel du marché	Atouts (FR/UE)	Niveau de dépendance
FAI	Croissance du revenu des acteurs français en croissance continue sur les segments fixes et mobiles depuis 2022.	Normalisation des télécommunications, gestion du spectre radioélectrique et des orbites satellites.	Membre actif, élu au Conseil, représenté par l'ARCEP et des opérateurs comme Orange, qui participent aux travaux techniques. La France est le 4e contributeur budgétaire.
Satellite	Marché en croissance, (193 Mds $ en 2024 vs 392 Mds € en 2023), à comparer toutefois à la croissance du *cloud* qui représente près du double (1 170 Mds € en 2023).	Initiative européenne Iris[2] pour des usages souverains et B2B. NB : la fourniture d'un nombre restreint de services critiques via le déploiement de 170 satellites orbite basse d'ici 2027.	Domination des États-Unis dans le lancement satellite (107, vs 66 en Chine, 3 en UE en 2023[234], grâce à un modèle de commande publique offensif et des lanceurs réutilisables.

[234] *Un regard sur la Terre, « bilan des lancements orbitaux », janvier 2024.*

	Potentiel du marché	Atouts (FR/UE)	Niveau de dépendance
Câbles sous-marins	Marché essentiel aux communications mais dont la taille reste moindre comparé à d'autres maillons (4 Mds € C.A. monde en 2023).	Études sur les infrastructures numériques et leurs implications économiques et sécuritaires.	Participation par des représentants du gouvernement (Direction générale des entreprises) et des experts privés.
Équipementiers	Forte croissance du marché portée par la hausse du recours à la 4G/5G (80 % de la population en 2025 vs. 60 % en 2020).	Prise en compte du risque de dépendance et d'intelligence économique par les acteurs européens (principe de « précaution active » pour Huawei).	Les acteurs européens n'ont pas pris le tournant du PaaS pour commercialiser les fonctionnalités 4G/5G, ce que font très bien les acteurs américains.
Hébergement *Cloud* IaaS/ Paas	Marché massif (392 Mds € en 2023) et le plus dynamique des infrastructures numériques (23 % de croissance annuelle selon Gartner) Mouvement progressif des *hyperscalers* vers les solutions PaaS.	Mise en place de solution de *cloud* dit « souverain » par le biais de partenariats avec les *hyperscalers* (SENS, Bleu).	Domination des *hyperscalers* qui traduit la « convergence » réseaux-cloud-edge-IoT.
Constructeurs de DC	250 *data centers* en France et leur nombre croît de 13 % à 14 % par an, avec une croissance de 18 % par an prévue à partir de 2030.	Le mix énergétique décarboné et peu cher de la France et sa connexion aux *backbones* sous-marins sont des atouts stratégiques.	Enjeu d'attractivité du territoire français par rapport aux autres pays européens, en raison d'un environnement normatif trop instable et lourd.
Microprocesseurs	Marché mondial de production de près de 420 Mds $ en forte croissance, dominé par les acteurs asiatiques (près de 80%). NB : la conception est dominée par les acteurs américains (48 % de parts de marché en 2022).	Expertise européenne sur la fabrication de puces de plus de 18 nm (11 % de parts de marché dans la production).	Investissements insuffisants pour être compétitif sur la fabrication de puces de moins de 18 nm. Concurrence européenne pour attirer des usines, mais marché intérieur insuffisant pour les rentabiliser.

	Potentiel du marché	Atouts (FR/UE)	Niveau de dépendance
Informatique industrielle	Marché ayant souffert du ralentissement de la production en période Covid (-5,8 en 2020[235]). Reprise progressive pour atteindre les 10 % de croissance moyenne entre 2022 et 2029.	Un seul acteur européen dans le top 10 mondial (ABB - Suisse) en raison d'une politique de délocalisation.	Dépendance forte vis-à-vis des constructeurs japonais. Fortes barrières à l'entrée pour les acteurs européens en raison de l'avance prise par les acteurs à la pointe sur la R&D.

Annexe 6
Cartographie de l'offre de formation initiale certifiante aux infrastructures numériques

	Niveau 3 (niveau CAP, BEP)	**Niveau 4** (niveau bac)	**Niveau 5** (niveau DEUG, BTS, DUT, DEUST)	**Niveau 6** (niveau licence)	**Niveau 7** (niveau master)
Nombre d'établissements proposant des formations initiales ou continues diplômantes	47	1 003	607	151	214

[235] *Xerfi, « Le marché mondial de la robotique industrielle », juin 2021.*

	Niveau 3 (niveau CAP, BEP)	**Niveau 4** (niveau bac)	**Niveau 5** (niveau DEUG, BTS, DUT, DEUST)	**Niveau 6** (niveau licence)	**Niveau 7** (niveau master)
Types de métiers infrastructures du numérique	Monteur(se) - Raccordeur(se).	Monteur(se) de réseaux électriques / électricien(ne) Monteur(se) - Installateur(trice) Technicien(ne) fibre optique / radio Technicien(ne) maintenance.	Technicien(ne) fibre optique / cuivre / radio Technicien(ne) maintenance fibre optique / cuivre / radio Technicien(ne) réseau Technicien(ne) datacenter Monteur(se) – Installateur(trice) d'équipements connectés.	responsable cybersécurité Technicien(ne) réseau Administrateur(trice) réseau Technicien(ne) datacenter Technicien(ne) de maintenance d'équipements connectés.	Architecte système Ingénieur(e) infrastructures télécom Ingénieur(e) système & réseaux Data analyste Ingénieur(e) maintenance IT Ingénieur(e) BIM Ingénieur(e) CVC Ingénieur(e) électricien(ne).
Types de certifications	Le CQP Monteur Raccordeur FTTH/ Le TP Installateur de Réseaux de Télécommunications.	Bac Pro Métiers de l'électricité et de ses environnements connectés (MELEC). Bac pro Systèmes numériques option C réseaux informatiques et systèmes communicants (RISC).	BTS (enseignement très large).	BUT et Licences Pro.	Écoles d'ingénieurs généralistes.

Remerciements

L'Institut Montaigne remercie l'ensemble des personnes ayant contribué à l'élaboration de ce travail :

PRÉSIDENTS DU GROUPE DE TRAVAIL

- **Nicolas Guérin,** secrétaire général, Orange
- **Gérard Memmi,** professeur, Télécom Paris
- **Nicolas Bohy,** vice-président Cloud Practice de Kyndryl France

L'Institut Montaigne tient à adresser ses plus sincères remerciements à **Philippe Roncati**, président de Kyndryl jusqu'en mars 2024 et membre du groupe de travail, pour son engagement dans l'élaboration de ce rapport. Son expertise, sa disponibilité et son implication sans faille ont constitué un apport déterminant pour la qualité et la rigueur des analyses développées tout au long de ce travail.

MEMBRES DU GROUPE DE TRAVAIL

- **Louise Frion,** rapporteure, responsable de projets Nouvelles technologies, Institut Montaigne
- **Matthieu Bourguignon,** vice-président Europe, Nokia France
- **Phédon Cacouros,** rapporteur, pilote d'Innovation chez Orano
- **Nicolas David,** rapporteur, consultant senior en stratégie IT, BearingPoint France
- **Jérôme Martin,** rapporteur, associé, BearingPoint France
- **Paul Monnier,** rapporteur, associé, BearingPoint France
- **Stéphane Perrin,** *Chief technology officer,* Nokia France
- **Marc Petitier,** *Partner*, White & Case
- **Milo Rignell,** directeur des opérations, LightOn

- **Philippe Roncati,** *ex-CEO*, Kyndryl
- **Stephen Shibel,** *directeur de la décarbonation et de la transformation "as a service"*, Atos

RELECTEURS

- **Charlotte Baylac,** directrice France des affaires publiques, AWS
- **Jean Philippe Bonnet,** directeur adjoint, pôle stratégie, prospective, évaluation, RTE
- **Christophe Cousin,** responsable affaires publiques chez Amazon France, AWS
- **Mathieu Duchatel,** directeur des études internationales, Institut Montaigne
- **Godefroy Galas,** directeur de cabinet adjoint auprès du directeur général des entreprises, ministère de l'économie, des finances et de la souveraineté industrielle et numérique
- **Fred Geraud,** directeur affaires publiques et politiques publiques, Google Cloud
- **Daniel Kofman,** Professeur Telecom Paris, Codirecteur du PEPR Réseaux du Futur (France 2030)
- **Alexandra Laffitte,** directrice des affaires publiques, Lenovo ISG France
- **Antoine Lesserteur,** chargé des relations institutionnelles, France Data Center
- **Philippe Limantour,** *Chief Technology and cybersecurity officer*, Microsoft
- **Olivier Micheli,** président directeur général, Data4
- **Jean-Christophe Morisseau,** directeur général, Lenovo ISG France
- **Julien Nicolas,** directeur numérique, groupe SNCF
- **Alexandre Pébereau,** fondateur et membre du conseil d'administration, Tofane
- **Henri Pidault,** président 574 Invest, directeur des actifs Numériques du Groupe SNCF

- **Corentine Poilvet Clediere,** directrice France, LSEG
- **Guillaume Poupard,** directeur général adjoint, Docaposte
- **Stéphane Requena,** directeur technique et innovation, GENCI
- **Milo Rignell,** directeur des opérations, LightOn
- **Philippe Roncati,** ex CEO, Kyndryl
- **Jean Pierre Sabio,** directeur général, Gigalis
- **Arthur Sauzay,** associé, Allen & Overy Shearman
- **Stephen Shibel,** directeur de la décarbonation et de la transformation "as a service", Atos
- **Alain de Thomasson,** responsable comptes monde, Hitachi
- **Jérôme Totel,** directeur de l'innovation et de la stratégie de groupe, Data4

PERSONNES AUDITIONNÉES

- **Neil Abroug,** ex-coordinateur national de la stratégie pour les technologies quantiques auprès du SGPI
- **Henri d'Agrain,** délégué général du Cigref
- **Ombeline Bartin,** directrice des affaires publiques, groupe Iliad
- **Gilles Babinet,** entrepreneur et président du Conseil national du numérique
- **Jean Barrere,** *Partner*, Accuracy
- **Charlotte Baylac,** directrice politiques publiques France, AWS
- **Rodolphe de Beaufort,** Délégué général adjoint, GIMELEC
- **Jean Philippe Bonnet,** directeur adjoint, pôle stratégie, prospective, évaluation, RTE
- **Matthieu Bourguignon,** *Senior Vice President*, *head of Europe market*, Nokia
- **Yves Caseau,** directeur général et systèmes d'information du groupe Michelin
- **Laurent Celerier,** *Executive vice-president Central Europe & International business*, Orange Cyberdéfense

- **Miguel Cereijo,** *Enterprise software practice manager*, Hitachi
- **Roland Chedlivili,** co-directeur général, BU TowerCO de TDF
- **Béatrice de Clermont-Tonnerre,** investisseur, ex *General Manager* Secteur Public Microsoft France
- **Christophe Cousin,** responsable affaires publiques chez Amazon France, AWS
- **Dr Agnès Delaborde,** responsable évaluation IA, LNE
- **Emmanuel Dotaro,** *VP, Fellow* 5G-6G expertise, Thalès
- **Camille Dumouchel,** consultante chez Anthenor Public Affairs, pour le secrétariat général de l'OFITEM
- **Julien Duvaud-Schelnast,** *Partner*, Arthur D. Little
- **Blandine Eggrickx,** responsable des affaires publiques OVH
- **Hugues Even,** *Chief data officer*, Groupe BNP Paribas
- **Antoine de Fleurieu,** délégué général, GIMELEC
- **Antoine Fournier,** président de Thésée data centers
- **Philippe Herbert,** président de Mission 5G industrielle
- **Jason HSU,** *Senior Fellow* at Hudson Institute
- **Yosra Jarraya,** co-fondatrice et *CEO*, Astran
- **Thomas Jeanneret,** directeur général adjoint, LNE
- **Francis Jutand,** directeur exécutif adjoint des Mines Télécom
- **Aloïs Kirchner,** *Senior Fellow Industrie*, Institut Montaigne
- **Daniel Kofman,** professeur Telecom Paris, Codirecteur du PEPR Réseaux du Futur (France 2030)
- **Nicolas Kozakiewicz,** *Innovation executive advisor*, Wordline
- **Dr Agniezska Kupzok,** *IP Policy & Advocacy*, Nokia Technologies
- **Paul Labrogère,** directeur général, IRT System X
- **Alexandra Lafitte,** directrice des Affaires Publiques, Lenovo
- **Philippe Laval,** *CTO & Managing partner*, Jolt Capital
- **Philippe Legrand,** vice-président d'InfraNum et président du groupe Teleos
- **Arnaud Lucaussy,** secrétaire général TDF, président OFITEM
- **Olivier Michelier,** président Data4, président de France datacenter
- **Stella Morabito,** déléguée générale, AFNUM
- **Jean-Christophe Morisseau,** directeur général, Lenovo ISG France

- **Jean-Louis Mounier,** directeur général de la business unit Towerco, TDF
- **Aliette Mousnier-Lompré,** directrice générale, Orange Business
- **Michel-Marie Maudet,** directeur général du groupe, Linagora
- **Jean-Noël Patillon,** directeur adjoint de l'institut CEA List
- **Mathieu Pauwels,** *Chief operating officer*, Zurich
- **Pierre Peladeau,** *Partner*, Arthur D. Little
- **Thierry Plouvier,** président, Hitachi Energy France
- **Vincent Pointcheval,** directeur juridique et affaires publiques ATC, Secrétaire OFITEM
- **Guillaume Poupard,** directeur général adjoint, Docaposte
- **Arno Pons,** délégué général, Digital New Deal
- **Mahasti Razavi,** avocat associé, Managing partner d'August Debouzy
- **Stéphane Requena,** directeur technique et innovation, GENCI
- **Jean-Louis Rougier,** enseignant chercheur, Télécom Paris
- **Guillaume de Saint Marc,** *VP Engineering*, Outshift by Cisco
- **Christophe Samson,** DG Peaksys, DSI Cdiscount
- **Arthur Sauzay,** associé, Allen & Overy Shearman
- **Hubert Tardieu,** ancien président, actuellement administrateur indépendant de Gaia-X
- **Alain de Thomasson,** responsable comptes monde, Hitachi
- **Aurélien Vigano,** *SVP International Infrastructures*, Orange
- **Joël Vormus,** délégué *Data Centers*, GIMELEC

Les rapporteurs remercient **Marie-Pierre de Bailliencourt**, directrice générale de l'Institut Montaigne, pour son suivi attentif tout au long de ce projet ainsi que l'ensemble des équipes de l'Institut Montaigne qui ont contribué à l'élaboration de ce rapport, notamment **Lisa Thomas-Darbois**, **Catherine Merle-du-Bourg**, **Nicolas Laine**, **Luna Vauchelle**, **Clara Yazi** et **Brian Ndungo Quiassata**.

Les opinions exprimées dans ce rapport n'engagent ni les personnes précédemment citées ni les institutions qu'elles représentent.

Retrouvez nos autres notes et rapports sur les mêmes sujets :

Nouvelles technologies

- **Quantique : vers une logique de marché**
  (Note d'action • Octobre 2024)

- **Données de santé : libérer leur potentiel**
  (Note d'action • Février 2024)

- **Pour une Autorité française de l'IA**
  (Note d'action • Janvier 2024)

- **Cybersécurité : passons à l'échelle**
  (Rapport • Juin 2023)

- **Blockchain : consolider nos atouts**
  (Rapport • Juin 2023)

- **Mobiliser et former les talents du numérique**
  (Note d'action • Mai 2023)

L'ensemble de nos travaux et publications
est disponible sur notre site institutmontaigne.org

Président

Henri de Castries Président, Institut Montaigne

Membres

Emmanuelle Barbara *Senior Partner*, August Debouzy
Jean-Pierre Clamadieu Président du Conseil d'Administration, ENGIE
Paul Hermelin Président du Conseil d'administration, Capgemini
Marwan Lahoud Directeur général délégué de Tikehau Capital, Président du Private Equity
Natalie Rastoin Présidente, Polytane ; *Senior Advisor*, WPP
Jean-Dominique Senard Président du Conseil d'administration, Groupe Renault
Arnaud Vaissié Président-directeur général, International SOS
Natacha Valla Économiste ; doyenne de l'École de Management et d'Innovation, Sciences Po
Florence Verzelen Directrice générale adjointe, Dassault Systèmes
Philippe Wahl Président-directeur général, Groupe La Poste

Président d'honneur

Claude Bébéar Fondateur et président d'honneur, AXA

Institut Montaigne
59 rue La Boétie, 75008 Paris
Tél. +33 (0)1 53 89 05 60
institutmontaigne.org

Imprimé en France
Dépôt légal : mars 2025
ISSN : 1771-6764

Les adhérents

INSTITUT Montaigne

ABB France
AbbVie
Accenture
Accor
Accuracy
Actual Group
Adeo
ADIT
Air Liquide
Allianz
Amazon
Amber Capital
Amundi
Antidox
Antin Infrastructure Partners
ArchiMed
Ardian
Arquus
Arthur D. Little
AstraZeneca
August Debouzy
AXA
AXA IARD
A&O Shearman
Bain & Company France
Baker & McKenzie
BearingPoint
Bessé
BNP Paribas
Bolloré
Bouygues
Bristol Myers Squibb
Brousse Vergez
Brunswick
Capgemini
Capital Group
CAREIT
Carrefour
Chubb
CIS
Clariane
Clifford Chance
CNP Assurances
Cohen Amir-Aslani
Conseil supérieur du notariat
D'Angelin & Co.Ltd
Dassault Systèmes
Delair
Deloitte
De Pardieu Brocas Maffei
Domia Group
Edenred
EDF
EDHEC Business School
Edmond de Rothschild
Ekimetrics France
Engie
EQT
ESL & Network
Eurogroup Consulting
FGS Global
Forvis Mazars
Getlink
Gide Loyrette Nouel
Gigalis
Google
Groupama
Groupe Bel
Groupe M6
Groupe Orange
Hameur et Cie
Henner
Hitachi Energy France
Hogan Lovells
Howden
HSBC Continental Europe
IBM France
IFPASS
Incyte Biosciences France
Inkarn
Institut Mérieux
International SOS
Interparfums
Intuitive Surgical
Ionis Education Group
iQo
ISRP
Jeantet associés
Johnson & Johnson
Jolt Capital
Katalyse
Kea
Kearney
KPMG S.A.
Kyndryl
La Banque Postale
La Compagnie Fruitière
LCH SA
Lenovo ISG
Linedata Services
Lloyds Europe
L'Oréal
LVMH - Moët-Hennessy - Louis Vuitton
M.Charraire
MACSF
Média-Participations
Mediobanca
Mercer
Meridiam
Microsoft France
Mitsubishi France S.A.S
Moelis & Company
Moody's France
Morgan Stanley
Natixis
Natural Grass
Naval Group
Nestlé
OCIRP
ODDO BHF
Ondra Partners
Optigestion
Orano
PAI Partners
Pelham Media
Pergamon
Polytane
Publicis
PwC France & Maghreb
Qualisocial
Raise
RATP
Renault
Ricol Lasteyrie
Rivolier
Roche
Roche Diagnostics
Rokos Capital Management
Rothschild & Co
RTE
Safran
Sanofi
SAP France
Schneider Electric
ServiceNow
Servier
SGS
SIER Constructeur
SNCF
SNCF Réseau
Sodexo
SPVIE
SUEZ
Synergie
Teneo
The Boston Consulting Group
Tilder
Tofane
TotalÉnergies
TP ICAP
Transformation Factory
Unicancer
Veolia
Verian
Verlingue
VINCI
Vivendi
Vodafone Group
Wavestone
Wendel
White & Case
Willis Towers Watson France
Zurich

Les infrastructures numériques évoluent vers une convergence entre les réseaux de communication et le traitement des données. Autrefois distincts, ces deux univers s'intègrent aujourd'hui grâce à la virtualisation et au *cloud*, et cette tendance s'accélère avec le développement du *edge computing*. **Concrètement, cela signifie que les acteurs du numérique ne se limitent plus à fournir une simple connectivité : ils proposent aussi des services de traitement et d'analyse des données directement sur leurs infrastructures**.

Par exemple, un opérateur télécom ne se contente plus d'assurer l'accès à internet. Il permet aussi aux entreprises de traiter leurs données au plus près des utilisateurs, en intégrant des fonctionnalités de *cloud* et de calcul distribué directement dans son réseau *(Network-as-a-Service)*. De même, une voiture autonome ne peut pas attendre qu'un serveur distant analyse en temps réel les informations de son environnement. Grâce au *edge computing*, ces calculs se font localement, sur des serveurs situés à proximité des antennes réseau, réduisant ainsi la latence et améliorant la réactivité du véhicule.

Cette transformation impose de repenser les infrastructures numériques de manière globale, car réseau et traitement des données sont désormais indissociables. Cela nécessite une approche stratégique, fonction des usages actuels et futurs des infrastructures numériques, qui va au-delà de l'approche actuelle de « service public » qui a été adoptée pour le déploiement de la fibre ou pour le raccordement des *data centers* sur le territoire français au réseau électrique.

Dans ce contexte, ce rapport explore plusieurs questions clés :

1. Quels usages devons-nous maîtriser de bout en bout, et comment adapter nos infrastructures en conséquence ?
2. Quelle puissance de calcul devons-nous garantir sur le territoire pour rester compétitifs à l'échelle mondiale ?
3. Comment organiser le maillage des *data centers* pour répondre aux besoins économiques et sociaux ?
4. Comment former les talents à l'interconnexion croissante entre métiers du réseau et du traitement des données ?
5. La 5G privée suivra-t-elle la trajectoire du *cloud* dans les années 2000, et comment accompagner son adoption ?
6. Comment valoriser l'excellence de nos infrastructures matérielles – câbles terrestres et sous-marins, satellites ?
7. Comment tirer parti de l'échelle européenne pour mieux intégrer les enjeux de cybersécurité et limiter les dépendances critiques ?

10 €
ISSN : 1771-6764
RAP2503-01