

Évaluer la sobriété d'un service numérique : proposition d'une grille d'évaluation multicritère adaptée au « *fuzzy front end* »

Robert Viseur¹

1 UMONS, Faculté Warocqué d'Économie et de Gestion

robert.viseur@umons.ac.be

Résumé

La digitalisation est moins une dématérialisation qu'un transfert de support : terminaux, réseaux, datacenters, fabrication et fin de vie portent une part « invisible » des impacts. Nous soutenons que l'enjeu n'est pas seulement d'optimiser, mais de définir ce qui mérite d'être produit et consommé. Nous proposons une grille de sobriété numérique (consommation et production) et discutons le cas des services d'IA générative.

Mots-clés

Sobriété numérique, Green IT, Green IS, IA générative.

Abstract

Digitalisation is less dematerialisation than a transfer of support: devices, networks, data centres, manufacturing and end-of-life account for much of the 'invisible' footprint. We argue that the key challenge is not only optimisation, but deciding what deserves to be produced and consumed. We propose a digital sobriety scoring framework (use and supply) and discuss the specific pressures introduced by generative AI services.

Keywords

Digital sobriety, Green IT, Green IS, GenAI.

1 Introduction

En 2021, le nombre de terminaux connectés s'élevait à 27,1 milliards d'unités [13], à comparer aux 13,4 milliards relevés en 2016, soit une augmentation de 102,24 % sur cinq ans. Dans cet ensemble, les *smartphones* représentent près de 25 % du matériel tandis que les objets connectés comptent pour plus de la moitié [13]. Bordage voit dès lors la digitalisation moins comme une dématérialisation que comme une forme de « *transfert de support* » [8]. Berthoud et Parry évoquent pour leur part des « *transferts de pollution* » (p. 16) [6]. Cette matérialité des TIC passe par les machines détenues par les particuliers et les organisations mais aussi celles associées aux services en ligne sur lesquels elles se connectent. Les services numériques questionnent donc quant à leur impact sur les écosystèmes naturels. Cette recherche, dédiée à l'évaluation de la sobriété d'un service numérique, répond dès lors à l'appel d'universitaires pour positionner les sciences de gestion dans l'ère de l'anthropocène [1]. S'inscrivant dans un naturalisme aménagé [1], elle propose, premièrement, de

documenter les impacts écologiques des services numériques, deuxièmement, de structurer une grille d'évaluation de la sobriété d'un service numérique, troisièmement, de discuter plus spécifiquement les impacts liés aux services d'intelligence artificielle (IA) générative.

2 Impact environnemental des équipements informatiques

Dans cette section, nous introduisons la problématique des impacts du numérique. Nous traitons ensuite de leur réduction et, enfin, de leur mesure.

2.1 Ordres de grandeur

Dans son rapport intitulé « *Déployer la sobriété numérique* », le *think tank* The Shift Project présente une ventilation de la consommation d'énergie finale du numérique [19]. La consommation d'énergie se répartit globalement entre l'utilisation (55 %) et la production (45 %) des équipements (cf. Tableau 1). Ces chiffres appellent quelques constatations. Premièrement, la consommation à l'utilisation se répartit de manière relativement équilibrée entre les terminaux, les réseaux et les centres de données. Près des deux tiers de la consommation énergétique, associée au traitement des données et à leur transfert, est donc peu ou prou visible pour l'utilisateur. Elle doit cependant être comptabilisée dans l'impact environnemental lié à l'utilisation d'un service numérique. Deuxièmement, près de la moitié de l'impact environnemental est déterminé par la production des équipements. Troisièmement, ces chiffres n'incluent pas la consommation associée au traitement en fin de vie.

Tableau 1. Consommation d'énergie par poste.

| | | | | |
|--------------------|-------------|------|------|---------|
| Utilisation | Terminaux | 20 % | 55 % | 36,36 % |
| | Réseaux | 16 % | | 29,09 % |
| | Datacenters | 19 % | | 34,55 % |
| Production | Ordinateurs | 17 % | 45 % | 37,78 % |
| | Smartphones | 11 % | | 24,44 % |
| | Télévisions | 11 % | | 24,44 % |
| | Autres | 6 % | | 13,33 % |

Ces chiffres se concentrent sur la consommation énergétique, qui ne reflète cependant qu'une partie de l'impact environnemental du numérique [17]. La production s'accompagne ainsi d'une consommation de

métaux, d'eau ainsi que de rejets de déchets et de gaz à effet de serre (GES). Quant au traitement en fin de vie, il se caractérise par un recyclage limité, entraînant la production de déchets, parfois exportés, ainsi que des rejets de polluants. Par ailleurs, ces équipements, et en particulier les *smartphones* [53], pèsent lourd dans la consommation de certains métaux. Or, pour la grande majorité des métaux, les réserves estimées, parfois géographiquement concentrées, se situent entre 30 et 60 ans [3]. De plus, le recyclage de ces métaux s'avère faible pour les biens numériques, car leur séparation est souvent difficile.

2.2 Réduction de l'impact environnemental

En pratique, la réduction de l'impact environnemental du numérique passe classiquement par deux mesures, d'une part, l'allongement de la durée de vie des équipements, d'autre part, l'optimisation de l'infrastructure technique [54]. La première concerne tant les consommateurs que les organisations, privées ou publiques, tandis que la seconde est essentiellement l'apanage des organisations. Ces dernières sont encouragées à réduire leur impact environnemental dès lors qu'elles supportent directement la consommation de ressources. Dans le cas où l'obésité logicielle est essentiellement supportée par l'utilisateur (cas des systèmes d'exploitation ou des suites bureautiques par exemple), l'incitation à optimiser pesant sur l'entreprise éditrice est cependant réduite, ce qui suppose une démarche volontaire de la part de celle-ci. Sur un plan financier, la gestion du Coût Total de Possession (TCO) des systèmes informatiques encourage l'achat de matériel fiable et réparable tandis que, du fait des cycles d'innovation (modèle A-U), les produits et services numériques matures tendent à privilégier l'innovation de processus, avec les économies de coûts de production associés. Dans ce contexte, réduire les coûts s'accompagne souvent d'une moindre consommation de ressources [47].

La lutte contre l'obsolescence des terminaux représente un enjeu important compte tenu de l'impact lié à la fabrication [8, 54]. Les types de machines doivent être distingués. Ainsi, sur le plan des émissions de CO₂, la fabrication d'un *smartphone*, d'un ordinateur portable et d'un serveur compte respectivement pour 72 %, 52 % et 20 % des émissions sur le cycle de vie de l'appareil [54]. Ces différences s'expliquent notamment par le taux d'utilisation des équipements, plus élevé pour les serveurs du fait notamment de l'usage de technologies comme la virtualisation. Les mesures favorables à l'allongement de la durée de vie du matériel sont cependant connues [38]. Elles incluent le réemploi (réparation et réutilisation de l'appareil ; p. ex. Back Market et ITJustGood anciennement Tradediscount), la réutilisation (réutilisation de sous-ensembles) et le recyclage (réutilisation des matières premières ; p. ex. Fairphone). La réparabilité, incluant une plus grande modularité des appareils [8], permet par ailleurs de faciliter le réemploi. Elle peut être favorisée par l'allongement des durées de garantie (mesure légale), mais aussi par une meilleure information au consommateur (voir, par exemple, les indices de réparabilité portés par iFixIt ou Spareka).

L'optimisation de l'infrastructure technique s'appuie sur

quatre leviers [42]. Le premier est le logiciel, incluant en particulier le modèle sous-jacent dans le cas des systèmes d'intelligence artificielle. Le second est le matériel permettant son exécution. Le troisième est le centre de données hébergeant le matériel. Le quatrième concerne les sources d'énergie utilisées par le centre de données. Ainsi, les entreprises gérant les grands centres de données veillent à décarboner l'alimentation en électricité, soit par leurs propres capacités de production, soit par la souscription de contrats d'approvisionnement d'électricité verte [29]. Les centres de données peuvent être, d'une part, localisés dans des zones froides, d'autre part, conçus pour abaisser les coûts de refroidissement. Les équipements informatiques peuvent être adaptés aux usages (p. ex. élimination des composants inutiles), et des puces spécialisées utilisées sur certaines tâches [55]. Enfin, les logiciels exécutés dans ces centres peuvent être optimisés, par le choix d'algorithmes et de langages de programmation plus efficaces [55].

Deux exemples permettent d'illustrer ces stratégies d'optimisation. Le premier cas, celui de Meta (avec Facebook), se traduit par l'innovation collaborative sur la conception des centres de données au sein de la plateforme [Open Compute Project](#). Créée en 2011, elle se définit comme « une communauté collaborative axée sur la refonte de la technologie matérielle afin de répondre efficacement aux demandes croissantes en matière d'infrastructure informatique ». Meta revendique une économie d'énergie de l'ordre de 38 % à la construction et une réduction du coût de fonctionnement de l'ordre de 24 %. De plus, après avoir utilisé PHP, un langage interprété, Meta a développé un traducteur de code source vers le C++, un langage compilé (projet open-source « HipHop for PHP »). Les limites de cette approche ayant été atteintes, malgré une réduction de 50 % de la charge processeur, un nouveau langage de programmation hybride baptisé [Hack](#) (projet open-source « HHVM ») a finalement été lancé. Le second cas, celui de Netflix, s'appuie plutôt sur une externalisation de son infrastructure auprès d'une entreprise d'hébergement dont c'est la spécialité (Amazon Web Services). Netflix concentre par contre ses ressources sur son cœur de métier (le *streaming* vidéo) et n'hésite pas à ouvrir sa R&D à d'autres partenaires. Publié en 2018, le codec [AV1](#) revendique ainsi un gain de bande passante de l'ordre de 30 % [27]. En outre, Netflix collabore avec les FAI via son programme [Open Connect](#). Cette politique s'inscrit par ailleurs dans le cadre d'un accord entre Netflix et le régulateur européen visant à modérer les volumes de données transférés [21].

2.3 Comprendre et mesurer l'impact environnemental des datacenters

Au-delà de l'empreinte écologique à la fabrication, les centres de données consomment généralement deux types de ressources : l'électricité et l'eau. Le refroidissement de machines peut recourir à différentes techniques [18]. Le refroidissement par détente directe, plus énergivore, recourant à la climatisation (groupe frigorifique), est généralement réservé aux petites unités modulaires. Le refroidissement gratuit (*free cooling*), très répandu, et moins énergivore, recoure, du plus économe au moins

économique en eau, soit à de l'air froid (rafraîchissement direct), soit à de l'air humidifié (rafraîchissement adiabatique), soit à de l'eau (rafraîchissement induit). Les développements récents se portent sur le refroidissement liquide direct où les processeurs sont par exemple immergés dans un fluide caloporteur. Dans le cas du rafraîchissement induit, l'impact environnemental dépend également de la provenance de l'eau. S'agit-il d'eau potable, d'eau usée ou d'eau de surface ? Dans le cas de Google, près de 80 % de l'eau prélevée est de l'eau de distribution (potable) mais la *bigtech* possède également des centres de données utilisant les eaux usées (Douglas, USA), l'eau de mer (Hamina, Finlande) ou l'eau de canal industriel (Saint-Ghislain, Belgique) (Google, 2025). Cette eau est ensuite rejetée (à une température supérieure) ou consommée (c'est-à-dire évaporée et réintroduite dans le grand cycle de l'eau [9]). La localisation du centre de données joue donc sur son impact. Premièrement, la présence d'un important réservoir (et d'un exutoire) d'eau (p. ex. mer) permet par exemple de le réduire. Globalement, la consommation des centres de données reste faible comparé à la consommation totale [40]. A contrario, la concentration des installations peut conduire à un impact local significatif, assorti de mesures visant à réduire la consommation. C'est par exemple le cas des centres situés le long de la Tamise à Londres [49]. Plus globalement, 14 % des prélèvements des centres de données de Google se déroulent dans des zones présentant un risque élevé de stress hydrique [22]. Deuxièmement, en fonction du mix énergétique, la disponibilité d'énergie décarbonée (énergie renouvelable, type solaire ou éolien, et nucléaire) permet d'en réduire sensiblement l'empreinte carbone. Là aussi, la concentration des grands centres de données peut venir perturber l'approvisionnement électrique. C'est notamment le cas en Irlande où des politiques d'atténuation des impacts ont été prises suite au succès des mesures incitatives à l'installation [49].

Dès lors, deux indicateurs principaux vont être suivis : le PUE et le WUE [18, 40]. Le PUE (*Power Usage Effectiveness*) est défini comme le rapport entre l'énergie totale annuelle consommée par le centre de données et l'énergie totale annuelle consommée par les seuls équipements IT. Le PUE moyen des centres de données industriels est de 1,58 contre 1,1 environ pour les fournisseurs d'infrastructures publiques de *cloud computing* [42]. La réduction du PUE passe notamment par le *free cooling*, soit que le centre de données soit refroidi naturellement par air froid, soit que l'eau soit utilisée comme fluide caloporteur. Le WUE (*Water Use Effectiveness*) est défini comme le ratio entre la quantité d'eau consommée par l'ensemble du centre de données et la consommation énergétique totale (L/kWh). Un indicateur supplémentaire peut être pris en compte pour évaluer la réutilisation de l'énergie. L'ERE (*Energy Reuse Effectiveness*) représente ainsi le rapport entre l'énergie totale consommée moins l'énergie réutilisée et l'énergie consommée par les équipements informatiques. Par exemple, Infomaniak a documenté son projet [D4](#), un modèle ouvert de centre de données où la chaleur est réutilisée pour assurer du chauffage urbain. Son installation

genevoise, qui vise une capacité de 10.000 serveurs, fournira un réseau de chaleur avec une capacité de 1,7MW, soit de quoi chauffer 6.000 logements (type Minergie-A) et éviter 3600 tCO₂eq chaque année. Au premier trimestre 2026, le PUE et le ERE de D4 tournaient respectivement autour de 1,1 et 0,25. Près de 90 % de l'énergie y était réutilisée tandis que le WUE se révélait très faible (pas d'utilisation d'eau additionnelle).

2.4 Mesurer les impacts environnementaux des services numériques

Plusieurs méthodes permettent de mesurer les impacts environnementaux d'un produit ou d'un service [6]. La première est le bilan carbone. Ce dernier ne se focalise pas contre que sur un seul indicateur : les émissions de gaz à effet de serre. Le second est l'analyse économique intrants-sortants. Les flux monétaires sont alors analysés pour estimer les ressources et l'énergie utilisées par un produit ou un service, puis pour évaluer les émissions. Troisièmement, l'analyse de cycle de vie vise à évaluer tous les impacts environnementaux potentiels d'un produit, d'un service ou d'un procédé sur l'ensemble de son cycle de vie.

La comptabilisation au sein d'un bilan carbone distingue en pratique trois « scopes » (définis au sein de la norme ISO 14064). Le « scope 1 » concerne les émissions directes de gaz à effet de serre. Le « scope 2 » concerne les émissions indirectes liées à l'énergie tandis que le « scope 3 » concerne les émissions indirectes de l'entreprise liées à ses achats. Dans le cas des centres de données, on distinguera ainsi les émissions directes du centre de données (« scope 1 »), les émissions liées à la production électrique nécessaire pour le faire fonctionner (« scope 2 ») et les émissions indirectes (« scope 3 ») liées, par exemple, à la production et la fin de vie des équipements informatiques.

L'Analyse de Cycle de Vie (ACV) se déroule en quatre phases [6]. La première phase consiste à définir les objectifs et le champ de l'étude. En particulier, les frontières du système y sont déterminées. La seconde phase porte sur l'inventaire du cycle de vie, ce qui inclut le recensement et la quantification des flux de matière. La troisième phase concerne l'évaluation de l'impact du cycle de vie. Elle produit des indicateurs permettant de traduire les impacts du système considéré. La quatrième phase est dédiée à l'interprétation du cycle de vie. Elle veille à rendre les résultats compréhensibles et à documenter les limites de l'étude. L'évaluation nécessite la disponibilité de données statistiques, associées à « leur lot d'incertitudes, d'approximations et d'hypothèses » (p. 4) [6]. Rassemblées en bases de données (p. ex. [EcoInvent](#)), traitées par des logiciels spécialisés (p. ex. [SimaPro](#)), leur traitement conduit à des écarts parfois importants. La mise en œuvre de l'ACV nécessite donc du temps et des expertises pointues.

En informatique, l'optimisation renvoie à l'ensemble des techniques, méthodes et algorithmes conçus pour réduire l'utilisation des ressources tout en maintenant ou améliorant les performances du système. Elle consiste plus précisément à analyser puis ajuster le logiciel et l'environnement d'exécution afin d'éliminer le travail

inutile et d'améliorer les performances [25]. Les gains d'efficacité ne conduisent pas nécessairement à la réduction des usages dès lors qu'ils s'accompagnent d'un effet rebond [51]. La notion d'effet rebond trouve son origine dans le paradoxe de Jevons, selon lequel l'amélioration de l'efficacité d'une technologie conduit paradoxalement à une augmentation de la consommation de la ressource concernée plutôt qu'à sa réduction [23]. La littérature distingue l'effet rebond direct (la réduction de prix, provoquée par celle du coût, entraîne un accroissement de la demande pour le produit moins cher) et l'effet rebond indirect (l'économie faite sur un produit induit l'accroissement de la consommation d'un autre produit). Nous nous concentrerons dans cette recherche sur l'effet rebond direct. La frugalité numérique suppose l'atteinte des mêmes résultats avec moins d'énergie en étant plus économe en ressources informatiques [54]. L'on peut dresser un parallèle avec l'innovation frugale [30], qui cherche à atteindre les performances attendues par le marché, avec des produits moins complexes, centrés sur les fonctionnalités cœurs, à moindre coût, de manière à atteindre une clientèle précédemment non solvable. A contrario la sobriété numérique s'oppose à la croissance et à la (sur-)consommation. Elle vise ainsi à réduire l'usage des technologies numériques, d'une part, en se limitant à ce qui est réellement utile, d'autre part, en promouvant la longévité des équipements existants [8, 51]. Tant la frugalité que la sobriété numérique peuvent s'appuyer sur des *low-tech*. L'innovation *low-tech* vise à « *réduire la taille, l'intensité et la complexité du système technique (...) [de manière à entraîner] une diminution de l'empreinte environnementale d'un objet particulier* » (p. 179) [34]. Comparativement à la frugalité, l'atteinte de la sobriété suppose un effort [26] de manière à éviter la production de nouveaux produits et services. Elle nécessite donc aussi un changement de comportement de consommateur, en particulier pour éviter l'effet rebond [34].

3 Construire une grille d'évaluation de la sobriété numérique

Quelle pourrait être la « *sobriété par design* », au sens de Franck Aggeri [2], d'un service numérique ? Ce que nous proposons de construire n'est pas un substitut aux démarches d'ACV, mais une grille d'évaluation, basée sur une échelle de notation, destinée à outiller la discussion au sein des équipes produit dès les premières phases de l'innovation (c'est-à-dire le *fuzzy front end*). Cette proposition répond à un angle encore peu couvert par la littérature : disposer d'un instrument suffisamment simple pour interroger le sens de l'innovation [2], sans exiger d'emblée la complexité, les données et l'expertise associées à un chiffrage précis de type ACV [6]. De tels chiffrages s'exposent par ailleurs à l'indisponibilité des chiffres ou à l'incomplétude des méthodologies de mesure ([37] ; cf. page 15). Sur le plan des usages, les impacts se révèlent différenciés si l'on compare, par exemple, d'une part, le Bitcoin (compensation par conception des gains d'efficacité matérielle) et la publicité ciblée (inflation dans la collecte de données à caractère personnel, accélération des flux

physiques), d'autre part, la vidéo en ligne et la visioconférence (réduction des déplacements ou des supports physiques) [55]. La littérature tend d'ailleurs à séparer les usages propices à la durabilité des optimisations des infrastructures numériques. Deux disciplines sont ainsi distinguées : le « Green IT » et le « Green IS ». Le terme « Green IT » renvoie à « *la réduction des effets environnementaux négatifs des technologies de l'information en utilisant et en éliminant les ressources informatiques de manière écoénergétique et rentable, en améliorant l'efficacité énergétique, en diminuant les émissions, ainsi qu'en réutilisant et en recyclant les matériaux* » tandis que le terme « Green IS » concerne « *la configuration et l'application des systèmes d'information afin d'atteindre des objectifs environnementaux, en réduisant l'empreinte écologique des entreprises et en soutenant la prise de décision organisationnelle vers la durabilité, tout en améliorant la performance économique* » (p. 938) [32]. Aligner durabilité (« *doing good* ») et profitabilité (« *doing well* ») pourrait ainsi relever du challenge pour les entreprises [58]. Nous distinguons dès lors deux types de critères, soit liés à la **consommation** (quel est l'objectif du service numérique ?), soit liés à la **production** (quelle est l'efficacité des infrastructures permettant au service de fonctionner ?).

3.1 Critères liés à la consommation

Premièrement, le numérique constitue un adjuvant pour la mise en œuvre d'une économie régénérative, non seulement capable de réduire les dommages sur les écosystèmes mais aussi de les restaurer et de les revitaliser [11]. Le numérique est ainsi vu comme un moyen de réduire l'empreinte environnementale des réseaux électriques, des transports, de l'agriculture ou de l'habitat [21]. Ces usages régénératifs coexistent cependant avec des usages extractivistes classiques tels que la publicité ciblée ou d'autres à visée récréative [55]. Un critère lié au **potentiel de régénération** peut dès lors être introduit. Dans quelle mesure le service rendu contribue-t-il aux durabilités environnementale, sociale et économique promues dans les Objectifs de Développement Durable ?

Deuxièmement, les services numériques présentent des conséquences en sens divers sur le plan des flux physiques induits par la technologie. Le numérique est souvent associé à une accélération des flux physiques. Peut ainsi être citée la publicité ciblée qui entraîne non seulement un traitement massif de données mais aussi une consommation accrue de biens physiques [55]. A contrario, certains usages relèvent d'une « *sobriété par le numérique* » [55], c'est-à-dire d'un changement de paradigme du physique vers le numérique [33]. Ainsi, et paradoxalement, la vidéo en ligne ou la visioconférence présentent, malgré la sollicitation des infrastructures numériques, une réduction substantielle des flux physiques. Idéalement, la tendance à la hausse ou la baisse doit cependant être estimée par ACV dès lors que toute forme de dématérialisation n'est pas nécessairement intéressante sur le plan des impacts environnementaux. Berthoud et Parry démontrent ainsi les limites de la dématérialisation dans les cas de l'affranchissement de colis (affranchissement traditionnel vs affranchissement

dématérialisé) et de la lecture de documents (feuilles papier vs clefs USB) [6]. Cela nous permet d'introduire comme critère le **potentiel de substitution**. Le service numérique évalué conduit-il de manière directe à une réduction ou à une augmentation des flux physiques ?

Troisièmement, les services numériques peuvent augmenter les flux physiques de manière directe, au travers des acquisitions de matériel, soit que le matériel courant devienne obsolète, soit que le service nécessite l'achat d'objets connectés. C'est le critère des **consommations induites**, soit le pendant des investissements induits côté fournisseur. Ce critère s'attellera à déterminer si le nouveau service contribue à l'obsolescence de terminaux existants (voir par exemple la polémique autour de l'incompatibilité à Windows 10) ou à l'achat d'équipements complémentaires incluant des objets connectés ou des périphériques [24]. Or, les impacts liés aux objets connectés, dont le nombre croît rapidement, reste sous-exploré, en particulier dès lors qu'il concerne leur fabrication ou leur démantèlement [21].

Quatrièmement, les services numériques s'accompagnent souvent d'un effet rebond, en particulier dès lors que sa production est optimisée et en facilite la disponibilité, notamment sur un plan financier [20]. Cependant, le concepteur peut faire le choix de déployer des mécanismes incitatifs, soit pour encourager l'usage du service, soit pour en limiter l'utilisation au strict nécessaire [55]. Les interfaces peuvent ainsi encourager les usages compulsifs à l'aide de *dark patterns* ou de designs cornucopiens [5], ou au contraire pousser à la modération des usages à l'aide de *green nudges* [44, 57]. Ceux-ci peuvent passer par des indicateurs, soit consécutifs (si l'impact est directement proportionnel à la consommation), soit attributifs (si l'impact est extrapolé depuis un calcul par unité fonctionnelle). Dans le second cas, la corrélation entre consommation et impact peut être faible. C'est notamment le cas des volumes de données où la consommation des équipements est essentiellement déterminée à leur mise sous tension. Cependant, même si l'indicateur est techniquement faux, il fournit une incitation utile à la tempérance des usages qui permet de limiter la pression sur les capacités des infrastructures. Ces dimensions sont couvertes par le critère de **tempérance des usages**.

3.2 Critères liés à la production

Premièrement, le déploiement d'un nouveau service numérique pose la question des **investissements induits**, soit le pendant des consommations induites côté utilisateur. Peut-il fonctionner sur des équipements pré-existants ou nécessite-t-il l'investissement dans de nouvelles capacités ? L'utilisation de grands modèles de langage s'accompagne ainsi d'investissements massifs en capacité de calcul destinées à l'entraînement des modèles et à l'inférence [28], ce qui occasionne des tensions d'approvisionnement [41].

Deuxièmement, le développement des services en ligne en général, et plus récemment des fournisseurs de *cloud computing* (Amazon Web Service, Google Cloud, Microsoft Azure...) a vu le développement des centres de données [55]. Se pose donc la question de l'**optimisation des infrastructures**. Cette optimisation passe par, d'une

part, le matériel utilisé dans le centre de données, incluant le taux d'utilisation des machines et leurs performances [4], d'autre part, la conception du centre de données, incluant les dispositifs de refroidissement, son alimentation en énergie ou encore sa localisation [40, 42, 46].

Troisièmement, avoir des infrastructures efficaces n'est pleinement satisfaisant que si les logiciels qu'elles font tourner ont eux-mêmes été optimisés. Ce critère est donc relatif à l'**optimisation des logiciels**. Cette dernière passe par le choix de langage de programmation efficaces [43], généralement compilés (C, C++, Rust, Go...) ou hybrides (Java, Scala...), la sélection d'algorithmes performants [14], mais aussi la mise en place de tests de performance [7] ou encore la limitation du périmètre fonctionnel du logiciel [5, 8].

Quatrièmement, la réduction du poids des infrastructures passe par un **traitement raisonné des données** [55]. Plus de données signifie davantage d'espace de stockage, davantage de capacités de calcul et des mises à jour des infrastructures réseaux. Limiter la taille des données traitées permet donc de travailler à équipements constants (terminaux, réseaux, centres de données), dès lors d'éloigner la nécessité de renouvellement des matériels ou d'accroissement des infrastructures.

Cinquièmement, les modèles d'intelligence artificielle deviennent un sujet d'inquiétude sur le plan environnemental. Schwartz et ses co-auteurs opposent ainsi ce qu'ils appellent la « Red AI », caractérisée par des coûts prohibitifs, un impact environnemental négatif et une élévation des barrières à l'entrée pour la recherche, et la « Green AI », plus inclusive et respectueuse de l'environnement [45]. Les poids environnementaux et financiers sont les deux versants d'une même pièce. En effet, les modèles, de plus en plus grands, nécessitent des capacités croissantes de calcul pour leur entraînement et leur utilisation, ce qui pousse ainsi à l'utilisation de modèles plus spécialisés et plus petits [56]. Les chercheurs appellent ainsi à moins se focaliser sur la précision et davantage sur l'efficacité des modèles. Cela permet d'introduire un dernier critère : la **frugalité du modèle d'IA**. Le service numérique s'appuie-t-il sur un modèle connexionniste de grande taille ou sur un modèle de *machine learning* plus classique ? L'utilisation de modèles au plus juste est-elle encouragée ?

4 Analyser la sobriété d'un portefeuille de services numériques

Les critères ci-dessus permettent de composer une grille synthétique où les deux branches (consommation, production) sont ensuite décomposées en critères évaluables sur une échelle de Likert [-2, +2] documentée. Associés à une échelle de notation (cf. « [Annexe 1](#) »), ces critères, combinés à une pondération ajustable, permettent de construire une grille permettant d'obtenir un score global de sobriété numérique, associé à des scores partiels liés à la dimension « consommation » et à la dimension « production » du service numérique. L'exemple ci-dessous porte sur l'espace de stockage en ligne kDrive (cf. Tableau 2), avec une pondération des critères égale à 1.

Tableau 2. Évaluation de la sobriété numérique du service numérique kDrive (Infomaniak).

| Critère | Score | Justification |
|---|-------|---|
| kDrive est un espace de stockage en ligne basé sur un <i>cloud</i> souverain suisse performant, prolongeant le matériel utilisateur, mais sans dispositif limitant automatiquement l'usage. | | B |
| Consommation | | B |
| Potentiel de régénération | 0 | Potentiel régénératif limité. Impact positif sur la collaboration, et amélioration de la souveraineté sur les données (base open-source, respect du RGPD et du nLPD). |
| Potentiel de substitution | 1 | Réduction du besoin en disques externes individuels. Limitation des impressions et doublons. Moindre transport physique de documents. |
| Consommations induites | 1 | Pas de nouveau matériel requis, support multi-OS. Besoin moindre en stockage local, ce qui peut prolonger la durée de vie des équipements existants. |
| Tempérance des usages | 0 | Interface neutre ; aucun <i>dark pattern</i> ni fonctionnalité limitant automatiquement l'usage. |
| Production | | A |
| Investissements induits | 0 | Utilisation de centre de données existants ou mutualisation de plusieurs services, d'où un investissement supplémentaire marginal. |
| Optimisation des infrastructures | 2 | Effort dans la conception des centres de données (cf. D4) : faible PUE, refroidissement par air extérieur, utilisation d'énergie renouvelable, réutilisation de la chaleur. |
| Optimisation des logiciels | 1 | Code optimisé (éco-conception), compression automatique, profilage continu, dette technique maîtrisée selon documentation interne. |
| Traitement raisonné des données | 1 | Volumes dictés par les utilisateurs. Conservation paramétrable. Pas de collecte systématique mais données potentiellement volumineuses. |
| Frugalité du modèle IA | 2 | Aucun modèle IA gourmand. Aucune nécessité de processeurs spécialisés. |

Tableau 3. Quadrant magique de la sobriété numérique.

| | | Sobriété apparente | Sobriété exemplaire |
|---------------------|-----------------|--|---|
| Consommation | Usage tempéré | Expérience utilisateur vertueuse masquant une empreinte infrastructurelle lourde. | Idéal-type caractérisé par la limitation des usages et la frugalité de l'infrastructure. |
| | Usage compulsif | Voracité numérique Usages et infrastructures cumulativement problématiques avec un impact global défavorable. | Frugalité paradoxale Système technique optimisé mais usages compulsifs, soit une sobriété partielle, contrastée. |
| | | Faible optimisation | Forte optimisation |
| | | Production | |

Avec un score de consommation égal à 0,5 (B) et un score de production égal à 1,5 (A), kDrive ressort comme un

service numérique plutôt sobre. Les scores de consommation et de production, suivant qu'ils sont en moyenne inférieurs ou supérieurs à zéro, permettent de positionner chaque service numérique évalué dans un quadrant (cf. Tableau 3). Ce quadrant permet par exemple de visualiser l'état d'un portefeuille de services numériques au sein d'une organisation de manière à identifier les priorités d'amélioration. Cet outil reprend la logique similaire et complémentaire à celle adoptée par Sneed [48] pour planifier les opérations d'amélioration sur des systèmes hérités (priorisation fonction du caractère critique de l'application et de sa qualité, notamment structurelle).

La situation, d'une part, côté consommation, d'autre part, côté production, permet d'identifier quatre quadrants. Le plus problématique, soit la **voracité numérique**, se caractérise par des usages compulsifs et une lourde infrastructure. Le plus vertueux, soit la **sobriété exemplaire**, ressort comme un idéal-type de la sobriété numérique, marqué par la sobriété des usages et la frugalité des infrastructures utilisées. Deux cas intermédiaires sont présentés : la **sobriété apparente** et la **frugalité paradoxale**. Pour le premier, les usages sont limités mais l'infrastructure devrait être davantage optimisée. Dans le second, l'usage est compulsif mais est partiellement compensé par l'efficacité du système technique sous-jacent.

5 Cas particulier des services numériques basés sur l'IA générative

Étudier l'impact environnemental des IA génératives nécessite de choisir des indicateurs. Deux sont couramment utilisés : les émissions de CO₂ (bilan carbone) et la consommation d'eau. Combien tel ou tel modèle a-t-il provoqué d'émissions de CO₂ pour son entraînement ? La découpe en « scopes » (« scope 1 », « scope 2 », « scope 3 ») se retrouve tant pour le calcul du bilan carbone que pour l'estimation des impacts sur la ressource « eau » [9]. Ainsi, l'entraînement du modèle GPT-3 aurait nécessité 1,3 GWh, soit la consommation annuelle moyenne de 120 foyers étasuniens, occasionnant l'émission de 522 tonnes de CO₂ [50]. Quant à BLOOM, l'énergie consommée pour son entraînement est estimée à 433 MWh [39]. Concernant l'eau, Li et ses co-auteurs attribuent ainsi à une conversation avec ChatGPT une consommation de l'ordre de 50cl [35]. Plus précisément, ce chiffre de 50cl y représente la quantité d'eau nécessaire, dans le « scope 1 » et le « scope 2 », pour entraîner le modèle puis produire les réponses à un nombre de *prompts* compris entre 10 et 70 requêtes (cet intervalle de valeurs a légèrement évolué au gré des versions successives de la prépublication). Cette valeur peut être comparée au chiffre de 0,26 millilitres (soit environ 5 gouttes d'eau) revendiqué par Google pour son modèle Gemini [16]. Cette différence s'explique par trois facteurs. Premièrement, le calcul de la consommation ne couvre que l'inférence. Cela s'explique notamment par le fait que l'étude de Elsworth et ses co-auteurs [16] vise aussi, sur un plan plus marketing que scientifique, à mettre en évidence l'efficacité revendiquée des puces TPU conçues par Google comme alternative aux GPU de NVIDIA, le fournisseur dominant actuellement le marché

des puces dédiées à l'IA. Deuxièmement, leur étude se restreint au « scope 1 » pour le calcul de la consommation en eau (au contraire de l'étude de Li *et al.* [35], prenant en compte le « scope 1 » et le « scope 2 »). Troisièmement, alors que l'étude de Li et ses co-auteurs concerne GPT-3 [35], un modèle sorti en 2020, celle de Elsworth et ses co-auteurs [16] exploite les données internes relatives aux modèles les plus récents, sur lesquels des optimisations ont été réalisées pour en améliorer l'efficacité. Le secteur de l'IA générative est en effet marqué par les économies d'échelle [56]. Les coûts financiers liés à l'entraînement doivent être amortis sur une base d'utilisateurs la plus large possible (plus de 700 millions d'utilisateurs actifs pour ChatGPT [12]) car il s'agit d'un coût fixe. Les coûts financiers associés à l'inférence constituent par contre un coût variable : plus la base d'utilisateurs augmente, et plus le coût d'inférence l'emporte sur le coût d'entraînement. Google met donc en évidence l'innovation typique d'une phase transitionnelle visant à réduire le coût de production du service plutôt que d'en améliorer les performances pour l'utilisateur [55].

Ces deux exemples permettent par ailleurs d'interroger la provenance des données nécessaires à ce genre de comparaison. Les rapports environnementaux des *bigtechs* (p. ex. Google, Meta et Microsoft) sont riches en informations. Cependant, la transparence totale y est peu ou prou faite sur le coût environnemental des modèles. Les analyses d'impact s'appuient donc sur trois stratégies, non sans conséquence sur leur fiabilité. Premièrement, le calcul d'impact peut être réalisé par des chercheurs indépendants, suite à un lent travail de collecte de données et d'approximations, permettant d'élaborer un modèle de calcul de coûts, bien documenté mais nécessairement imprécis. Deuxièmement, le calcul d'impact peut être estimé par des équipes liées aux industriels (Google, Microsoft, OpenAI...). Cette configuration conduit à un accès privilégié à des données internes, ce qui autorise une analyse actualisée, plus proche de la réalité, mais aussi à de possibles conflits d'intérêt, par exemple pour présenter certaines informations sous un jour plus favorable. Troisièmement, le calcul d'impact peut être réalisé sur des modèles ouverts développés au sein de consortiums de recherche. La transparence est dès lors de mise. Cependant, ce type de modèle n'est pas soumis aux contraintes, notamment d'optimisation, d'un modèle réellement utilisé en production. L'analyse d'impact de GPT-3 [35, 36] correspond à la première configuration, celle de GLaM [42] ou de Gemini [16], à la seconde, et celle de BLOOM [39], à la troisième. L'évaluation chiffrée des impacts se révèle donc complexe du fait, d'une part, du manque de chiffres [37], d'autre part, des difficultés de comparaison.

Sur le plan de la consommation du service, le potentiel de substitution des IA génératives mériterait un examen attentif dès lors que l'automatisation de la production de textes ou d'images pourrait, dans certains contextes, conduire à des économies de ressources [52]. Le secteur de la production cinématographique, qui envisage déjà la production virtuelle comme moyen de réduire les transports, le temps de tournage et les besoins logistiques [31], pourrait par exemple bénéficier de modèles génératifs

performants dédiés à la production de vidéos. Force est cependant de constater que les modèles génératifs font davantage naître des espoirs de gains de productivité, par exemple en programmation, même si ces derniers tardent à se concrétiser dans la pratique [10].

Sur le plan de la production du service, l'IAG appelle une attention particulière en matière d'investissements induits, d'optimisation des infrastructures et de frugalité des modèles IA. Premièrement, l'essor de l'IA générative entraîne de nombreux investissements pour l'acquisition de puces permettant l'entraînement et l'inférence [41]. Ainsi OpenAI est par exemple engagé dans un partenariat stratégique avec NVIDIA tandis qu'il a annoncé acheter pour 10 milliards de dollars de puces auprès de Cerebras. Deuxièmement, et même si l'optimisation des infrastructures fait partie des bonnes pratiques sectorielles, l'essor de l'IA connexionniste fait naître de nouveaux besoins en capacités de stockage et, surtout, de calcul, d'une part, pour l'entraînement des modèles, d'autre part, pour leur utilisation (inférence) au sein de systèmes techniques (agents conversationnels, moteurs de recherche ou applications spécialisées). L'IEA prévoit ainsi une croissance régulière de la consommation électrique des centres de données, autour de 15 % par an (ce qui aboutirait à un doublement de la consommation après 5 ans). Par ailleurs, le renforcement de capacités existantes conduit à de fortes disparités locales ou régionales [28]. Conséquence de leur concentration et de leur gigantisme, les installations les plus massives pèsent ainsi sur les infrastructures d'approvisionnement en électricité et en eau potable. Ainsi, la municipalité de West Des Moines, où Microsoft possède un centre de données sur lequel a été entraîné GPT-4, et ses services de distribution d'eau potable conditionnent l'approbation de tout futur centre de données à l'adoption de technologies réduisant drastiquement les pics de consommation d'eau par rapport aux projets actuels de Microsoft. Troisièmement, la réduction des impacts est possible à condition d'utiliser des modèles frugaux. Les grands modèles de langage (LLM, *Large Language Models*) coûtent cher, tant pour l'entraînement des modèles (p. ex. GPT) que lors de leur utilisation (inférence) au sein d'une application (p. ex. ChatGPT). L'optimisation des modèles peut donc porter sur l'entraînement (p. ex. nettoyage des jeux de données, réduction du nombre de paramètres et réduction du poids des paramètres) et l'inférence (p. ex. exécution partielle des modèles, utilisation de modèles plus petits pour des besoins plus simples et utilisation de puces spécialisées comme les LPU). Actuellement, trois types de modèles coexistent : les LLM (*Large Language Models*), les LRM (*Large Reasoning Model*) et les SLM (*Small Language Models*). Les LRM sont des LLM, généralement de grande taille, et optimisés pour le raisonnement. Ce sont les vaisseaux amiraux de modèles proposés. À l'opposé, les SLM sont des LLM plus légers, avec peu de paramètres, adaptés à la réalisation de tâches spécifiques. L'optimisation de l'impact environnemental des IA génératives passe ainsi par l'utilisation d'un modèle au plus juste (p. ex. routeur dans GPT-5) et la spécialisation de modèles. Les IAG se développent ainsi selon deux trajectoires [56]. La première

comporte des IA génératives généralistes (p. ex. ChatGPT) proposant souvent l'accès à deux types de modèles (un LLM et un LRM ; p. ex. GPT-4o et o1 ou DeepSeek-V3 et DeepSeek-R1 : cf. [15]), lourdes mais sensibles aux économies d'échelle (logique « *winner-takes-all* »). La seconde comporte des IA spécialisées, notamment basées sur des modèles ouverts (open-weight, open-source), plus rudimentaires mais sensiblement moins lourds. Par exemple, Perplexity a progressivement construit une expertise interne en matière d'hébergement de modèles ouverts (API interne « *pplx-api* ») puis de *finetuning* (modèles Sonar basé sur des modèles ouverts Mistral, Llama ou DeepSeek). Perplexity s'appuie sur l'infrastructure *cloud* d'Amazon Web Service ainsi que sur des prestataires spécialisés comme Cerebras, en plus de disposer d'un soutien technique de NVIDIA. Cet investissement s'est fait progressivement et a abouti à la famille de modèles Sonar, proposés comme LLMAaaS mais aussi utilisés en interne. En particulier, l'utilisation de modèles orientés « réponses factuelles » a permis une réduction des coûts d'inférence sans perte de satisfaction des utilisateurs comparativement aux modèles propriétaires précédemment utilisés.

Tableau 4. Évaluation de la sobriété numérique du service numérique Euria (Infomaniak).

| Critère | Score | Justification | |
|---|-------|---|----------|
| Euria est un agent conversationnel associé à la suite collaborative en ligne kDrive basé sur des LLM ouverts. | | | C |
| Consommation | | | B |
| Potential de régénération | -2 | Potential régénératif absent. Amélioration de la productivité et substitution de services existants. | |
| Potential de substitution | -1 | Optimisation de la recherche d'information mais nouvel usage numérique intensif. | |
| Consommations induites | 0 | Fonctionnement en ligne sur des terminaux existants. | |
| Tempérance des usages | 0 | Interface neutre ; aucun <i>dark pattern</i> ni fonctionnalité limitant automatiquement l'usage. | |
| Production | | | C |
| Investissements induits | -1 | Utilisation de centre de données existants mais nécessité de capacités GPU supplémentaires. | |
| Optimisation des infrastructures | 2 | Effort dans la conception des centres de données (cf. D4). | |
| Optimisation des logiciels | 1 | Peu d'informations publiques sur l'éco-conception logicielle spécifique d'Euria. Bonnes pratiques standards. | |
| Traitement raisonné des données | -1 | Minimisation de la collecte de données (p. ex. données à caractère personnel) mais pas d'absence de données massives. | |
| Frugalité du modèle IA | -1 | Orchestration au plus juste de grands modèles de langage. | |

Le Tableau 4 propose une évaluation de la sobriété numérique du service [Euria](#), un agent conversationnel basé sur des modèles ouverts tels que Qwen, intégré à la suite collaborative [kDrive](#) (cf. Tableau 2).

6 Conclusion

La croissance des terminaux et la part « invisible » des impacts (réseaux, *datacenters*, fabrication et fin de vie) confirment que la digitalisation relève souvent d'un transfert de support plutôt que d'une dématérialisation. Dès lors, l'enjeu central n'est pas uniquement d'optimiser l'existant, mais d'abord de définir ce qui mérite d'être produit et consommé : quelles fonctions sont réellement utiles, quels usages doivent être encouragés, lesquels doivent être évités. Dans cette perspective, cette recherche propose une grille d'évaluation (par critère de notation) de la sobriété numérique mobilisable en amont de l'innovation (*fuzzy front end*), structurée en deux dimensions : consommation (caractère régénératif, substitution des flux physiques, consommations induites et tempérance des usages) et production (investissements induits, optimisation des infrastructures et des logiciels, minimisation des données et frugalité du modèle d'IA). Le quadrant associé facilite la comparaison entre services et la priorisation des actions, notamment face aux situations de voracité numérique. Les résultats soulignent que l'efficacité ne garantit pas la sobriété : sans tempérance, l'optimisation alimente l'effet rebond. L'IA générative cristallise ce risque : malgré des gains techniques, elle induit souvent des investissements en calcul et des pressions locales sur l'énergie et l'eau. D'où l'intérêt de privilégier le modèle au plus juste, des architectures permettant de réduire la taille des modèles (p. ex. RAG) et des infrastructures performantes, afin d'aligner l'innovation numérique avec les limites planétaires. L'outil proposé doit encore faire l'objet d'une validation du choix, de la couverture et de l'indépendance des critères. Sa mise à l'épreuve dans des cas d'usage concrets permettrait en outre de préciser les modalités d'application de la grille et de réduire la part de subjectivité dans l'évaluation.

7 Références

- [1] Acquier, A., Mayer, J., & Valiorgue, B. (2024). Introduction. Anthropocène, limites planétaires et nouvelles frontières des sciences de gestion. *Revue Française de Gestion*, 315(2), 11-36. <https://doi.org/10.1684/rfg.2024.18>.
- [2] Aggeri, F. (2023). Quelles innovations pour un monde de post-croissance ? *L'Économie politique* 2023/2 N° 98, Éditions Alternatives économiques, pp. 42-57. https://shs.cairn.info/article/LECO_098_0042/pdf.
- [3] Aurez, V. et Georgeault, L. (2016). Économie circulaire : système économique et finitude des ressources. De Boeck Supérieur. ISBN : 978-2-8073-0148-1.
- [4] Barroso, L. A., & Hölzle, U. (2007). The case for energy-proportional computing. *Computer*, 40(12), 33-37. <https://doi.org/10.1109/MC.2007.443>.
- [5] Beignou, A., Tabard, A., & Maudet, N. (2026, April). How Interface Design Choices Lead to Indirect Environmental Impacts Through Use Intensification. In CHI Conference on Human Factors in Computing Systems. <https://dx.doi.org/10.1145/3772318.3790305>.

- [6] Berthoud, F., & Parry, M. (2010). Évaluation des impacts environnementaux de l'informatique. Quels outils ? Quelles limites ? Terminal. Technologie de l'information, culture & société, (106-107). <https://doi.org/10.4000/terminal.1794>.
- [7] Bondi, A. B. (2015). Foundations of software and system performance engineering: process, performance modeling, requirements, testing, scalability, and practice. Pearson Education. ISBN : 978-0-321-83382-2.
- [8] Bordage, F. (2019). Sobriété numérique : Les clés pour agir. Buchet-Chastel. ISBN : 978-2283032152.
- [9] Bouveret, S., Bugeau, A., Orgerie, A. C., & Quinton, S. (2024). De l'eau dans les nuages. Annales des Mines-Enjeux Numériques, 27(3), 41-48. <https://doi.org/10.3917/ennu.027.0041>.
- [10] Calvino, F., Reijerink, J., & Samek, L. (2025). The effects of generative AI on productivity, innovation and entrepreneurship (OECD Artificial Intelligence Papers, No. 39). OECD Publishing. <https://doi.org/10.1787/b21df222-en>.
- [11] Čegar, S., Drezgić, S., & Čišić, D. (2024). Exploring the Regenerative Economy: A Comprehensive Review of Literature. Journal of Regenerative Economics, 1(1), 1-40. <https://crossreis.com/journal/JRE-Vol1-Issue1-2024.pdf>.
- [12] Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). How people use chatgpt (No. w34255). National Bureau of Economic Research. <https://doi.org/10.3386/w34255>.
- [13] Cisco (2021). VNI Complete Forecast Highlights. https://www.cisco.com/Global_2021_Forecast_Highlights.pdf.
- [14] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). Introduction to algorithms. MIT press. ISBN : 978-0-262-03384-8.
- [15] Deng, Z., Ma, W., Han, Q. L., Zhou, W., Zhu, X., Wen, S., & Xiang, Y. (2025). Exploring DeepSeek: A Survey on Advances, Applications, Challenges and Future Directions. IEEE/CAA Journal of Automatica Sinica, 12(5), 872-893. <https://doi.org/10.1109/JAS.2025.125498>.
- [16] Elsworth, C., Huang, K., Patterson, D., Schneider, I., Sedivy, R., Goodman, S., ... & Manyika, J. (2025). Measuring the environmental impact of delivering AI at Google Scale. arXiv preprint arXiv:2508.15734. <https://doi.org/10.48550/arXiv.2508.15734>.
- [17] Faucheux, S., Hue, C., et Nicolaï, I. (2010). TIC et développement durable : Les conditions du succès. De Boeck. ISBN : 978-2804103712.
- [18] FEMP (2024). Best Practices Guide for Energy-Efficient Data Center Design. Federal Energy Management Program, 26 juillet 2024. <https://www.energy.gov/femp/articles/best-practices-guide-energy-efficient-data-center-design>.
- [19] Ferreboeuf, H., Efoui-Hess, M., Marraud, L. et Lescop, C. (2020). Déployer la sobriété numérique. The Shift Project. https://theshiftproject.org/.../2020/10/Deployer-la-sobriete-numerique_Resume_ShiftProject.pdf.
- [20] Flipo, F., & Gossart, C. (2009). Infrastructure numérique et environnement. L'impossible domestication de l'effet rebond. Terminal. Technologie de l'Information, Culture & Société, (103-104). <https://doi.org/10.4000/terminal.3093>.
- [21] Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G. S., & Friday, A. (2021). The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. Patterns, 2(9). <https://doi.org/10.1016/j.patter.2021.100340>.
- [22] Google (2025). 2025 Environmental Report. Google, Juin 2025. <https://sustainability.google/reports/google-2025-environmental-report/>.
- [23] Gossart, C. (2015). Rebound Effects and ICT: A Review of the Literature. In: Hilty, L., Aebischer, B. (eds) ICT Innovations for Sustainability. Advances in Intelligent Systems and Computing, vol 310. Springer, Cham. https://doi.org/10.1007/978-3-319-09228-7_26.
- [24] Gould, P., Song, G., & Zhu, T. (2024). Environmental and Economic Impact of I/O Device Obsolescence. arXiv preprint arXiv:2412.20655. <https://doi.org/10.48550/arXiv.2412.20655>.
- [25] Gregg, B. (2014). Systems performance: enterprise and the cloud. Pearson Education. ISBN : 978-0-13-339009-4.
- [26] Guillard, V. (2021). Towards a society of sobriety: Conditions for a change in consumer behavior. Field Actions Science Reports. The Journal Of Field Actions, (Special Issue 23), 36-39. <https://journals.openedition.org/factsreports/pdf/6590>.
- [27] Han, J., Li, B., Mukherjee, D., Chiang, C. H., Grange, A., Chen, C. et Bankoski, J. (2021). A technical overview of AV1. Proceedings of the IEEE, 109(9), 1435-1462. <https://doi.org/10.1109/JPROC.2021.3058584>.
- [28] IEA (2025b), Energy and AI, International Energy Agency. <https://www.iea.org/reports/energy-and-ai>.
- [29] IEA (2025a). Data Centres and Data Transmission Networks, International Energy Agency. <https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks>.
- [30] Kammoun, S., & Le Bas, C. (2020). L'innovation frugale peut-elle être disruptive?. Innovations, 63(3), 179-199. <https://doi.org/10.3917/inno.063.0179>.
- [31] Khairulbahri, M. (2025). The application of the system archetypes in understanding GHG emissions: A case study in the US movie industry. Journal of Cleaner Production, 505, 145419. <https://doi.org/10.1016/j.jclepro.2025.145419>.
- [32] Kotlarsky, J., Oshri, I., & Sekulic, N. (2023). Digital sustainability in information systems research: Conceptual foundations and future directions. Journal of the Association for Information Systems, 24(4), 936-952. <http://doi.org/10.17705/1jais.00825>.
- [33] Kurp, P. (2008). Green computing. Communications of the ACM, 51(10), 11-13.

<https://doi.org/10.1145/1400181.1400186>.

[34] Le Bas, C. (2024). Les approches « low-tech » de l'innovation: une analyse critique des narratifs. *Revue d'Économie Industrielle*, 186(2), 175-197. <https://doi.org/10.3917/rei.186.0175>.

[35] Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). Making ai less 'thirsty'. *Communications of the ACM*, 68(7), 54-61. <https://doi.org/10.1145/3724499>.

[36] Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models. *arXiv 2023. arXiv preprint arXiv:2304.03271*. <https://doi.org/10.48550/arXiv.2304.03271>.

[37] Ligozat, A.-L., & Bugeau, A. (2025). Méthodes d'évaluation de l'empreinte de l'IA [Conference presentation]. *GreenDays 2025*, Rennes, France. HAL. <https://hal.science/hal-05317063>.

[38] Lombès, T., & Poubeau, B. (2014). Obsolescence programmée: mythes et réalité. *Mines Paristech*. [http://www.annales.org/\(...\)/Livre-Obsolescence-\(...\).pdf](http://www.annales.org/(...)/Livre-Obsolescence-(...).pdf).

[39] 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. <https://doi.org/10.48550/arXiv.2211.02001>.

[40] Mytton, D. (2021). Data centre water consumption. *npj Clean Water*, 4(1), 11. <https://doi.org/10.1038/s41545-021-00101-w>.

[41] OECD (2025). Competition in artificial intelligence infrastructure (OECD Roundtables on Competition Policy Papers, No. 330). OECD Publishing. <https://doi.org/10.1787/20758677>.

[42] Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L. M., ... & Dean, J. (2022). The carbon footprint of machine learning training will plateau, then shrink. *Computer*, 55(7), 18-28. <https://doi.org/10.1109/MC.2022.3148714>.

[43] Pereira, R., Couto, M., Ribeiro, F., Rua, R., Cunha, J., Fernandes, J. P., & Saraiva, J. (2017). Energy efficiency across programming languages: how do energy, time, and memory relate?. In *Proceedings of the 10th ACM SIGPLAN international conference on software language engineering* (pp. 256-267). <https://doi.org/10.1145/3136014.3136031>.

[44] Schubert, C. (2017). Green nudges: Do they work? Are they ethical?. *Ecological economics*, 132, 329-342. <https://doi.org/10.1016/j.ecolecon.2016.11.009>.

[45] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54-63. <https://doi.org/10.1145/3381831>.

[46] Sheme, E., Holmbacka, S., Lafond, S., Lučanin, D., & Frashëri, N. (2018). Feasibility of using renewable energy to supply data centers in 60° north latitude. *Sustainable Computing: Informatics and Systems*, 17, 96-106. <https://doi.org/10.1016/j.suscom.2017.10.017>.

[47] Sinkin, C., Wright, C. J., & Burnett, R. D. (2008). Eco-

efficiency and firm value. *Journal of accounting and public policy*, 27(2), 167-176. <https://doi.org/10.1016/j.jaccpubpol.2008.01.003>.

[48] Sneed, H. M. (1995). Planning the reengineering of legacy systems. *IEEE software*, 12(1), 24-34. <https://doi.org/10.1109/52.363168>.

[49] Soares, I. V., Yarime, M., & Klemun, M. (2024). Balancing the trade-off between data center development and its environmental impacts: A comparative analysis of Data Center Policymaking in Singapore, Netherlands, Ireland, Germany, USA, and the UK. *Environmental Science & Policy*, 157, 103769. <https://doi.org/10.1016/j.envsci.2024.103769>.

[50] Sundberg, N. (2023). Tackling AI's Climate Change Problem. *MIT Sloan Management Review*, 65(2), 38-41. <https://sloanreview.mit.edu/article/tackling-ais-climate-change-problem/>.

[51] Szilas, N. (2024). Digital Sobriety: from Tips to Values. *LIMITS workshop*. <https://computingwithinlimits.org/2024/papers/limits2024-szilar-digital-sobriety.pdf>.

[52] Tomlinson, B., Black, R. W., Patterson, D. J., & Torrance, A. W. (2024). The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports (Sci Rep)*, 14(1), 3732. <https://doi.org/10.17605/OSF.IO/YHTMQ>.

[53] Vaija, S. et Philipot, É. (2020). L'importance des métaux rares pour le secteur des technologies de l'information et de la communication, le cas d'Orange. *Annales des Mines-Responsabilité et environnement*, 2, 24-28. <https://doi.org/10.3917/re1.099.0024>.

[54] Vanderbauwhede, W. (2023). Frugal Computing--On the need for low-carbon and sustainable computing and the path towards zero-carbon computing. *arXiv preprint arXiv:2303.06642*. <https://doi.org/10.48550/arXiv.2303.06642>.

[55] Viseur, R. (2025a). Sobriété numérique et traitement des données massives: vers des stratégies d'innovation plus durables. *Terminal. Technologie de l'information, culture & société*, (140), 61-83. <https://doi.org/10.4000/13vvh>.

[56] Viseur, R. (2025b). Des fonctionnalités à coût maîtrisé ? Le modèle A-U appliqué à l'IA générative. *APIA (Applications Pratiques de l'Intelligence Artificielle)*, Dijon, France. <https://hdl.handle.net/20.500.12907/52466>.

[57] Viseur, R. (2026). Éthique de la gestion du consentement au traitement des données à caractère personnel: Comment les dark patterns permettent-ils d'orienter la prise de décision des internautes ? *Revue ouverte d'ingénierie des systèmes d'information*, Vol. 5. <http://doi.org/10.21494/ISTE.OP.2025.1301>.

[58] Yang, X., Li, Y., & Kang, L. (2020). Reconciling "doing good" and "doing well" in organizations' green IT initiatives: A multi-case analysis. *International Journal of Information Management*, 51, 102052. <https://doi.org/10.1016/j.ijinfomgt.2019.102052>.