IA, données, calculs : quelles infrastructures dans un monde décarboné ? | Synthèse des ateliers collaboratifs du 6 mars 2025

Contexte : la présentation du rapport intermédiaire

Le 6 mars 2025, l'équipe Numérique du Shift Project a eu le plaisir de présenter en ligne son rapport intermédiaire. Ce webinaire a été l'occasion de mettre en discussion les premiers résultats des travaux lancés fin 2024 sur **les impacts et trajectoires énergie-climat des infrastructures et capacités de calcul et de l'intelligence artificielle.**

L'ensemble des documents (rapport intermédiaire, support de présentation, replay) sont disponibles sur cette page : https://theshiftproject.org/article/rapport-intermediaire-ia/

Introduction : les ateliers collaboratifs

La présentation du rapport a été suivie d'ateliers collaboratifs en ligne de 15h15 à 17h30 réservés aux professionnels du numérique (réservation obligatoire), afin d'approfondir certaines thématiques de l'étude. Ce document est la synthèse de ce qui a été échangé lors de ces 4 ateliers :

- Atelier 1 : Les dynamiques mondiales d'évolution des capacités de stockage et de calcul : Quelles implications sur les centres de données, le système numérique et leurs impacts énergétiques et climatiques ?
- Atelier 2 : Les centres de données en France : Quelles modélisations énergétiques et climatiques pour quels enjeux ? Quels inventaires pour quels suivis ?
- Atelier 3 : Cas d'usages de l'IA : Quelles conditions pour engendrer des dynamiques énergétiques et climatiques bénéfiques ?
- Atelier 4 : Qualifier les impacts directs et indirects de l'IA sur le système numérique : un exercice indispensable pour détourer les impacts carbone d'une adoption de l'IA

Avis de lecture : Les éléments rapportés sont des quasi verbatim, non nécessairement vérifiés.

Nous remercions l'ensemble des participantes et participants, ainsi que les personnes qui nous ont apporté leurs retours en asynchrones.

Table des matières

Atelier n°1 : Les dynamiques mondiales d'évolution des capacités de stockage et de calcul : Quelles implications sur les centres de données, le système numérique et leurs impacts énergétiques et climatiques ?

L'objectif de cet atelier est d'interroger la partie p17-37 du rapport intermédiaire.

Ont participé 18 professionnelles et professionnels du secteur sur 26 confirmations sur 117 demandes. Parmi les profils présents : centre national de recherche, du numérique et de l'énergie, agence nationale, direction générale, entreprise de l'énergie, entreprise du numérique, représentant de filière numérique, acteur de la normalisation, cabinet de responsable politique, membre de conseil, enseignant d'école, d'université, think tank climat, retraité et enfin shifter.

L'échange a pris lieu autour de 4 grandes thématiques :

- Empreinte des data centers, lors des phases de production et de fonctionnement.
- Autres impacts environnementaux.
- Cas des pays en développement.
- Estimation de l'évolution de la consommation électrique des data centers.

Impacts sur l'eau :

Ce n'est a priori pas vraiment un sujet, puisque peu de volume d'eau est prélevé par les data centers, la problématique reste très localisée sur les zones en stress hydrique.

Attention, le problème est plutôt que nous n'avons pas de mesure donc les consommations sont difficiles à estimer, et les besoins en eau augmentent.

Une approche transversale est nécessaire car l'impact va dépendre du mix énergétique présent sur les territoires et de l'augmentation simultanée des autres besoins en eau liée au réchauffement climatique, avec des conflits d'usage à prévoir.

Ex de l'inde très illustratif, mais même en France, par exemple à Marseille.

En France, la majorité des data centers est pour l'instant à refroidissement à air, l'augmentation du volume d'eau pour le refroidissement est moins un sujet qu'ailleurs, y compris avec la montée en densité des DC.

Empreinte carbone des data centers :

Discussions sur le poids de l'empreinte carbone de fabrication du data center Vs d'utilisation / d'exploitation.

On sait estimer l'empreinte de la construction et de la consommation d'énergie. Les bâtiments et installations techniques représentent de l'ordre de 20% de l'empreinte. On peut prendre une analyse de cycle de vie de 30 ans pour les bâtiments.

On a plus de mal à estimer l'empreinte carbone des équipements informatiques, qui est supérieure à l'empreinte du bâtiment. Ceci notamment car les équipements IT sont renouvelés plus souvent.

Sujet des nouvelles typologies de data centers IA Vs data centers Telecom classique ; avec des typologies de bâtiments différentes (densité de puissance, distribution énergie, climatisation, …). Ces nouveaux data centers IA s'additionnent à l'existant (pas de remplacement).

Interaction data centers – réseau de transport d'électricité :

Comment évolue l'alimentation électrique dans les data centers ?

Avec l'augmentation des capacités de production « on site » développées par ex par les hyperscalers, les data centers peuvent-ils jouer un rôle de résilience dans les systèmes énergétiques du futur ?

Google a signé un accord avec kairos power qui développe un SMR.

Est-ce que les DC vont devenir des centrales d'énergie qui auto-consomment une partie de l'énergie qu'ils produisent ?

On constate beaucoup de dynamiques sur l'optimisation énergétique des data centers avec des notions de réutilisation de l'énergie produite, de gestion de l'intermittence et d'effacement. Par ex RTE met en place un système d'heures pleines / heures creuses pour les data centers (source req).

Foncier :

Différences de dynamique entre Europe et USA.

En Europe, on joue avec les contraintes plus fortes (Plan Local d'Urbanisme), pas de problématique de gestion des grandes surfaces. Différent des USA ou il y a de très grandes surfaces plates. On ne voit pas arriver les grandes surfaces types USA qui consomment beaucoup de fonciers.

En France on optimise la puissance informatique / les ratios d'utilisation pour utiliser moins de foncier.

Charge :

On estime le ratio entre la puissance réellement utilisée et la puissance maximum à 70% pour le dimensionnement.

Ressources métalliques pour construction des data center et terminaux :

Pourquoi ne pas éclairer ce point pour le rapport final.

Discussion parallèle dans le chat :

Nouveaux styles de data centers :

Sous-marins : Natick par Microsoft, Naval Group en 2015, HiCloud en Chine plus récemment.

Flottant : data center flottant à Nantes.

Taux d'utilisation des data centers :

A pleine capacité actuellement pour les data centers des GAFAM. En revanche, les serveurs locaux sont utilisés de manière sous optimale. L'utilisation pour l'entraînement permet aussi de mieux anticiper l'allocation des calculs et d'optimiser l'utilisation contrairement à l'inférence. Ces deux justifications (data center central géré par un GAFAM et utilisation pour l'entraînement) pourraient se recouper.

Estimation de l'évolution de la consommation électrique des data centers :

Quelles sont les hypothèses pour traduire une trajectoire de demande de puissance en consommation. Echanges sur les délais réglementaires Vs transformation du projet en opérationnel, délais de montée en charge du data center.

Difficile d'estimer les délais de montée en charge car les constructeurs / opérateurs de data center ne sont pas ceux qui vendent les services. Ce qu'on voit ce sont des montées en charge à 80%, en quelques mois, parfois en quelques semaines.

Délais de livraison bout en bout d'un DC :

- 18/24 mois pour les phases réglementaires / permis de construire
- 18/24 mois pour construction du site / bâtiment
- 18/24 mois pour les installations dans les locaux
- Donc entre 4 ans et 6 ans, ceci si RTE sait livrer l'énergie, avec des tensions constatées par ex aujourd'hui en région parisienne
- + rapide en Angleterre et Italie (pas à Rome car forte contrainte sur l'obtention des permis de construire)
- En Allemagne et en Suisse c'est + long

Les moyennes constatées :

- Temps de construction d'un bâtiment = 2 ans

- Déploiement de solutions dans une salle déjà équipée = 5 mois

+ rapide aux USA avec des délais de réalisation de 12 à 18 mois au global

En Amérique du Nord, avec la standardisation des solutions, on accélère les délais de traitement pour atteindre entre 9 à 12 mois

En France on est + contraints sur l'urbanisme + étude environnementale avec la loi ZAN (zéro artificialisation sol). Précision que le PJL simplification ne vient pas supprimer cette étude mais modifie son processus.

Procédure spéciale en IDF (1/3 des DC en France).

Méthodologie pour le dimensionnement des data center Edge, quelles sources d'information ?

France data center a fait une étude d'impact de la filière DC, le phénomène Edge reste compliqué à caractériser.

Voir le magazine global security qui est la meilleure carte des data center en France

Voir immobilier pappers qui liste toutes les surfaces de parcelles, à retravailler avec les données de vente de matériel

Atelier n°2 : Les centres de données en France : Quelles modélisations énergétiques et climatiques pour quels enjeux ? Quels inventaires pour quels suivis ?

L'objectif de cet atelier est de renforcer la partie p38-63 du rapport, principalement p53-57.

Le format de cet atelier était en 3 grands temps, introduits par un support de présentation (intégré ci-dessous).

- Contexte et introduction : typologie de centres de données (15 min)
- Modélisation : leviers de dimensionnement, de contraintes, d'impact (35 min)
- Scénarisation : modélisation de l'offre et de la demande et scénarios possibles (30 min)

Ont participé 15 professionnelles et professionnels du secteur sur 26 confirmations sur 68 demandes. Parmi les profils présents : entreprise de services du numérique, dont en impacts environnementaux du numérique, acteur de l'électricité, du réseau de transport, du gaz, opérateur télécom, direction générale, direction régionale, agence nationale, centre national de l'énergie, du numérique, organisme de formation, eco-organisme, académique et enfin étudiant.

Avis de lecture : Les éléments ci-dessous sont des quasi verbatim des participantes et participants et ne sont pas vérifiés.

Introduction :

Atelier n°2 | Les centres de données en France : Quelles modélisations énergétiques et climatiques pour quels enjeux ? Quels inventaires pour quels suivis ?

Suite à un tour de table (20 min)

Cet atelier vise à échanger sur le contenu de la partie *« Une première réflexion sur les modélisations possibles et leurs objectifs »* du rapport intermédiaire, présenté précédemment. Nous espérons vos retours et contributions pour aller plus loin pour le rapport final (critiques constructives et données).

Règles de prise de parole : main levée, les animateurs distribuent, faire des interventions les plus concises possibles pour laisser la parole aux autres. Vous êtes 27, il y a environ 20 min par partie (=> < 1 min / intervention).

Le format de cet atelier est semi-directif en 3 grands temps autour des thématiques suivantes :

- **Contexte et introduction :** typologie de centres de données (15 min)
- **Modélisation :** leviers de dimensionnement, de contraintes, d'impact (35 min)
- **Scénarisation :** modélisation de l'offre et de la demande et scénarios possibles (30 min)

The Shift Project – IA, données, calculs : quelles infrastructures dans un monde décarboné ? Ateliers collaboratifs du 6 mars 2025 - Atelier n°2 1

Atelier n°2 - Partie 1 :

Partie 1/3 | Contexte : typologie de centres de données (15 min)

Constat : La catégorie « centre de données » est large : elle recoupe différentes tailles/concentrations (en m^2, GW d'IT), différents modèles d'affaires (on-prem, fournisseurs de services, fournisseurs d'infrastructures de colocation), différentes localisations, pour différents usages (polyvalents, spécialisés (CDN, IA, etc)).

Nous considérons les catégories de centres de données suivants :

- edge
- entreprise
- fournisseurs de services de télécommunications
- colocation ou co-hébergement
- hyperscale

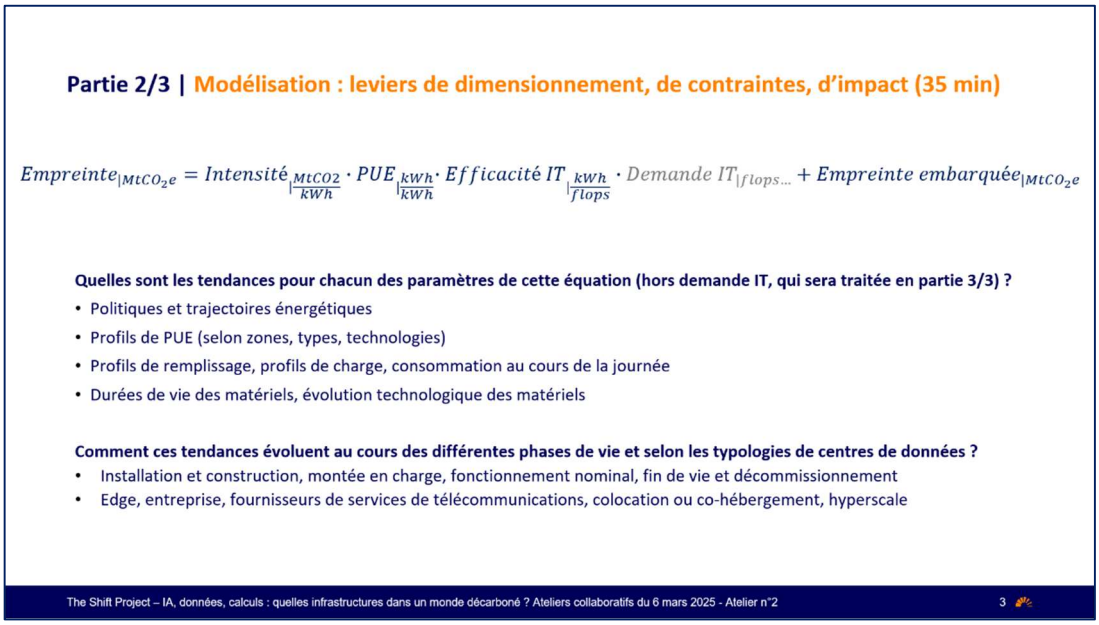
Quelles sont leurs spécificités : matériels, systèmes de refroidissement, approvisionnement électrique ?

Quelles sont les sources de données disponibles ?

The Shift Project – IA, données, calculs : quelles infrastructures dans un monde décarboné ? Ateliers collaboratifs du 6 mars 2025 - Atelier n°2 2

- Tous les data centers (DC = Data Centers) sont importants pour compter la consommation électrique et les émissions de gaz à effets de serre et pas seulement les data centers dédiés à l'IA.
- Proposition d'avoir une nomenclature selon taille et localité.
- Idée de partir de la taxonomie DEE (par taille de DC) + séparation retail/whole sale.
    - Hyperscale : 0 en France à date mais une vingtaine en projets.
    - Edge : ne savent pas à quoi raccrocher, plutôt une tendance de marché => des DC retail ou des telcos.
- Penser à différencier selon les systèmes de refroidissement. Pour les HPC, ce n'est pas si clair si les DC IA feront du DLC (Direct Liquid Cooling) du fait de la densification. C'est difficile d'extrapoler des DC actuels aux DC IA qui feront du DLC.
- HPC historiques assez proches des futurs DC IA ? C'est déjà du DLC. A partir de 40 kW (entre 40 et 70 kW) on passe à du DLC. Il est possible d'avoir du door cooling : le DLC sur la porte de la baie et ventilos sur les serveurs. Sorte d'intermédiaire pour les petites IA.
- Le DLC va faire baisser le PUE. Etude de Berkeley qui donnerait des PUE et des WUE pour les US. Mais c'est pour les US, donc géographie spécifique.
- Besoin d'eau pour évacuer la chaleur par évaporation. Donc risque avec les fortes chaleurs : ça va faire augmenter la consommation d'eau.
- Sur le secours, sujet sur la taille du DC vs. la puissance des générateurs (augmenter le nombre d'unité prend de la place ou il faut stocker beaucoup de fuel). Piles à combustible (méthane/biométhane => raccordement au réseau de gaz pour éviter les problèmes de stockage). BESS : pb sur le volume des batteries et risques de combustion vs. autonomie recherchée (2 jours) mais pas trop vu.
- Les moyens de secours actuels pour les data center cloud permettent une autonomie de 2/3 jours. Des rapports soulignent que la continuité d'alimentation est moins critique pour les data centers dédiés à l'entraînement de modèles d'IA et donc que le rapport des puissances peut être différent (à vérifier avec des données réelles des Etats-Unis par exemple, il est possible que cet argument soit relayé par des hyperscaler pour réduire l'estimation de leurs consommations).

Atelier n°2 - Partie 2 :



- Dans la SNBC, approche hybride qui part de la demande et qui reboucle avec la surface des DC.
- La géographie est importante à prendre en compte ici, ne serait-ce pour prendre en compte l'intensité carbone de l'électricité.
- Cette équation pourrait amener à une estimation élevée car des effets sont masqués (l'intensité carbone de l'électricité n'est pas la même le jour et la nuit, etc.). Elle ne prend pas en compte des effets d'optimisation. Calculs sur la base de séries temporelles pour 1 DC.
- Question de séparer selon Cloud souverain ou pas
- Question d'avoir une approche du type ACV « une heure d'IA c'est tant », ce qui n'est certes pas l'objet de cette modélisation qui cherche à restituer les émissions totales
- La logique et le rythme d'implantation des DC sont décorrélés de la demande, donc pas évident qu'il faille utiliser la demande dans l'équation. Par ailleurs, manque d'information sur les workloads qui sont vieilles et qui viennent d'on ne sait pas où => approche sur les m² qui sont +/- publics.
- PUE : beaucoup d'incohérences et de zones d'ombres. A l'échelle locale, quand on n'a pas le PUE, utilisation d'un PUE local majorant, mais cette approche semble difficile à appliquer à l'échelle d'un pays voire au-delà.
- (NB : dans les ACV de services numériques, il y aussi la virtualisation à prendre en compte, et si celle-ci n'est pas efficace, il n'y a pas l'effet d'échelle escompté. => l'écoconception de services numériques ne s'arrête pas au bâtiment.)
- Chaudière, chaleur fatale => Qarnot. Donc plutôt exploitation pour le edge. Impact sur l'ERF et pas le PUE.
- Surtout fabrication des GPU ? 2 projets en cours mais pas de résultats encore.
- Etude Google récente sur les TPU
- Question du potentiel de récupération de chaleur fatale. Ça reste compliqué. Comparaison d'un calcul fait en edge vs. en hyperscale.
- Où trouver les profils de charge des DC ?
- Beaucoup de difficultés pour trouver ça. 2 usages typiques : HPC avec consommation plate grâce aux schedulers, et commerce (car c'est un site d'e-commerce qui a publié). Les acteurs ne communiquent pas beaucoup sur ce genre de choses. Amazon fait payer en fonction des temps de calcul (avec des seuils) => les coûts peuvent donner une idée de la charge, mais ces coûts sont fonction du type de service acheté (disponibilité).
- Quels types de DC fournissent l'IA aujourd'hui ? Est-ce qu'on en a en France qui accueille de l'IA ?
- Fournisseurs de services d'IA (ex. Mistral AI) => quels fournisseurs ? (ex. Corewidth) => localisation. Uptime Institute et ses sondages sur la densité par baie. Attention : pas de lien entre lieu d'entraînement et lieu d'inférence (l'intensité carbone de l'électricité n'est pas la même).
- A l'ADEME : deux travaux en cours : un prospectif (scénarios ADEME) et un sur les enjeux territoriaux d'implantation.

Atelier n°2 - Partie 3 :

- Données sur les ventes et stocks de matériel de mauvaise qualité.
- Mais l'approche m² est privilégiée chez certains.
- Attention aux graphes d'efficience énergétique publiés par NVIDIA en W/FLOP. D'année en année, ce ne sont pas les mêmes FLOP. Ex. l'inférence nécessite moins de précision et donc communiquer sur l'inférence ou faire plus d'inférence que d'apprentissage ça améliore l'efficacité.
- Les acteurs de colocation donnent leurs m². API : immobilier.papers.fr => m² des DC + un ratio pour passer aux m² IT.
- Base de données « institut Paris données région » pour avoir des infos DC en IDF.
- Base de données en open source.
- Scenarios selon la distinction des offres en France et à l'étranger.

Conclusion :

- Consommation d'eau. On pourra peut-être faire baisser les indicateurs énergétiques mais en augmentant la consommation d'eau.
- PUE et WUE peuvent être des vases communicants. RCP Cloud v2 qui sortira avec des facteurs d'impacts.
- openinframap => localisation des DC : https://openinframap.org/#5.88/47.442/2.958, cocher Telecommunications et Electicité
- On a une recommandation d'allonger la DDV des terminaux mais quid des DC ? Avantage : on pourrait s'adresser à des entreprises et pas seulement au grand public.
- Proposition de reprendre l'équation de Kaya et de travailler sur l'un des facteurs
- Moyens de secours des DC à prendre en cours. Il y aura de + en + de coupures de courant ou de l'effacement, donc de plus en plus recours aux moyens de secours qui restent très carbonés.
- Efficacité des groupes électrogènes : 35% à 40% (gaz ou fioul). 60% piles à combustible sans émissions de CO2 (sauf démarrage).
- Consommation en eau. Les enjeux et les conflits d'usage seront surtout à l'échelle des territoires. Questionnement du caractère spéculatif de l'IA en ce moment donc attention à comment que cela pourrait déformer les projections.
- Attention aux usages et les biais que cela peut introduire dans la modélisation
- https://www.extendo-datacenter.fr/actualites/services/la-cartographie-2021-des-215-datacenters-en-france-est-sortie/

Atelier n°3 : Cas d'usages de l'IA : Quelles conditions pour engendrer des dynamiques énergétiques et climatiques bénéfiques ?

L'objectif de cet atelier est de renforcer la partie p64-73, principalement la description des cas d'usage.

Ont participé 12 professionnelles et professionnels du secteur sur 26 confirmations sur 263 demandes. Parmi les profils présents : centre national de l'énergie, du numérique, entreprise (start-up), entreprise de services du numérique, agence nationale, académique, professeur (philosophie), think tank climat et enfin opérateur télécom.

Cet atelier a doublement fait varier :

- Les cas d'usage étudiés (assistant intelligent, robot intelligent, assistant de compte-rendu intelligent)
- Les angles d'études (fonctionnalités et briques technologiques, infrastructures numériques (réseaux, terminaux, centres de données), pertinence/intérêt des cas d'usage vis-à-vis des indicateurs énergétiques et climatiques)

Ci-dessous sont présentés successivement les supports utilisés pendant l'atelier, mis à jour ensuite des réflexions majeures pendant l'atelier.

Atelier 3 - Déroulé

Début de la session - 15h15

15h15 - 15h20 - Introduction & explication du déroulé

15h20 - 15h40 - Tour de table : Nom, Prénom, Poste, Organisation, Intérêts sur le sujet

15h40 - 16h05 - Cas d'usage 1 - Fonctionnalités et briques technologiques

16h05 - 16h30 - Cas d'usage 2 - Infrastructures, Réseaux & Terminaux

16h30 - 16h55 - Cas d'usage 3 - Analyse sous le prisme {Energie - Climat}

16h55 - 17h15 - Tour de table ouvert

17h15 - Fin de la session - Retour en plénière

Atelier n°3 – Cas d'usage 1 en support à une réflexion sur les fonctionnalités et les briques technologiques :

15h40 - 16h05 - Cas d'usage 1

Nous disposons d'un assistant intelligent qui sait tout faire

Quelles sont les fonctionnalités disponibles de cet assistant ?

Exemples de fonctionnalités : capacité décisionnelle, rapidité, polyvalence, sécurité, …

Quelles sont les briques technologiques nécessaires à son existence ?

Exemples de briques technologiques : IAG, RAG, Algorithmie, …

Synthèse express :

Cas d'usage 1

Nous disposons d'un assistant intelligent qui sait tout faire

Prisme: fonctionnalité / briques technologiques

En bref

- Nous avons exploré les fonctionnalités de cet assistant, avec des échanges approfondis notamment sur 3 fonctionnalités:
  → **Souveraineté / confidentialité / transparence**
  → (sûrement biaisé par le contexte géopolitique actuel)
- 1 bloqueur remonté : la difficulté à identifier / monitorer les briques technologiques
  → les évolutions à vitesse grand V rendent obsolètes de nombreuses modélisations / réflexions

Atelier n°3 – Cas d'usage 2 en support à une analyse sous le prisme des infrastructures (réseaux, data centers) et des équipements numériques :

16h05 - 16h30 - Cas d'usage 2

Nous disposons d'un robot intelligent autonome avec une capacité de décision
(Robot agricole dans le Nord pour la surveillance de la pousse des betteraves à sucre)

Quelles sont les infrastructures, les réseaux et les terminaux mis en jeu ?

Exemples d'infrastructures : centre de donnée hyperscale, edge, IA, …

Exemples de réseaux : 5G, antennes, …

Exemples de terminaux : batteries, puces embarquées, capteurs, …

Synthèse express :

Cas d'usage 2

Nous disposons d'un robot intelligent autonome avec une capacité de décision
(Robot agricole dans le Nord pour la surveillance de la pousse des betteraves à sucre)

Prisme: infrastructures

En bref, au delà de la nécessité de permettre l'essor de ces robots, nous avons exploré les impacts infra / réseaux / terminaux avec 3 axes principaux:

- La notion d'hypervision / supervision de nuages de drônes
- Les impacts sur les infrastructures hors "IT" : fin de vie, bornes de recharge
- Le besoin d'équilibre de la charge de calcul (edge / embarqué / federated)

Atelier n°3 – Cas d'usage 3 en support à une analyse sous le prisme de la pertinence/de l'intérêt énergétique et climatique :

Cas d'usage 3

Nous disposons d'un assistant de compte-rendu intelligent

Prisme: leviers & maximisation des bénéfices (énergie / climat)

- En bref:
  - *Utilité* **vs** *éthique*
  - *Gain / Productivité* **vs** *Coûts environnementaux / sociaux*
  - *Humain* **vs** *Machine*
  - *Difficulté à borner l'échange sur la dimension {énergie - climat} très restrictive*

Synthèse express :

16h30 - 16h55 - Cas d'usage 3

Nous disposons d'un assistant de compte-rendu intelligent

Pour ce cas, **sous le prisme {Energie - Climat}**, quels sont les leviers qui nous permettent de

- Comprendre et évaluer sa pertinence ?
- Maximiser ses bénéfices ?

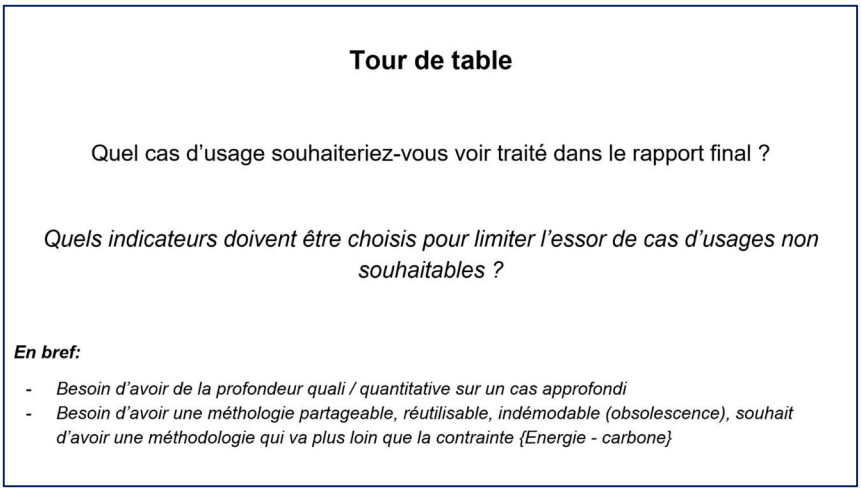
Ensemble des prises de parole :

- En question énergie/climat : vaut-il mieux le faire faire par une personne ? Entre le temps que je passe à le faire, est-ce que mon assistant est capable de me dire si le coût est plus intéressant ? Est-ce lié au temps de travail ?
- Pour maximiser le bénéfice, ça doit être hybride, mais donc il devrait être capable de me dire ce qu'il doit utiliser, avec une approche ACV comparative et ensuite l'adapter au juste besoin.

Pour évaluer sa pertinence, il faudrait que le demandeur soit compétent pour être capable de discriminer à l'intérieur du compte rendu quelle en est la valeur. En plus, si on parle de réunion, il serait préférable que le demandeur ait assisté à la réunion. Il y a des requêtes d'expertise par rapport à des expériences pratiques.
Il semble difficile d'évaluer la pertinence uniquement sous le prisme énergie/climat. Se dire : "voilà la quantité d'énergie dont on dispose, ensuite quels sont nos besoins ?"

- En fait, il faudrait faire un premier prompt pour identifier ce que l'on veut, et qu'ensuite, l'IA évalue le temps et l'énergie que cela va lui prendre, avec une première balance bénéfice/impact. Puis, avant d'effectuer le travail, valider cette évaluation soi-même. Cela paraît très compliqué à quantifier.
- La vitesse à laquelle on évalue est également un facteur clé.
- Il y a un backlash en ce moment, on mise plutôt sur l'inverse. Drill, baby, drill, il faut accélérer, car il y a une lutte entre les grandes puissances.
- En fait, ça avance trop vite pour qu'on puisse quantifier.
- Avoir une méthodologie pourrait être une solution. La question est : qu'est-ce qu'on veut faire avec cette IA ?
- Quelles tâches ?
- Qu'est-ce que ça apporte ?
- Quel impact économique et climatique ?
- Faut-il réfléchir par secteur, en fonction des besoins et des contraintes ?
- Pour les data centers, la question de l'utilité réelle de l'IA et de l'éthique pour l'utiliser se pose.
- D'un point de vue purement pragmatique, ce qui est demandé, c'est combien cela va coûter en euros, en impact écologique et social. Sinon, l'analyse se focalise uniquement sur l'aspect financier, ce qui est réducteur.
- Si une méthodologie validée peut être intégrée dans les modèles et transmise aux utilisateurs, cela faciliterait l'évaluation de l'impact réel.
- L'IA est un accélérateur. Mais qu'est-ce qu'on veut accélérer ?
- Qu'est-ce qu'on fait des gains de productivité permis par l'IA ? C'est sur cette question qu'il faut collectivement poser des conditions si l'on veut orienter l'IA vers des bénéfices climatiques, écologiques et sociaux.
- Et clairement, le prérequis dont il est question ici, c'est la bonne méthode de calcul
- Multicritère via :
    - Taches
    - Environnement
    - Social
    - Économique
    - Prise en compte des effets rebonds
    - Contraintes & limites
    - Stockholm résilience center = IA & éthique pour utiliser l'IA -> lien philosophique
- Clients demandent de l'éthique, mais le business model, c'est d'aller plus vite
    - Combien ça coûte (finance / écolo / social)
    - Quel ROI (finance / écolo / social)
    - Il faudrait une méthodo comme celle ci validée par l'écosystème à développer en entreprise pour les utilisateurs
    - L'IA est un accélérateur, qu'est qu'on veut accélérer

Tour de table conclusif :



Tour de table

Quel cas d'usage souhaiteriez-vous voir traité dans le rapport final ?

Quels indicateurs doivent être choisis pour limiter l'essor de cas d'usages non souhaitables ?

En bref:

- *Besoin d'avoir de la profondeur quali / quantitative sur un cas approfondi*
- *Besoin d'avoir une méthologie partageable, réutilisable, indémodable (obsolescence), souhait d'avoir une méthodologie qui va plus loin que la contrainte {Energie - carbone}*

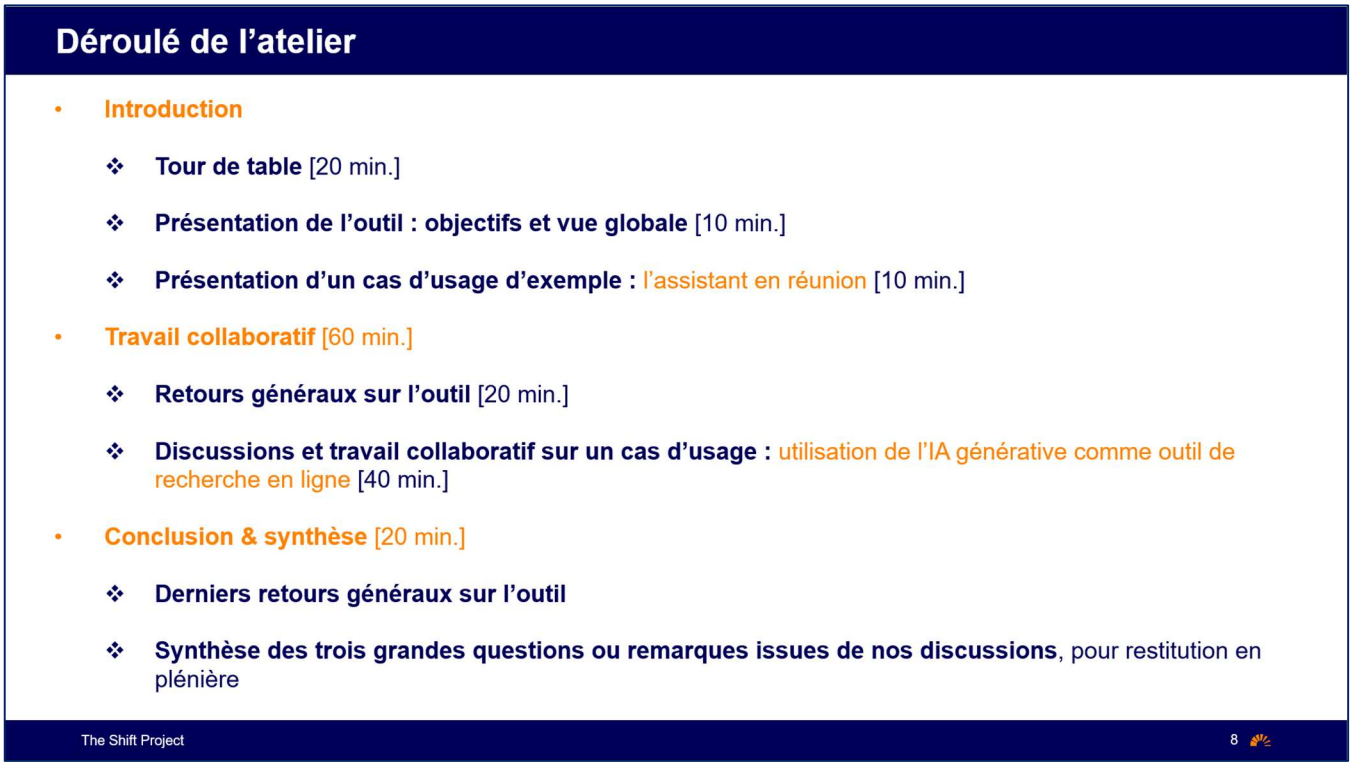
- Il est essentiel de prendre le temps de se poser les bonnes questions.
- Une grande partie des usages concernent l'IA générative.
- Il pourrait être intéressant d'aller sur des modèles plus petits. Le secteur de l'énergie pourrait être optimisé, avec des gains significatifs.
- L'optimisation avec PAC permet par exemple un gain de 40% sur les logements collectifs.
- Les cas dans le RI sont intéressants, mais ils devraient être plus contextualisés. Quitte à en avoir moins, mais mieux les contextualiser.
- Un agent IA "vert" pourrait prendre en compte la dimension écologique dans ses propositions d'action. Par exemple, une IA qui planifie les voyages en favorisant les mobilités douces et le covoiturage. L'exercice de pensée est intéressant, car il s'agit d'un cas d'usage avec un bénéfice concret, mais qui a un coût (entraînement, fine-tuning, inférence…).
- Les cas d'usage en optimisation logistique, visant à remplir au maximum les véhicules et les conteneurs, sont également pertinents.
- L'IA non générative peut aussi être utilisée pour des solutions pro-environnementales et sociales, comme l'imagerie médicale ou la mise en place de plans biodiversité via l'imagerie satellite.
- Prioriser les projets IA repose sur deux critères principaux :
    - Est-ce que les briques technologiques nécessaires sont disponibles ?
    - Est-ce que cela génère des bénéfices financiers ?
- Pour les applications de covoiturage, certaines études montrent que les bénéfices supposés ne sont pas toujours confirmés dans la réalité.
- Sur les indicateurs, il faudrait permettre aux utilisateurs de comprendre le coût réel de l'inférence d'un algorithme en énergie ou en eau. L'inférence pourrait servir d'unité de mesure, ou alors le token dans le cas des LLM.
- Une application IA pourrait être développée pour limiter l'usage de certains services, comme la voiture.
- Une autre idée serait de voir en temps réel le coût d'une IA lorsqu'on lui pose une question.
- Tout comme la pollution a été exportée à l'étranger, l'impact de la consommation numérique a été invisibilisé.
- L'agriculture semble être un bon sujet d'exploration, car il est très complet en termes de cas d'usage, avec des implications sur les facteurs de production et les terminaux.
  Le transport est également un sujet clé en matière de décarbonation, mais il subsiste encore beaucoup d'incertitudes sur les gains réels.
- L'optimisation de l'usage des voitures par le covoiturage peut être efficace, mais il faudrait évaluer précisément comment cela fonctionne en pratique.
- Pour aborder ces sujets d'un point de vue scientifique, la meilleure approche est de travailler directement avec les entreprises qui reçoivent ces prompts : Comment leurs outils sont-ils utilisés ? Quelle est la part des prompts inutiles ?

Atelier n°4 : Qualifier les impacts directs et indirects de l'IA sur le système numérique : un exercice indispensable pour détourer les impacts carbone d'une adoption de l'IA

L'objectif de cet atelier est d'aller plus loin que la partie p64-73, en réfléchissant à une façon de synthétiser et de vulgariser les impacts directs et indirects de l'IA.

Ont participé 20 professionnelles et professionnels du secteur sur 27 confirmations sur 227 demandes. Parmi les profils présents : éditeur de logiciel, expert impacts environnementaux du numérique, assureur, banquier, psychologue du travail, ingénieur agronome, centre national de l'énergie, du numérique, laboratoire d'innovation, agence nationale, ingénieur université, école d'ingénieur et association de recherche sur les algorithmes.

Le déroulé de l'atelier était le suivant :



L'introduction avait pour objectif de présenter « la boussole de l'IA », objectifs et intentions :

Présentation : les grands objectifs

La « Boussole de l'IA », prochaine grande étape du travail :

Produire un support et une méthode
Permettre la discussion éclairée entre les sphères décisionnelles, techniques de conception et techniques de l'impact environnemental

Analyse qualitative des cas d'usage
Quels effets des choix de fonctionnalités sur le système numérique et ses impacts ?

Analyse quantitative d'un cas d'usage
Quantifier les impacts carbone-énergie d'un cas d'usage et de son déploiement

Prendre en compte les effets indirects et systémiques
Impacts engendrés sur le système numérique, le système d'usage, effets rebonds et autres impacts

Présentation : les grands objectifs

La « Boussole de l'IA », permet de poser trois grandes questions :

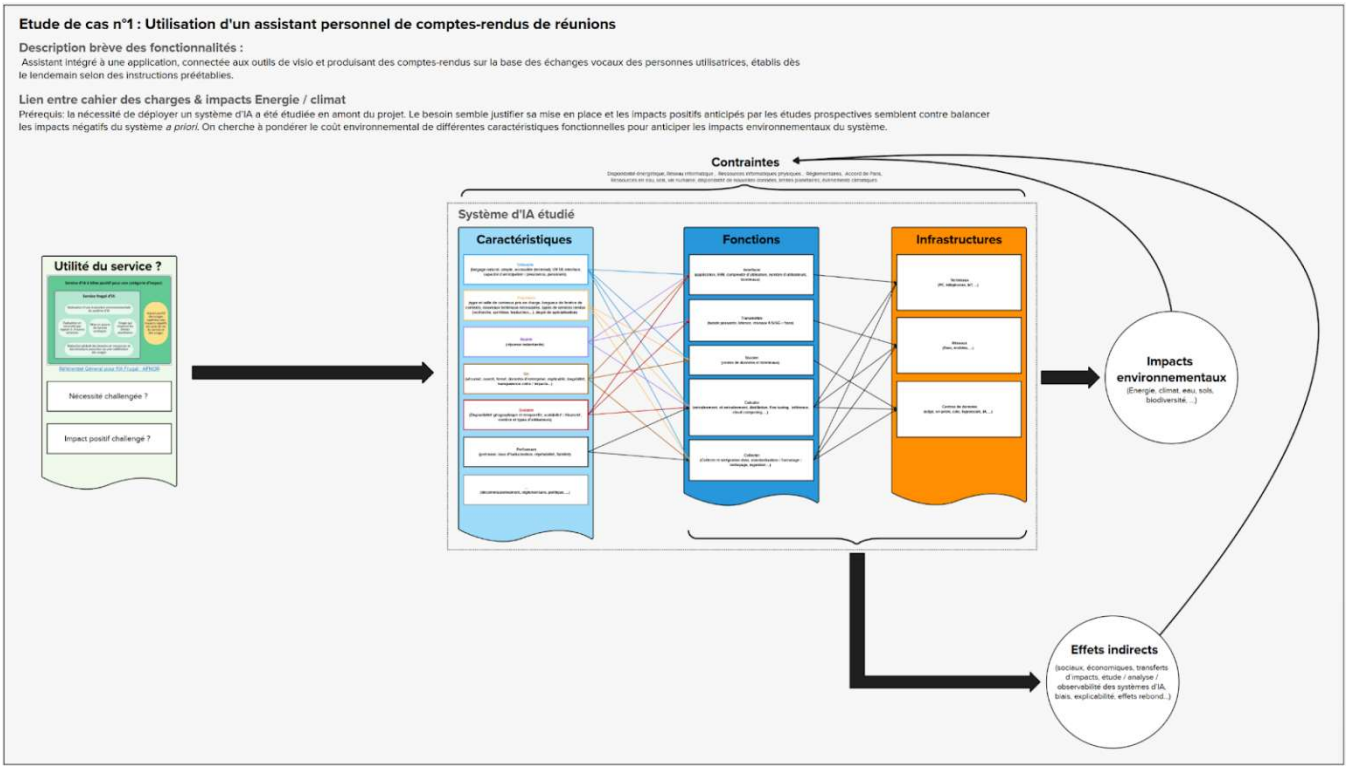
Objectiver le besoin : quelles sont les caractéristiques du besoin auquel on souhaite répondre avec ce système d'IA ?

Un système d'IA est-il nécessaire ou existe-t-il d'autres solutions ?

Quelle traduction des choix du cahier des charges fonctionnel (caractéristiques, solutions techniques) en impacts infrastructures et environnementaux ?

A l'aune de deux cas d'usage différents (assistant personnel de comptes-rendus de réunion, moteur de recherche augmenté), une proposition méthodologique a été travaillée puis challengée. Dans laquelle, il s'agit de partir d'un cas d'usage, puis de successivement aborder :

- Une brève description des fonctionnalités
- Les questions à se poser pour jauger de l'utilité du service
- Quel lien entre cahier des charges et impacts énergétiques et climatiques via une réflexion sur les caractéristiques des services d'IA, les fonctions et infrastructures sollicitées.



Premiers retours avant le travail sur les cas d'usage :

- Parle-t-on d'IA ou d'IA générative ?
- Est-ce que l'objectif est de réaliser un arbre des conséquences détaillées ?
- Est-ce que l'outil doit traiter les impacts indirects ?
- Choix du vocabulaire « caractéristiques » vs « fonctionnalités » vs « fonctions » vs « solutions technologiques »

Le premier cas présenté, le second étudié en séance sont disponibles ici :

- https://app.mural.co/t/alex8698/m/alex8698/1740762738110/65fb6c19a6f92854936ae7c9d9c630d343aefd05

Retour lors du travail sur le cas « moteur de recherche augmenté » :

- L'outil doit permettre de « dé-polluer » ce service
- Et d'améliorer son modèle économique (abonnement ou pub), puisque la pub, cela signifie des effets indirects qui ne sont pas mis en valeur par l'outil
- Le modèle économique, selon les cas d'usage peut avoir une importance capitale dans la détermination des impacts (directs et indirects) et semble donc un axe majeur dans l'outil
- Cela pose question sur le fait d'étudier un cas d'usage de "moteur de recherche" alors que la GenAI permet d'aller plusieurs étapes plus loin. Est-ce pertinent puisqu'on change complètement l'usage ?
- Pouvoir facilement représenter dans l'outil l'importance des impacts sur les Fonctions et Infrastructures serait un vrai plus
- Pouvoir comparer des options aussi (plaide pour un outil modulaire)

Conclusions :

- Quid d'éclater l'outil pour en faire un arbre de décision prédéfini ?
- Ou p*lutôt un arbre de décision qui projette les conséquences, et plus proche des caractéristiques et fonctions que l'on retrouve dans les systèmes d'IA ?*
- Les impacts indirects pouvant être bien plus importants que les impacts directs, il faudrait pouvoir les intégrer ou ajouter un autre outil pour cela (ex : https://docs.google.com/document/d/1s-nx3-66LWcMX3En4w7cw7JvmaPyRniTaYGM7EwFo1o/edit)
- La possibilité d'identifier facilement les éléments dimensionnants serait un vrai plus
- La question du modèle économique est à intégrer dans l'outil car les impacts ensuite peuvent être fondamentalement différents
- L'arbre de conséquence (ADEME etc.) est déjà très performant. Faudrait-il plutôt réfléchir à une aide à la production d'un arbre de conséquence spécifique à l'IA : décrire des causes-conséquences spécifiques à l'IA ? Ou en fait, peut-être besoin de dérouler l'outil caractéristique par caractéristique ? Pour vraiment faire apparaître les choix technologiques possibles, qui sont dimensionnants. Un arbre de conséquence caractéristique par caractéristique. Ou faire apparaître les effets indirects sur l'outil ?
- A qui cet outil se destine ? Un public divers avec une personne qui puisse décrire le CdC et traduire les nuances de choix possibles en impact sur la stratégie du service (modèle économique, public visé, valeur ajoutée/pertinence), une personne qui puisse le décliner en briques techno (compréhension technique des solutions) et une personne qui puisse traduire en impacts environnementaux (et liens avec les contraintes environnementales) ? Ou autre ?