

Allouer le kilowattheure-IA : pourquoi le marché énergétique n'est pas un protocole

De la régulation des modèles à l'explicitation des arbitrages matériels.
Second volet d'un diptyque sur les contraintes physiques de l'IA

I. Un seuil franchi : la ressource sous contrainte est en train d'être allouée

L'article qui ouvre ce diptyque, *La virtualité supposée de l'IA face au mur du réel* (février 2026), établissait que l'IA générative est entrée dans un régime d'industrie lourde, contrainte par un faisceau convergent de ressources finies : énergie, eau, silicium, données. Trois mois plus tard, le diagnostic est confirmé empiriquement plus vite qu'attendu.

L'Agence internationale de l'énergie projette désormais 1 100 TWh de consommation par les datacenters dès 2026 (révision haussière de 18 % par rapport à l'estimation de décembre 2025), avec environ 5 à 15 % imputables à l'IA aujourd'hui et 35 à 50 % anticipés à l'horizon 2030. Aux États-Unis, *Fortune* (20 avril 2026) documente que ces datacenters absorbent à eux seuls environ la moitié de la *croissance* de la consommation électrique nationale ; le prix de détail s'est accru de l'ordre de 40 % sur 2021-2026 ; le suivi de Data Center Watch recense, sur la période mai 2024 à mars 2025, seize projets affectés par l'opposition locale (six entièrement bloqués, dix retardés) pour 64 milliards de dollars, chiffre auquel le Q2 2025 ajoute à lui seul vingt projets supplémentaires pour 98 milliards de dollars, le cumul 2025 atteignant au moins quarante-huit projets pour un volume total dépassant largement 100 milliards de dollars. Goldman Sachs estime un impact de 0,1 point sur la core inflation US en 2026 et 2027.

Ces chiffres ne sont pas des prévisions. Ce sont des arbitrages déjà engagés.

La question pertinente n'est donc plus : *l'IA consomme-t-elle trop ?* Elle est : *selon quel mécanisme cette consommation est-elle distribuée entre usages concurrents ?* L'objet a changé. La matérialité n'est plus l'hypothèse à démontrer ; elle devient la prémisse de l'arbitrage. Et l'arbitrage actuel est triplement distribué (marché énergétique, contrats long terme, oppositions locales), sans qu'aucun de ces trois mécanismes ne soit conçu pour délibérer.

Domaine de validité : l'analyse qui suit s'appuie principalement sur des données et des cas issus des États-Unis et de l'Union européenne. La transposition à d'autres régions (Asie de l'Est, Amérique latine, Afrique) appelle une actualisation contextuelle distincte que cet article ne prétend pas conduire.

II. Où se prend, aujourd'hui, la décision d'allouer un MWh à l'IA

Trois niveaux à séparer

Avant cartographie, une distinction analytique s'impose. Trois niveaux sont fréquemment confondus dans le débat public, et la confusion bloque l'argument :

- *Allocation physique* : flux d'énergie effectivement livrés à un usage (kWh transitant du réseau au datacenter).
- *Allocation économique* : signal-prix qui hiérarchise les capacités à payer pour cette ressource (€/MWh, contrats, primes).
- *Arbitrage normatif* : décision collective sur la valeur sociale relative des usages auxquels la ressource est affectée.

L'allocation physique se fait en temps réel et ne pose pas problème : le réseau assure la livraison physique de la puissance appelée. L'allocation économique se fait via le marché énergétique et fonctionne sur sa dimension propre. L'arbitrage normatif n'est, lui, exécuté par aucun dispositif. Le glissement implicite par lequel on assimile les trois niveaux (*le marché alloue, donc le marché arbitre*) est précisément le sophisme dominant. Le prix agrège des préférences solvables sous contrainte locale ; il n'agrège pas des préférences collectives non solvables, ni des externalités différées, ni des rendements sociaux différenciés. Cette proposition déplace la discussion du jugement normatif vers la description mécanique : elle ne juge pas les marchés mais décrit ce qu'ils font et ne font pas.

Cartographie en trois couches

Le marché énergétique. Première couche. Le signal-prix agrège localement l'information sur la rareté, forme un prix qui circule entre producteurs, transporteurs et consommateurs, et permet une réallocation continue. Ce mécanisme fonctionne sur sa dimension propre, l'allocation économique. Il suppose toutefois deux conditions pour produire incidemment un arbitrage normatif acceptable : que les acteurs en compétition pour la ressource paient un prix marginal effectivement comparable, et que leur capacité à payer corrèle avec la valeur sociale de leurs usages. Pour les datacenters IA, la première condition est seulement nominalement remplie ; la seconde ne peut pas être présumée.

Les contrats long terme. Seconde couche. Les accords d'achat d'électricité (Power Purchase Agreements / PPA) signés sur des horizons de quinze à vingt ans soustraient des volumes durables au mécanisme spot. Microsoft a contracté la remise en service de l'unité 1 de Three Mile Island avec Constellation Energy (environ 835 MW), et plus de 10 GW de capacités renouvelables additionnelles à l'horizon 2030. Amazon Web Services vise plus de 5 GW alimentés par SMR d'ici 2039. Google a signé un contrat de 500 MW avec Kairos Power et s'engage publiquement sur une alimentation 24/7 décarbonée. Ces contrats ne sont pas une distorsion du marché ; ils en sont l'emploi rationnel par les acteurs qui peuvent payer la prime de sécurisation. Leur effet net est néanmoins celui d'une *soustraction* : les volumes ainsi engagés sont retirés du mécanisme spot pour la durée du contrat. La temporalité contractuelle excède la temporalité électorale d'un facteur trois ou quatre.

À cette soustraction privée, les cadres régulateurs en vigueur n'opposent aucune procédure d'arbitrage d'usage. L'AI Act européen, dont les pouvoirs d'exécution sur les modèles GPAI s'activeront le 2 août 2026, ne dispose d'aucune disposition de hiérarchisation d'usage sur critère ressource. Le NIST AI RMF Profile *Critical Infrastructure*, publié en concept note le 7 avril 2026, concerne la gestion du risque pour les opérateurs d'infrastructure critique, pas la stratification publique de leurs usages d'IA.

Les blocages locaux. Troisième couche. Quarante-huit projets bloqués ou retardés en 2025 aux États-Unis selon Data Center Watch, pour un volume excédant largement 100 milliards de dollars ; plafond irlandais sur la part datacenter de la consommation nationale ; moratoires de Singapour et des Pays-Bas sur les nouveaux projets ; controverses autour des méga-bassines en France, déjà mobilisées comme analogie procédurale dans l'article-mère. Ces oppositions produisent un effet régulateur réel mais imprévisible. Leur logique n'est pas l'arbitrage entre usages : elles ne disent pas à *quoi* le projet contesté aurait dû céder la place, mais opposent un veto à un usage particulier dans un lieu particulier. C'est une *sanction ex post*, pas un *arbitrage ex ante*.

Démonstration mécaniste de l'aveuglement

L'aveuglement du marché énergétique aux rendements sociaux n'est pas un défaut accidentel. Il est mécanique, et il l'est dans ce cas précis pour trois raisons qui composent un système causal fermé.

D'abord, *non-différenciation par usage final*. L'électricité est fongible au point de livraison. Le réseau ne sait pas, par construction, distinguer un MWh affecté à un IRM d'un MWh affecté à un agent conversationnel récréatif : il livre des électrons au compteur. Cette fongibilité est techniquement nécessaire ; elle a pour conséquence directe que toute différenciation par usage final relève du dispositif institutionnel, pas du dispositif technique.

Ensuite, *asymétrie de capital*. La capacité à sécuriser l'accès énergétique long terme dépend du volume capitalistique des acteurs en compétition. Un hyperscaler engage des milliards en CapEx énergétique ; une PME hospitalière, un laboratoire académique, un opérateur d'infrastructure critique non solvable au même standard ne peuvent pas déployer ces dispositifs.

Enfin, *horizon temporel*. Les contrats de quinze à vingt ans excèdent la temporalité du signal spot. Une fois engagés, ils retirent durablement du marché les volumes les plus stratégiques. Le signal-prix continue de fonctionner, mais sur un sous-ensemble de la ressource, celui qui n'a pas été pré-alloué.

La proposition formelle qui en résulte tient en une phrase : *dans un système où la ressource est fongible au point de livraison, le capital concentré et les engagements intertemporels dominants, le signal-prix converge structurellement vers une allocation par solvabilité et non par valeur sociale*. Ce n'est pas une critique du marché en général. C'est une description de ce que ce marché fait, ici, dans ces conditions précises.

Nous avons construit une métrique d'aptitude fonctionnelle. Nous n'avons jamais construit une métrique d'efficacité socio-technique, et c'est pourquoi l'allocation suit la solvabilité.

Trois propriétés émergentes du système

Le système marché-contrats-blocages n'est pas seulement une cartographie. Il possède trois propriétés émergentes négatives, observables et nommables : *absence de réversibilité collective*, les volumes pré-alloués par contrats long terme excèdent la temporalité de tout dispositif politique ordinaire ; *absence de comparabilité des usages*, le mécanisme prix égalise des solvabilités, pas des valeurs sociales ; *absence de point d'entrée institutionnel*, aucun mécanisme explicitement conçu n'existe pour arbitrer, en amont, l'allocation de la ressource entre usages concurrents.

La formule qui résume ce système distribué tient en une phrase : **le marché alloue, les contrats verrouillent, les oppositions sanctionnent ; aucun des trois ne délibère**. L'allocation existe. La procédure n'existe pas. Le système est stable, non mandaté, non révisable collectivement, non instruit par une métrique publique.

III. Défaillance de l'instrument : le benchmark sans empreinte

Le défaut de procédure est solidaire d'un défaut métrologique. Le pont entre les deux n'est pas analogique. Il est nécessaire.

Un système de décision ne peut arbitrer que ce qu'il mesure. En l'absence de métrique couplant performance et coût ressource, toute allocation est mécaniquement découplée

de la valeur. La proposition ne dépend pas d'une école d'économie ou d'une option idéologique. Elle décrit une condition d'instrumentation du jugement. Pour qu'un arbitrage entre usages puisse advenir, il faut disposer d'une métrique qui couple ce qu'un usage produit à ce qu'il consomme. Cette métrique n'existe pas, au sens où elle n'a jamais été construite comme infrastructure publique, et au sens où les benchmarks qui font office de standards de fait sont précisément définis pour mesurer la performance *indépendamment* du coût ressource.

C'est ce que le corpus Twingital désigne, sous le nom RAISE, comme la critique du *benchmark sans empreinte*. Un MMLU à 89 %, un SWE-bench Verified à 84 %, un GAIA à 65 % décrivent une aptitude fonctionnelle ; ils ne disent rien de l'énergie, de l'eau, du silicium amorti ou du temps GPU mobilisés pour atteindre cette aptitude. Ce silence n'est pas neutre. Il rend les performances comparables uniquement sur la justesse fonctionnelle, et structurellement non comparables sur la dimension qui importe pour l'arbitrage : le rendement ressource pour une tâche donnée.

Le 12 avril 2026, le Center for Responsible Decentralized Intelligence de Berkeley a publié une démonstration empirique que les huit principaux benchmarks d'agents (SWE-bench, WebArena, OSWorld, GAIA, Terminal-Bench, FieldWorkArena, CAR-bench et un huitième) pouvaient être systématiquement saturés par reward hacking automatisé, obtenant des scores quasi-parfaits sans résoudre une seule tâche réelle. L'étude a porté sur huit benchmarks existants ; elle démontre la fragilité logique d'une *classe*, celle qui n'incorpore pas le coût ressource, plutôt qu'une vulnérabilité empirique universelle, mais le mécanisme sous-jacent s'applique à toute métrique conçue de la même manière. Un benchmark qui ne contraint pas son optimisation par un coût ressource est, par construction, exposé à toutes les stratégies qui maximisent le score sans produire l'effort de résolution correspondant dans des conditions opérationnelles. La fragilité observée est la conséquence prévisible d'une métrique conçue pour être maximisée plutôt que pour être instructive.

Cette fragilité de l'évaluation agentique ne prouve pas, à elle seule, l'incapacité à arbitrer un MWh. Elle le devient lorsque ces benchmarks sortent de leur fonction d'évaluation pour entrer en fonction d'allocation, c'est-à-dire lorsqu'ils deviennent signaux d'investissement, critères d'achat public, ou bases de certification sectorielle. Le pont entre défaillance métrologique et défaillance d'allocation n'est pas conceptuel. Il est institutionnel. Le glissement par lequel un test technique se transforme en clé d'admission, sans avoir été conçu pour porter cette charge, est ce que le corpus Twingital désigne sous le nom de *port de promotion* du benchmark. La métrique fait carrière au-delà de sa juridiction d'origine, et son emploi dans un cadre décisionnel pour lequel elle n'a pas été calibrée produit des dérives prévisibles.

Le AI Index 2026 de Stanford HAI documente un autre symptôme. Les taux d'hallucinations des 26 LLM frontière évalués s'étendent de 22 % à 94 % selon le domaine

; les benchmarks généralistes de type MMLU et MMLU-Pro sont fonctionnellement saturés au-delà de 88 %, ce qui rend les écarts entre modèles à ces scores statistiquement non exploitables. Gartner indique pour sa part que seulement 28 % des projets IA en infrastructure et opérations livrent le ROI promis, un projet sur cinq s'effondrant en cours d'exécution. La métrique de capacité corrèle faiblement, ou ne corrèle pas observablement, avec la métrique de valeur opérationnelle ; et aucune des deux n'est instrumentée par une métrique de coût ressource.

Le précédent est ailleurs. Les crises financières des années 2000 n'ont pas résulté d'une absence de marché. Elles ont résulté d'une métrique de risque institutionnalisée et inadéquate (VaR, ratings AAA sur subprimes) dont la défaillance a contaminé l'arbitrage des allocations à l'échelle du système. En santé, des décennies de surtraitement n'ont pas tenu à un défaut d'évaluation clinique. Elles ont tenu à une sélection d'endpoints primaires (critères intermédiaires insuffisamment reliés aux bénéfices cliniques finaux) dont l'inadéquation a orienté la décision médicale en aval. Le problème, dans les deux cas, n'est pas l'absence d'instrument. C'est son *inadéquation*. Et c'est cette même inadéquation qui agit dans l'écosystème IA actuel.

L'aveuglement métrologique et l'aveuglement politique partagent la même cause structurelle. La construction de l'instrument et la construction du dispositif décisionnel ont en commun d'avoir été réalisées hors de toute exigence de coût ressource. Corriger l'un sans l'autre revient à conserver la défaillance.

IV. Trois défaillances structurelles du système implicite

La cartographie de la section II décrit un système qui *fonctionne*. Il alloue effectivement les ressources rares à des usages, sans interruption fatale, en présence de chocs majeurs : bascule de l'opinion publique américaine, blocages locaux excédant 100 milliards. Cette robustesse fonctionnelle est précisément ce qui rend la critique difficile : les marchés énergétiques ne sont pas en panne, les contrats long terme ne sont pas illégaux, les oppositions locales ne sont pas pathologiques. Le système ne dysfonctionne pas sur sa dimension propre, technique et économique. Il décide de l'allocation, mais sans pouvoir trancher l'arbitrage normatif.

Trois défaillances en éclairent la cause unique.

Non-comparabilité des usages

Un mégawattheure n'est pas homogène à un mégawattheure. Un MWh affecté à la modélisation des trajectoires climatiques de l'Atlantique Nord, un MWh affecté à un dispositif de diagnostic assisté en oncologie pédiatrique, et un MWh affecté à un agent conversationnel récréatif n'ont pas les mêmes rendements sociaux. Le marché énergétique ne fait pas cette différence. Il ne le peut pas, par construction, les trois

mécanismes de la section II le verrouillent. Le marché ne compare pas des usages, il compare des solvabilités. Lorsque la capacité à payer corrèle avec la valeur sociale, le mécanisme produit incidemment un arbitrage acceptable. Lorsqu'elle en diverge (et la divergence est précisément la situation contemporaine, où des acteurs très capitalisés peuvent payer pour des usages dont la valeur sociale marginale est faible), le marché alloue de manière *efficente au sens de Pareto* mais *muette sur la valeur*.

Verrouillage intertemporel

Les contrats long terme, économiquement rationnels du point de vue de l'hyperscaler dont la marge dépend de l'accès garanti à plusieurs gigawatts pilotables, produisent une soustraction durable du marché spot. L'enjeu n'est ni la légalité ni l'efficacité du mécanisme : il est l'asymétrie de temporalité. Dix ans de mandat législatif, cinq ans de mandat exécutif, deux à trois ans de cycle régulateur, contre quinze à vingt ans de PPA. La géométrie est asymétrique, et cette asymétrie est elle-même un fait politique. La fenêtre de révision avant verrouillage contractuel se referme entre 2027 et 2028. Passé ce seuil, quinze à vingt ans de PPA hyperscalers dominant les volumes marginaux réduiraient à presque rien la capacité d'arbitrage public, non parce que les contrats l'interdiraient, mais parce qu'ils auraient préempté la matière de l'arbitrage.

Substitution métrique : le « green PPA »

L'article-mère établissait au paragraphe 8.4 qu'un PPA renouvelable peut, lorsque l'additionnalité physique et la concordance temporelle ne sont pas démontrées, opérer essentiellement une *réallocation comptable* de production existante. Tous les PPA ne sont pas dans ce cas : certains financent effectivement de la capacité bas-carbone qui n'aurait pas été construite autrement. Mais lorsque les conditions ci-dessous ne sont pas réunies, le risque est de déplacer la légitimité vers une comptabilité d'attribution, indépendamment de la réalité physique du flux d'électrons. Trois propriétés techniques précises ancrent la critique.

Non-additionnalité physique : lorsqu'un PPA n'est pas associé à la construction de capacité bas-carbone qui n'aurait pas existé sans lui, l'achat de garanties d'origine attribue comptablement à l'acheteur une production renouvelable disponible par ailleurs ; la production fossile résiduelle continue de couvrir la demande nette ailleurs dans le système.

Désynchronisation temporelle production / consommation : les datacenters consomment 24 heures sur 24 ; les sources renouvelables intermittentes (solaire, éolien) produisent selon des profils variables et non corrélés. Sans capacité de stockage longue durée à l'échelle, le matching annuel ou même mensuel masque un déséquilibre horaire que la production fossile pilotable continue de combler.

Illusion de traçabilité énergétique : l'attribution comptable d'un certificat à une consommation crée une représentation de chaîne énergétique propre (*24/7 carbon-free, 100 % renewable matched*) à laquelle ne correspond, dans le système physique, aucune relation directe et continue entre la production attribuée et la consommation effective. La métrique est juste sur son périmètre comptable propre ; elle est trompeuse sur la dimension physique qu'elle prétend résumer.

Ces trois propriétés sont admises dans la littérature technique. Leur effet, repris dans le cadre du présent article, est cumulatif : la métrique de décarbonation comptabilisée *tient lieu de débat* sur l'allocation. Elle déplace la conversation publique d'un terrain où le coût matériel serait explicite à un terrain où la performance environnementale, même partiellement fictive, occupe l'espace de la légitimité. Nous avons inventé le *24/7 carbon-free* avant d'inventer le *24/7 socially-justified*.

Trois défaillances, une cause unique : l'absence d'un cadre permettant de relier ressource, usage et valeur. Le système marché-contrat-blocage est stable et non instruit. Cette combinaison n'est pas indéfiniment soutenable. L'opinion publique américaine bascule à mesure que le prix retail monte et que les blocages se multiplient ; les juridictions territoriales improvisent des plafonds dont la valeur normative est faible mais dont la valeur signal est forte ; le verrouillage par contrats long terme s'élargit à mesure que les acteurs anticipent une régulation possible. Ce ne sont pas trois mouvements indépendants. Ce sont trois conséquences cohérentes de la même cause.

V. Conditions d'un protocole : conditions d'accès différenciées, non planification

L'objection libérale doit être adressée frontalement. Construire un protocole délibératif sectoriel autour de l'allocation du kilowattheure-IA, c'est ouvrir la porte à la nostalgie planificatrice, à la capture politique, à la lenteur arbitraire. C'est, plus brutalement formulé, troquer une décision distribuée et efficiente pour une décision centralisée et contestable.

L'objection est sérieuse. Elle est aussi, dans sa forme actuelle, mal calibrée. Elle suppose que l'alternative au protocole est un marché concurrentiel idéal. Or l'alternative au protocole est le marché captif décrit en section II : signal-prix borné par contrats long terme et corrigé par opposition NIMBY. Ce n'est pas la configuration canonique que les théories de l'efficacité des marchés mobilisent pour défendre la supériorité du signal-prix. Reformuler le débat en termes de protocole revient donc à comparer non *marché vs. planification*, mais *marché captif sans procédure vs. marché contraint par procédure*. Le protocole ne se substitue pas au mécanisme prix. Il l'encadre.

Verrouillage doctrinal de la distinction : **un protocole d'allocation IA ne classe pas les usages pour leur affecter directement la ressource. Il définit des conditions d'accès différenciées à la ressource.** Cette précision déplace l'objet du dispositif d'une logique d'allocation centrale (que personne de raisonnable ne propose) à une logique d'admissibilité contrôlée (que des dispositifs sectoriels existants pratiquent depuis des décennies). Concrètement, des conditions d'accès différenciées peuvent prendre la forme d'exigences de preuve plus fortes pour certains usages, de contraintes d'efficience minimale variables selon strate, d'obligations de transparence asymétriques selon le profil d'externalité. Aucune de ces formes ne consiste à prescrire qui consomme quoi.

Quatre précédents partiels, mobilisés comme repères et non comme modèles, éclairent les conditions de praticabilité. L'allocation des bandes de spectre par l'UIT et, en France, par l'ARCEP, organise des conditions d'accès différenciées à une ressource rare avec révisions périodiques, sans paralyser l'innovation industrielle. Les SDAGE français (Schémas Directeurs d'Aménagement et de Gestion des Eaux) articulent niveau européen, national et bassin pour hiérarchiser les conditions d'usage de l'eau sous contrainte. La gestion des slots aéroportuaires combine signal-prix et contraintes d'admissibilité dans un secteur où la rareté de la ressource est non-substituable. La Convention Citoyenne pour le Climat de 2019-2020 doit être mobilisée non comme inspiration naïve mais comme avertissement procédural : son taux d'application limité montre que la procédure délibérative n'est légitime *que* si elle est articulée à un mandat exécutif clair et à des mécanismes de suivi observables.

Aucun de ces dispositifs n'est transposable tel quel. Tous indiquent que l'idée d'un protocole sectoriel n'a rien d'utopique : elle a, en fait, plus d'antécédents historiques fonctionnels que l'alternative actuelle.

Quatre conditions paraissent nécessaires à la praticabilité d'un tel protocole.

Condition 1, couplage performance / empreinte. Toute performance déclarée par un benchmark utilisé comme gate institutionnel (audit de conformité, marché public, certification sectorielle) doit être accompagnée d'un coût ressource normalisé. Cette exigence est la sortie directe de la critique RAISE du benchmark sans empreinte. **Sans métrique couplée, le protocole est impossible. Avec métrique couplée, le marché devient arbitrable.** Le protocole ne remplace pas le marché ; il le rend opérant sur une dimension qu'il ne voit pas aujourd'hui. Cette inversion est le pivot conceptuel de toute la proposition. Une fois la métrique disponible comme infrastructure publique, les arbitrages peuvent à nouveau s'opérer en partie par le signal-prix, sur des dimensions qui agrègent performance et coût ressource, ce que ni l'un ni l'autre des trois mécanismes actuels n'agrège.

Condition 2, conditions d'accès différenciées explicites. Distinguer publiquement, par un dispositif lisible, des strates d'usage (santé et infrastructure critique, recherche scientifique et modélisation à externalité positive forte, automatisation productive,

agentique et inférence non productive) auxquelles s'appliquent des conditions d'accès distinctes : exigences de preuve d'utilité sociale, contraintes d'efficacité minimale, obligations de transparence opérationnelle. Cette stratification n'est pas une planification. Elle est l'explicitation d'arbitrages déjà effectués implicitement par les hyperscalers, les juridictions locales et les opérateurs de réseau. Le passage de l'implicite à l'explicite est le geste politique propre.

Condition 3, multi-niveau juridictionnel. Échelle européenne pour la grille d'usages, la prévention du carbon leakage et l'harmonisation minimale qui prévient la course au moins-disant ; échelle nationale pour la mise en œuvre territoriale et la planification de l'infrastructure ; échelle locale pour l'acceptabilité et les contreparties. Sans cette articulation, soit l'on tombe dans une harmonisation européenne déconnectée des terrains, soit l'on s'enferme dans des décisions locales contradictoires. Un protocole adopté unilatéralement par une seule juridiction (européenne par exemple) produirait inévitablement un *carbon / regulatory leakage* : déplacement des workloads d'IA les plus intensifs vers des juridictions sans protocole, concentration géographique amplifiée, accroissement des asymétries entre blocs. La multi-juridictionnalité n'est pas un raffinement ; elle est une condition de non-contournement.

Condition 4, réversibilité institutionnelle. Cycles de révision quinquennaux, observation des effets non anticipés, possibilité de redéfinir les strates et les conditions d'accès. Sans réversibilité programmée, le protocole reproduit le verrouillage qu'il prétend corriger. La temporalité contractuelle des PPA hyperscalers (15 à 20 ans) doit être absorbée par une temporalité de révision plus courte, faute de quoi la procédure devient elle-même un instrument de soustraction.

À échelle opérationnelle, des projets opérationnels d'allocation d'IA en contextes critiques (santé pédiatrique notamment) montrent qu'une stratification d'usages est techniquement opérable lorsque les boucles d'inférence à haute criticité, les workloads prédictifs et les tâches documentaires ne sont pas placés dans la même classe d'admissibilité, chaque classe portant ses propres exigences de preuve, de transparence et d'efficacité. Cette opérabilité interne, observée dans des contextes régulés (médecine pédiatrique, dispositifs sous contrainte HDS et conformité ISO), ne constitue pas une preuve générale du protocole sectoriel proposé. Elle réfute en revanche, par existence, l'objection que la stratification d'usages serait un objet conceptuel hors prise opérationnelle. Il est possible, dans un système contraint et régulé, de définir des conditions d'accès différenciées qui ne paralysent pas la production. Le passage de l'opérabilité interne au dispositif sectoriel public n'est pas trivial : il est le travail institutionnel propre que les régulateurs européens auront à conduire dans la fenêtre 2026-2028, avant que le verrouillage par contrats long terme ne ferme la fenêtre d'arbitrage.

VI. Conclusion : de la régulation à l'allocation

L'article-mère concluait sur une question : quelle part de nos ressources physiques finies nos sociétés accepteront-elles d'allouer à l'IA, et au détriment de quels autres usages. Le présent article ne répond pas à cette question. Il en déplace l'objet.

L'IA n'est plus seulement un objet à réguler. Elle est un vecteur d'allocation de ressources physiques rares. Réguler ses modèles, ses usages à risque, ses biais : l'AI Act le fait, l'AI SI britannique l'évalue, le NIST en construit les standards. Aucun de ces instruments ne décide, et aucun n'a vocation à décider, comment les kilowattheures, les hectolitres et les hectares affectés à cette industrie sont arbitrés contre leurs usages alternatifs. La régulation des modèles ne tient pas lieu d'arbitrage de l'allocation matérielle. Le diptyque que forment ces deux articles soutient une thèse simple : la doctrine doit explicitement séparer les deux, et instrumenter chacune par ses dispositifs propres.

Le coût réel de l'IA ne se mesure pas seulement à l'énergie qu'elle consomme. Il se mesure aussi aux trajectoires que cette énergie rend impossibles avant qu'elles aient été délibérées. Une société peut toujours brûler une ressource rare pour produire un bénéfice immédiat. La question politique commence lorsque cette même ressource aurait pu ouvrir une trajectoire plus longue, et que la décision d'y renoncer n'a jamais été formulée comme une décision.

Ce que la situation actuelle masque, à mesure qu'elle se stabilise, est une décision politique majeure prise sans procédure politique. Microsoft, AWS, Google et leurs homologues décident, par leurs PPA long terme, de la trajectoire énergétique des deux prochaines décennies. Les juridictions locales corrigent à coups de moratoires et de blocages ce qu'elles n'ont jamais débattu en amont. L'opinion publique américaine bascule sans avoir été consultée. Le système fonctionne et s'érode simultanément.

Une société ne choisit pas si elle paie le coût matériel de l'IA. Elle choisit si ce coût est subi par défaut ou distribué selon des critères explicites. **Nous avons industrialisé la production d'intelligence artificielle. Nous n'avons pas encore industrialisé la décision de savoir quels usages méritent d'en consommer la matière.**

Références

- Article-mère : Vetillard, J. (2026). *La virtualité supposée de l'IA face au mur du réel*. Twingital Institute, édition révisée février 2026. <https://twingital-ventures.com/fr/publications/ia-energie-contraintes-physiques/>
- International Energy Agency (2026). *Energy and AI / Global Energy Review 2026*. Projection 1 100 TWh datacenters 2026, part IA 5-15 % aujourd'hui / 35-50 % à 2030.
- *Fortune* (2026-04-20). « Data centers now account for half of all new U.S. electricity use, just as Americans start to sour on AI ».
- Data Center Watch (initial report mai 2024-mars 2025 ; Q2 2025 update). 16 projets / 64 Md\$ pour la première période ; +20 projets / 98 Md\$ en Q2 2025 seul ; 48+ projets / >100 Md\$ pour le cumul 2025.
- Goldman Sachs Research (février 2026). Note sur l'impact des datacenters IA sur l'inflation. +0,1 pt core inflation US 2026 et 2027.
- UC Berkeley Center for Responsible Decentralized Intelligence (2026-04-12). *Trustworthy benchmarks (cont.)*.
- Stanford HAI (2026). *AI Index Report 2026*.
- Gartner (2026-04-07). « AI projects in I&O stall ahead of meaningful ROI returns ».
- NIST (2026-04-07). *Concept note for AI RMF Profile on Trustworthy AI in Critical Infrastructure*.
- European Parliament (2024). Regulation (EU) 2024/1689 (AI Act).
- Sorrell, S. (2009). « Jevons' Paradox revisited ». *Energy Policy*, 37(4).
- Kaack, L. H. et al. (2022). « Aligning AI with climate change mitigation ». *Nature Climate Change*, 12.