



THESE PROFESSIONNELLE

MBA Management de l'Intelligence Artificielle

Institut Léonard de Vinci

Présentée par Vincent SIX

Le jeudi 30 juin 2022

L'IA FRUGALE, UN ENJEU INDISPENSABLE POUR LE DEPLOIEMENT DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE

Directeur de thèse : Joachim MASSIAS

Année universitaire 2021-2022



MBA Management de l'Intelligence Artificielle

L'IA FRUGALE, UN ENJEU INDISPENSABLE POUR LE DEPLOIEMENT DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE



Vincent SIX

Juin 2022

Directeur de thèse : Joachim MASSIAS

Remerciements

Mes remerciements vont tout d'abord à M. Joachim Massias, Directeur du MBA Management de l'Intelligence Artificielle à l'Institut Léonard de Vinci qui fut également mon directeur de thèse. Nos premiers échanges m'ont tout d'abord permis de préciser le sujet de ma thèse. Ce travail fut ensuite pour moi l'opportunité d'une véritable exploration, riche de découvertes et d'enseignements, sur un très vaste domaine relatif à l'environnement, au numérique, aux algorithmes et applications de l'intelligence artificielle, à l'innovation frugale et à la conduite du changement. Ses conseils avisés ont permis d'éclairer mon cheminement. Je remercie également M. Thierry Occre, Directeur Général de Junia ainsi que Mme Laurence Deboffe, Directrice des Ressources Humaines de m'avoir permis de suivre ce MBA. Je suis également très reconnaissant des apports et des échanges que j'ai pu avoir avec toutes les personnes que j'ai pu interviewer : M. Pierre-Louis Bescon, M. Laurent Cervoni, M. Côme Chatagnon, M. Vincent Courboulay, M. Serge Cousin, M. Manuel Davy, M. Jean-Denis Garo, M. Luc Julia, M. Arnault Loualalen, M. Bruno Maisonnier, M. Louis Naugès, M. Guillaume Navarre, M. Arnaud Soumet et Mme Caroline Truc. Je les remercie vivement pour le temps qu'ils ont bien voulu me consacrer. Je remercie également l'ensemble du corps professoral du MBA MIA pour leurs enseignements tout au long de l'année ainsi que mes camarades de promotion pour leurs apports et leur bonne humeur. Enfin, je remercie sincèrement mon épouse Nathalie pour son soutien ainsi que nos enfants Chloé et Manon pour leur patience durant mes longues heures d'isolement nécessaires à la réalisation de ce travail.

Listes des sigles et abréviations

ACV	Analyse du Cycle de Vie
ASIC	Application Specific Integrated Circuit (Circuit intégré à application spécifique)
COP	Conférence des Parties
CPU	Central Processing Unit (unité centrale de traitement)
DEEE	Déchets d'Équipements Électriques et Électroniques
DRAM	Dynamic Random Access Memory (mémoire vive dynamique)
ESG	Critères Environnementaux, Sociaux et de Gouvernance
ETI	Entreprise de Taille Intermédiaire
FGPA	Field Programmable Gate Array (réseau de portes programmables)
FLOPS	Floating-point Operation Per Second (opérations à virgule flottante par seconde)
GES	Gaz à Effet de Serre
GIEC	Groupe d'Experts Intergouvernemental sur l'Évolution du Climat
GPU	Graphics Processing Unit (processeur graphique)
IA	Intelligence Artificielle
IOT	Internet des Objets
MIPS	Material Input Per Unit of Service (Consommation de matériaux par unité de service)
ML	Machine Learning (apprentissage automatique)
NLP	Natural Language Processing (Traitement du Langage Naturel)
NPU	Neural Processing Unit (unité de traitement neuronal)
ODD	Objectifs de Développement Durable
ONU	Organisation des Nations Unies
OTI	Organisme Tiers Indépendant
PACTE	Plan d'Action pour la Croissance et la Transformation des Entreprises
PME	Petite ou Moyenne Entreprise
RSE	Responsabilité Sociale des Entreprises
SI	Système d'Information
TDP	Thermal Design Power (enveloppe thermique)
TEN	Technologies à Émissions Négatives
TPU	Tensor Processing Unit (unité de traitement tensoriel)

Résumé

Les impacts du réchauffement climatique se font de plus en plus sentir, les rapports du GIEC ainsi que les accords de Paris ont identifié l'urgence d'agir pour limiter les émissions de gaz à effet de serre et l'utilisation des énergies fossiles. Un certain nombre de recommandations et de feuilles de route voient le jour y compris pour tout l'écosystème du numérique. En effet, le développement exponentiel des technologies et des usages du numérique pose inévitablement la question de sa soutenabilité vis-à-vis de ses impacts environnementaux et de sa consommation énergétique. Ces dix dernières années, les puissances de stockage et de calcul ont permis le déploiement fulgurant d'algorithmes d'intelligence artificielle. Grâce à la facilité d'accès de ces technologies, de très nombreux cas d'usage ont été mis en place apportant ainsi de la valeur dans le modèle économique des entreprises. Ces nouveaux outils sont maintenant capables de grandes prouesses en termes d'optimisation, de prédiction, de reconnaissance d'images ou encore de traitement de langage naturel. Mais les algorithmes gèrent aujourd'hui de plus en plus de paramètres et appellent aussi à de plus en plus de ressources. Une approche plus frugale s'avère donc nécessaire pour pérenniser l'évolution des applications d'intelligence artificielle. Ce nouveau paradigme nécessite d'être porté au plus haut niveau d'une organisation pour parvenir à concilier profitabilité et responsabilité.

Mots clés : Intelligence artificielle – Frugalité – Environnement – Numérique responsable – Conduite du changement

Les personnes interviewées



Pierre-Louis BESCOND
Head of Data & Advanced Analytics – Roquette



Dr. Luc JULIA
Chief Scientific Officer – Renault Group



Laurent CERVONI
Directeur de la Recherche et de l'Innovation – Talan



Arnault LOUALALEN
CEO and R&D director - Numalis



Côme CHATAGNON
Partner Associé – AI Builders



Bruno MAISONNIER
CEO and Founder – AnotherBrain



Vincent COUBOULAY
Maitre de conférences HDR - La Rochelle
Université / Directeur scientifique - Institut du
Numérique Responsable



Louis NAUGES
Chief Strategy Officer - Wizy.io



Serge COUSIN
Président / Fondateur - CIV France



Guillaume NAVARRE
CCO & Cofondateur – Golem.ai



Manuel DAVY
PhD, Founder and CEO – VEKIA
Leader – Cité de l'IA Hauts-de-France



Arnaud SOUMET
Consultant Senior – Talan Consulting



Jean-Denis GARO
Head of Marketing – Golem.ai



Caroline TRUC
DRH et accompagnement du changement,
Coach certifiée HEC – Indépendante

Sommaire

L'IA FRUGALE, UN ENJEU INDISPENSABLE POUR LE DEPLOIEMENT DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE	1
REMERCIEMENTS	2
LISTES DES SIGLES ET ABREVIATIONS	3
RESUME	4
LES PERSONNES INTERVIEWEES.....	5
SOMMAIRE	6
INTRODUCTION	7
PARTIE 1 - LES ENJEUX ENVIRONNEMENTAUX	9
1. CONTEXTE ENVIRONNEMENTAL.....	9
2. LES DOCUMENTS CADRES POUR L'IA ET LE NUMERIQUE	13
3. TENDANCES SOCIETALES.....	18
PARTIE 2 - LES IMPACTS ENVIRONNEMENTAUX DU NUMERIQUE	21
1. L'EVALUATION DE L'IMPACT ENVIRONNEMENTAL	21
2. L'UNIVERS DU NUMERIQUE.....	27
3. L'EFFET REBOND.....	30
4. LES LABELS DU NUMERIQUE.....	31
PARTIE 3 - L'INTELLIGENCE ARTIFICIELLE ET SA SOUTENABILITE	32
1. QU'EST-CE QUE L'IA ?.....	32
2. L'ECOSYSTEME BIG DATA ET IA.....	38
3. L'EMERGENCE DE L'IA POUR L'ENVIRONNEMENT	43
4. LES IMPACTS ENVIRONNEMENTAUX DE L'IA.....	44
5. AMELIORATION DES PERFORMANCES ENERGETIQUES DE L'IA	46
PARTIE 4 - VERS UNE CULTURE DE LA FRUGALITE.....	52
1. L'INNOVATION FRUGALE.....	52
2. LE CADRE DU CHANGEMENT.....	53
3. LE PILOTAGE ET L'ACCOMPAGNEMENT AU CHANGEMENT.....	54
4. DU CHANGEMENT A LA TRANSFORMATION.....	61
5. IA FRUGALE, UN AUDIT DES PRATIQUES NECESSAIRE.....	63
CONCLUSION	65
BIBLIOGRAPHIE	66
ANNEXES	74
TABLE DES FIGURES	114
TABLE DES TABLEAUX.....	115
TABLE DES EQUATIONS.....	116
TABLE DES MATIERES.....	117

Introduction

Bien que les toutes premières approches d'Intelligence Artificielle (IA) remontent aux années cinquante, les technologies de l'IA connaissent depuis seulement une dizaine d'années un véritable essor. L'augmentation exponentielle des données produites, les capacités de stockage mais aussi les puissances de calcul ont rendu possible des applications qui n'étaient pas réalisables jusqu'à lors. L'IA est devenue porteuse de nombreuses sources de création de valeurs pour les entreprises. Personnalisation de l'expérience client, automatisation de flux, prédiction de commandes, analyse d'images, compréhension de texte et de langage, optimisation énergétique, jumeaux numériques, interaction homme-machine ne sont que quelques exemples. Les entreprises sont maintenant de plus en plus nombreuses à percevoir les potentialités de ces technologies (1). L'arrivée de l'IA dans notre vie personnelle et professionnelle représente une véritable révolution au même titre que les précédentes révolutions industrielles. Andrew Ng, cofondateur et directeur de Google Brain (un projet de recherche d'apprentissage profond), et professeur à l'université de Stanford va même jusqu'à la comparer avec l'arrivée de l'électricité dont les changements provoqués sur nos modes de vie ont été majeurs.

Pour répondre à ces enjeux, les acteurs du numérique mènent une véritable course à l'armement sur les moyens de collecte et de traitement des données, sur la complexité et la performance des algorithmes, sur le déploiement des infrastructures informatiques toujours plus puissantes. Toutes ces activités de recherche et de développement ont facilité l'accès à ces technologies et aux possibilités d'expérimentation. Les attentes sont fortes, il faut pouvoir disposer d'une qualité de service de haut niveau (sécurité des données, rapidité d'exécution, précision, ubiquité). Les progrès du secteur du numérique sont phénoménaux et bien plus importants que dans d'autres industries. Il s'agit là aussi de réels enjeux stratégiques et de souveraineté pour un pays.

Cependant, bien que riche de promesses, l'IA interpelle sur les risques qu'elle représente en termes d'éthique, de juridique, de confiance et d'environnement. Les préoccupations environnementales se font de plus en plus prégnantes et à juste titre. En effet, les derniers rapports du GIEC sont alarmants quant à la nécessité absolue d'agir pour le climat afin de limiter le réchauffement climatique. Si l'IA peut assurément contribuer à la réduction des émissions de gaz à effet de serre et des dépenses énergétiques de secteurs très polluants comme le transport, le bâtiment, l'agriculture, ne constitue-t-elle pas à la fois la solution et le problème ? Si dans la perception collective, le numérique est perçu comme peu impactant en raison d'une certaine immatérialité due à l'usage du cloud, qu'en est-il réellement ? Quels sont

en réalité les impacts de ces nouvelles technologies sur notre planète ? Comment peut-on les caractériser et les évaluer pour mieux agir et les minimiser ?

Pour venir étayer ces premiers éléments, le présent document est organisé en quatre parties.

La première partie intitulée « **Contexte environnemental** » vise à dresser un état des lieux au travers de l'analyse de rapports environnementaux, de documents cadres, de textes législatifs dont l'objectif est de mieux évaluer les enjeux climatiques, de mesurer la nécessité de s'engager dans une démarche éco responsable et d'anticiper ainsi des réglementations à venir vis-à-vis des technologies de l'IA.

La deuxième partie concerne « **les impacts environnementaux du numérique** ». Elle permet de caractériser les impacts directs et indirects du numérique et d'identifier la contribution du secteur vis-à-vis de la consommation électrique et des émissions de gaz à effet de serre.

La troisième partie, dédiée à « **l'intelligence artificielle et à sa soutenabilité** » présente tout d'abord les différents types d'IA et les algorithmes. Elle s'attache ensuite plus particulièrement à présenter des solutions d'optimisation matérielle et logicielle et d'identifier des outils d'évaluation de l'impact environnemental des applications.

Enfin la quatrième et dernière partie présentent les actions à mettre en place pour tendre « **vers une culture de la frugalité** ». Elle recouvre ainsi une dimension de conduite de changement pour faciliter l'adhésion des collaborateurs en s'appuyant sur des leviers de sensibilisation et de formation. Ces évolutions comportementales et culturelles reposent sur la mise en place d'un environnement propice à la co-construction ainsi qu'un véritable pilotage de la conduite du changement.

Sans nul doute l'IA constitue un outil technique porteur de progrès pensé dans l'intérêt général, mais comme le rappelle Luc Julia dans son dernier ouvrage (2), tout outil n'a de sens qu'en vue de ce pourquoi il a été conçu. Si un marteau peut constituer un outil extrêmement efficace pour enfoncer un clou, il peut également devenir une arme destructrice. En soit, le marteau n'est ni bon, ni mauvais, tout dépend de son usage. Pour éviter de voir tout problème comme un clou, il faut pouvoir disposer d'une multiplicité d'outils pouvant être associés, être en capacité de les utiliser avec discernement et agir ainsi en acteur responsable.



“ Il est grand temps de sortir la caisse à outils. Elle permet de ne pas voir « tout problème comme un clou » Aussi discret soit-il, chacun a un rôle à jouer. Le manche du marteau est à notre portée. ”

Dr Luc Julia - Chief Scientific Officer – Renault Group

Partie 1 - Les enjeux environnementaux

"Mieux vaut prendre le changement par la main avant qu'il ne vous prenne par la gorge"
Winston Churchill

1. Contexte environnemental

Les alertes sur le réchauffement climatique ont commencé à s'intensifier réellement dans les années quatre-vingt. La première conférence de l'ONU sur le climat (Conférence of the Parties - COP) s'est déroulée à Berlin en 1995. Bien que celle-ci n'ait pas abouti à des mesures concrètes, les 120 gouvernements présents prirent acte de la gravité de la situation. Deux ans plus tard, le protocole de Kyoto fixait comme objectif de réduire de 5% les émissions de gaz à effet de serre à l'horizon 2012. Aujourd'hui, les impacts du réchauffement climatique se font de plus en plus ressentir et il y a urgence à se tenir à des objectifs beaucoup plus ambitieux. Plusieurs éléments viennent renforcer la prise de conscience et incitent à l'action. De nombreux rapports, dont ceux du GIEC encore récemment publiés, ont permis de mieux comprendre l'impact des activités humaines sur la planète et ne permettent plus de nier le phénomène du réchauffement climatique. Régulièrement des accords sont établis par de nombreux états lors des différentes COP et des objectifs sont définis pour limiter notamment l'usage des énergies fossiles responsables des émissions de GES. Malgré cela, la situation environnementale n'est pas bonne et un certain nombre de feuilles de route et de textes de loi commencent à apparaître dans le domaine du numérique et de l'intelligence artificielle. C'est également sans compter sur les évolutions de la société où les citoyens plus sensibilisés et mieux formés aux problèmes environnementaux appellent à l'action et font preuve de militantisme. L'objectif de cette première partie est donc de dresser un tour d'horizon de ces différentes dimensions du contexte environnemental.

a. Les rapports et enquêtes

Rapport du GIEC

Créé en 1988 à l'initiative du G7, le Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC) réunit des centaines d'experts éminents chargés d'analyser scientifiquement les changements climatiques. Les rapports du GIEC fournissent aux gouvernements et aux décideurs les éléments pour définir leurs orientations et servent de base aux négociations des COP. Ils fournissent également des éléments pour l'évaluation des 17 objectifs de développement durable (ODD) des Nations unies. Cinq rapports d'évaluation ont déjà été publiés depuis 1990. Trois groupes de travail ont été mis en place au GIEC pour la rédaction du 6^{ème} rapport dont la synthèse est prévue pour le 3 octobre 2022 :

- Groupe 1 : Sciences physiques du changement climatique / rapport du 09 août 2021

- Groupe 2 : Impacts, adaptation et vulnérabilité / rapport du 28 février 2022
- Groupe 3 : Atténuation du changement climatique / rapport du 04 avril 2022

Le rapport du **groupe 1** se concentre sur 5 scénarios (cf. Annexe 1). Sur un scénario intermédiaire d'émission de CO₂, le rapport indique une forte probabilité que le seuil de +2°C de réchauffement soit dépassé au cours du 21^{ème} siècle. Le seuil de 1,5°C devrait être dépassé entre 2021 et 2040 et le réchauffement se poursuivra au moins jusqu'en 2050. Pour éviter un réchauffement de 2°C, il est indispensable de réduire immédiatement et fortement les émissions de GES mais également de façon durable et généralisée. Pour atteindre cet objectif, les pays industrialisés doivent diviser leurs émissions de gaz à effet de serre par **quatre** d'ici à 2100.

Le rapport du **groupe 2** porte sur **les impacts, les vulnérabilités et les capacités d'adaptation à la crise climatique**. Celui-ci met en évidence les impacts déjà observables du réchauffement climatique sur les écosystèmes et les sociétés humaines (cf. Annexe 1)

Le rapport du **groupe 3** traite plus particulièrement de **la lutte contre le réchauffement climatique**. L'accent est mis sur la nécessité d'agir au plus vite sur la réduction des gaz à effet de serre car ceux-ci n'ont cessé d'augmenter ces vingt dernière années (+12% de plus qu'en 2010, +54% de plus qu'en 2000). Si rien n'est fait l'augmentation de la température de la planète devrait atteindre +3,2°C et entraîner des catastrophes écologiques, économiques, sanitaires et sociales. Pour limiter le réchauffement sous des valeurs seuil de +1,5°C à 2°C, les émissions doivent être réduites en valeur absolue dès aujourd'hui et au plus tard dans les 2 à 3 prochaines années (cf. Annexe 1). Bien que les investissements et les efforts à consentir pour lutter contre le réchauffement climatique demeurent importants, les auteurs indiquent que ceux-ci seront bien **inférieurs aux coûts de l'inaction**.

Enquête sur la perception des risques mondiaux du World Economic Forum (3)

L'enquête sur la perception des risques mondiaux (Global Risks Perception Surver - GRPS) a été actualisée en Janvier 2022 afin de recueillir des informations nouvelles et plus larges auprès des quelques 1 000 experts et dirigeants mondiaux qui y ont répondu. Dans le cadre de cette enquête, les personnes interrogées ont classé en risque numéro un pour les dix prochaines années, l'échec à tenir les objectifs climatiques, arrivent ensuite les événements climatiques extrêmes et la perte de la biodiversité. Ce qui constitue une véritable prise de conscience des enjeux climatiques par les acteurs économiques.

Rapport Meadows

Un modèle de croissance insoutenable dénoncé depuis cinquante ans. Le 3 mars 2022 a été réédité le « rapport Meadows » (4) ou « Rapport du Club de Rome ». Publié pour la première fois en 1972 et plusieurs fois actualisé depuis, ce rapport démontrait pour la première fois que

la croissance économique ne pouvait continuer indéfiniment dans un monde fini. Pour répondre à la commande du Club de Rome, l'équipe de scientifiques du MIT (Institut de Technologie du Massachusetts) aux États-Unis a alors utilisé un modèle informatique pour montrer que la croissance continue de la production industrielle et de la pollution ainsi engendrée amèneraient à l'effondrement des ressources alimentaires et de la population. Cinquante ans plus tard, très peu des recommandations présentes dans le rapport ont été reprises par les dirigeants. Pour Dennis Meadows co-auteur du texte, il est indispensable de vivre avec moins et il nous faut réduire de moitié notre consommation énergétique. Selon lui, pour évoluer vers un monde « post croissance » plus sobre, il faut faire des changements culturels et cela doit passer par de la pédagogie et de l'apprentissage.

Le rapport de l'ADEME

Le rapport de l'ADEME (Agence de la Transition Énergétique) publié le 30 novembre 2021 (5) présente quatre scénarios pour parvenir à la neutralité carbone pour 2050. Si tous les scénarios visent le même objectif, les impacts sociétaux ne sont pas les mêmes suivant celui retenu. Facteur clé pour atteindre la neutralité carbone, la réduction de la demande en énergie peut aller de 23% à 55% par rapport à 2015. Les quatre scénarios s'intitulent : S1 – **Génération Frugale** (transition conduite par la contrainte et par la sobriété), S2 – **Coopérations territoriales** (transformation de la société par une gouvernance partagée), S3 – **Technologies vertes** (déploiement des technologies et du numérique pour l'efficacité énergétique), S4 – **Pari réparateur** (capacité à réparer les systèmes sociaux et écologiques)

The Shift Project

Créé en 2010, The Shift Project est un groupe de réflexion dont la mission est d'éclairer et d'influencer le débat sur la transition énergétique en France et en Europe. De nombreuses publications (rapports et notes d'analyse) sont réalisés par les membres de l'association présidée par Jean-Marc Jancovici. Parmi les thématiques traitées qui vont du bâtiment, de l'énergie, de l'industrie, du transport en passant par l'enseignement supérieur, les impacts environnementaux du numérique y trouvent également place. Le 7 février 2022, The Shift Project a présenté son « **Plan de transformation de l'économie française (PTEF)** » (6). Fruit d'un travail de deux ans, ce plan a été établi dans la perspective d'une neutralité carbone pour la France en 2050 et une réduction des émissions de GES de 5% par an pour limiter la hausse de la température à +2°C.

b. Les accords et objectifs

Accord de Paris

Adopté par 193 parties (dont l'Union Européenne) le 12 décembre 2015 à l'issue de la COP21, l'accord de Paris sur le climat est entré en vigueur le 4 novembre 2016. L'objectif de cet accord

est de lutter contre les changements climatiques et leurs effets nuisibles. Les objectifs à long terme visent à orienter les nations pour :

- Contenir d'ici à 2100 l'élévation de la température moyenne de la planète nettement en dessous de +2°C, soit 1,5°C par rapport aux niveaux préindustriels,
- Réévaluer à la hausse les engagements des pays signataires tous les 5 ans,
- Soutenir financièrement les pays en développement pour atténuer les changements climatiques,
- Renforcer la résilience et accroître les capacités d'adaptation aux effets induits par le réchauffement,

L'accord de Paris marque le début d'une transition vers la neutralité carbone pour limiter le réchauffement climatique. 150 états ont soumis en 2020 ou 2021 un nouvel engagement avec un objectif à la hausse mais bon nombre d'engagements de pays en développement sont conditionnés à un soutien financier. D'ici 2030, l'Union Européenne prévoit de réduire les émissions d'au moins 55% d'ici 2030 (l'objectif fixé en 2014 était de 40%). L'objectif final de l'Union Européenne est d'atteindre la **neutralité carbone d'ici 2050**. Les émissions mondiales restent cependant en augmentation avec une augmentation sensible d'environ 16% en 2030 par rapport à 2010 (59% de plus qu'en 1990). L'évaluation des progrès au niveau mondial est prévue en 2023 puis tous les 5 ans, la prochaine actualisation des plans en 2025.

COP26

Du 31 octobre au 12 novembre 2021 s'est tenue à Glasgow la COP26, l'occasion de vérifier les résultats des stratégies et des actions décidées dans le cadre de l'accord de Paris. Bien que les négociations aient été difficiles et que les ONG écologistes ont qualifié le « pacte de Glasgow » d'accord en demi-teinte, plusieurs engagements ont été pris pour limiter le réchauffement climatique. Les énergies fossiles, notamment le charbon, sont clairement ciblées. En effet, le pacte prévoit de diminuer le recours aux énergies fossiles (une première historique) et de ne plus les financer d'ici fin 2022 au profit des énergies renouvelables. L'objectif d'une neutralité carbone pour l'horizon 2050 est maintenu et plus de cent pays ont fait l'engagement de réduire leurs émissions de méthane de 30% minimum d'ici 2030.

Les objectifs de l'ONU sur le développement durable (ODD)

Le développement durable s'appuie sur 3 piliers que sont la dimension sociale, la préservation de l'environnement et la performance économique, repris également sous l'acronyme 3P (*People, Planet, Profit*). Les 17 Objectifs de Développement Durable (ODD), adoptés par l'Organisation des Nations Unies (ONU) en 2015, constituent l'Agenda 2030. A chaque objectif sont associées des cibles (169 au total) concernant l'éradication de la pauvreté, la protection de la planète et la garantie de la prospérité pour tous à atteindre d'ici 2030 (cf. Annexe 2). Chaque acteur (États, entreprises, investisseurs, ONG, citoyens, collectivités locales, écoles,

...) sont appelés à se mobiliser sur les ODD dans leurs domaines de compétences. Bien qu'adoptés en 2015, l'agenda 2030 et les ODD ne sont pas sur la bonne voie pour être atteints. En 2020, la crise sanitaire, sociale et a impacté fortement plusieurs ODD. (7)

2. Les documents cadres pour l'IA et le numérique

Depuis quelques années, plusieurs rapports et de feuilles de route sont établis pour cadrer les usages de l'IA et guider les stratégies d'investissement. Une IA sobre et au service de l'environnement devient de plus en plus présente dans les orientations.

a. Rapports et feuilles de route

Rapport de mission parlementaire – Cédric Villani – Mars 2018 (8)

En 4^{ème} partie de son rapport de mission intitulé « Donner un sens à l'intelligence artificielle - Pour une stratégie nationale et Européenne », Cédric Villani pose le cadre d'une « Intelligence artificielle au service d'une économie plus écologique ». En préalable, il indique que d'ici 2040, la production énergétique mondiale serait dépassée par l'énergie nécessaire pour les besoins de calcul. L'IA ne fera d'ailleurs que renforcer les besoins de stockage et d'échange de données (dont la croissance en volume est exponentielle), le besoin de puissance de calcul, le renouvellement des matériels pour améliorer les performances. L'IA représente cependant une opportunité pour optimiser la gestion des ressources énergétiques et préserver l'environnement. Mais une vision réellement ambitieuse de l'IA devra aller au-delà et permettre la conception de modèles techniques et économiques plus sobres. A retenir de la 4^{ème} partie du rapport de mission parlementaire – Cédric Villani :

- Mettre en place une politique claire en faveur de l'évaluation écologiques des solutions numériques
- Constituer une base nationale pour généraliser ce type d'évaluation
- Créer une plateforme pour la mesure de l'impact environnemental des solutions d'IA
- Encourager les entreprises à l'écoconception
- Penser l'innovation de rupture dans le domaine des puces électroniques
- Accompagner l'industrie du cloud européen dans son optimisation énergétique
- Mettre en place d'un label pour soutenir les démarches écologiques des fournisseurs de cloud européens
- Libérer les données écologiques pour la compréhension et la préservation de l'environnement
- Mettre à disposition les données publiques disponibles afin de permettre le développement de solutions d'intelligence artificielle au service de la transition écologique.

Stratégie Nationale - France 2030 : une nouvelle ambition en matière d'intelligence artificielle

La deuxième phase de la stratégie nationale d'intelligence artificielle, annoncée le 8 novembre 2021 (9), repose sur un plan d'investissement de 2,22 Mds€ (dont 1,5 Mds€ de financement publics) sur cinq ans (2021-2025). 56% des 1,5 Mds€ de financements publics concernent la formation, 33,5% le soutien à l'innovation et à l'économie et 10,5% pour la recherche scientifique et le transfert. Les principales mesures envisagées concernent :

- Un programme pour faire émerger des pôles de formation d'excellence et un plan de formation massif à l'IA (781M€)
- Le développement massif d'IA embarquée, d'IA responsable et de confiance (1,22Mds €)
- L'accompagnement de 500 PME et ETI pour l'adoption et l'intégration de solutions d'IA (40M€)
- Le passage à l'échelle des incubateurs/accélérateurs de startup issus de la recherche (40M€)
- Le lancement de démonstrateurs d'une IA frugale en énergie (120M€)



“ Au-delà de la dimension environnementale, l'IA frugale est un enjeu de souveraineté nationale et européenne face à l'avance prise par les américains et les chinois sur les technologies du Big Data. Mais c'est aussi un besoin de certaines industries et du domaine militaire qui n'ont d'ailleurs pas toujours accès à de nombreuses données. ”

Arnault Loualalen - CEO and R&D director - Numalis

Recommandation sur l'éthique de l'intelligence artificielle UNESCO Novembre 2021 (10) (11)

Le 25 novembre 2021, la Directrice générale de l'UNESCO a présenté la toute première norme mondiale sur l'éthique de l'intelligence artificielle qui a été adoptée officiellement par les 193 États membres. Fruit du travail depuis mars 2020 d'un groupe de centaines d'experts internationaux indépendants issus de tous les continents, ce texte n'est pas contraignant pour les États signataires. Néanmoins, ce cadre a le mérite d'exister et pourrait inspirer de futures réglementations nationales. La recommandation définit les valeurs et les principes communs pour guider la mise en place de l'infrastructure juridique nécessaire pour garantir un développement sain de l'IA afin de favoriser les droits de l'Homme et de contribuer à la réalisation des Objectifs de Développement Durable (ODD). Les quatre grands axes de la recommandation sont : Protection des données, Interdiction de la notation sociale et de la surveillance de masse, Aide au suivi à l'évaluation de l'impact des systèmes d'IA, **Protection de l'environnement**. Au moment de l'adoption de la recommandation, une vingtaine d'États ont indiqué avoir déjà commencé à mettre en place des mesures. Hors de l'UNESCO, Israël et les États-Unis ne sont pas signataires. A retenir de la recommandation sur l'éthique de l'intelligence artificielle – UNESCO :

- Favoriser le déploiement de solutions d'IA économes en données, en énergie et en ressources

- Mesurer l'impact environnemental direct et indirect tout au long du cycle de vie du système d'IA.
- Diminuer l'impact environnemental des infrastructures de données et des systèmes d'IA
- Investir dans les technologies vertes
- Ne pas utiliser les solutions d'IA dont l'impact négatif sur l'environnement est disproportionné

Feuille de route sur l'environnement et le numérique du Conseil national du numérique (Cnum)

Publiée le 11 février 2020, cette feuille de route (12), propose 50 mesures pour un numérique responsable au niveau national et européen. Les trois chantiers développés concernent la réduction de l'empreinte du numérique sur l'environnement, l'exploitation du potentiel du numérique pour la transition écologique et solidaire, l'accompagnement de l'ensemble de la société vers un numérique responsable. Le chantier 2 comporte un objectif sur la création d'une intelligence artificielle responsable en cohérence avec la transition écologique et solidaire et les ODD. A retenir de la feuille de route sur l'environnement et le numérique du Conseil national du numérique (Cnum) :

- L'ambition d'une IA responsable au niveau national, européen et international fait partie des mesures mais ne doit pas céder au solutionnisme. Les technologies low-tech et/ou non numériques doivent faire partie des réflexions.
- Parmi les mesures prioritaires figure le déploiement d'une stratégie pour une IA sobre et au service des ODD.

Feuille de route gouvernementale « Numérique et Environnement »

La feuille de route gouvernementale « Numérique et Environnement » (13) du 23 février 2021 vise à faire converger les transitions numérique et écologique et s'articule autour de 3 axes (14) : **Connaître pour agir** : Développer la connaissance de l'empreinte environnementale numérique / **Soutenir un numérique plus sobre** : Réduire l'empreinte environnementale du numérique / **Innover** : Faire du numérique un levier de la transition écologique. A retenir de la feuille de route gouvernementale « Numérique et Environnement » :

- La création d'un baromètre environnemental des acteurs du numérique
- L'initiation d'une mission sur l'évaluation de l'impact environnemental du numérique en France à horizon 2030.
- Une sensibilisation sur les pratiques numériques moins polluantes au travers d'une campagne nationale
- L'engagement, en concertation, des acteurs du numérique pour favoriser la collecte et le reconditionnement des smartphones et de l'amélioration de leur efficacité environnementale

- Le renforcement du financement des GreenTechs

Rapport du Sénat - Pour une transition numérique écologique

Publié en juin 2020 au nom de la commission de l'aménagement du territoire et du développement durable, le rapport du sénat n°555 (15) (16) est organisé en 4 grands axes. Il identifie ainsi 25 propositions pour limiter l'empreinte environnementale du numérique. Le premier axe concerne la **prise de conscience** du grand public (au travers d'actions de sensibilisation et de formation) et l'intégration de **l'enjeu du numérique responsable** dans les stratégies des acteurs économiques. Le deuxième axe s'attache à **limiter le renouvellement des terminaux** responsables principalement des émissions de GES du numérique. L'axe 3 vise au développement **d'usages plus vertueux en matière d'écologie** et prône une **gestion durable de la donnée**. L'axe 4 cible plus particulièrement la **réduction de la consommation énergétique des data centers et des réseaux**.

Feuille de route stratégique AFNOR

L'AFNOR a publié le 29 mars 2022 une feuille de route stratégique relative à la normalisation de l'intelligence artificielle (17). Cette démarche de normalisation pour une IA de confiance viendra en appui des directives actuellement en projet au sein de la Commission Européenne (AI Act). Les 6 axes stratégiques de normalisation portent sur :

- La confiance (explicabilité, robustesse, équité, responsabilité, supervision humaine, sécurité, sûreté)
- La gouvernance et le management de l'IA (compétences, méthodes, outils)
- La supervision et le reporting de l'IA
- Les compétences des organismes de certification
- Certains outils numériques (simulation, jumeaux numériques...)
- La simplification de l'accès et de l'utilisation des normes.

Plusieurs « angles morts » sont identifiés dans la feuille de route et devraient être précisés dans une prochaine version de la feuille de route nationale. Parmi ces « angles morts » figure notamment la dimension liée à une IA soutenant le développement durable.

b. Le cadre législatif

S'il n'y pas encore de loi régissant l'usage de l'IA d'un point de vue environnemental, un certain nombre de textes législatifs viennent poser les bases d'une IA dite de confiance et orientent les actions en termes de numérique responsable. Au-delà de l'incitation, ces textes visent à structurer, à cadrer mais aussi à valoriser les démarches des entreprises dans ce domaine.



“ Même si une réglementation n’est pas évidente à mettre en place et serait difficile à faire respecter, elle aurait l’intérêt d’éduquer et de sensibiliser les gens afin d’utiliser les ressources à bon escient et de ne pas faire n’importe quoi. ”

Dr Luc Julia - Chief Scientific Officer – Renault Group

Projet de règlement européen sur l’IA

Un premier projet de règlement européen a été dévoilé le 21 avril 2021 par Commission Européenne dont l’objectif est d’assurer une **IA de confiance**. Il classe les usages de l’IA en quatre catégories de risque (inacceptable, élevé, acceptable, minime). Il prévoit également la mise en place d’un conseil européen dédié à l’IA ainsi qu’un système de sanction dissuasif pouvant aller jusqu’à 30 millions d’euros ou 6% du chiffres d’affaire annuel mondial. Ce projet n’aborde pas la dimension environnementale et doit encore être soumis au Parlement Européen.

Loi REEN

La loi REEN visant à « Réduire l’empreinte environnementale du numérique en France » (Loi n°2021-1485) a été adoptée définitivement par le sénat le 15 novembre 2021. La France fut le premier pays au monde à définir un texte législatif visant à prévenir, encadrer et réduire les impacts du numérique sur l’environnement. La loi vise à responsabiliser tous les acteurs du numérique : consommateurs, professionnels du secteur et acteurs publics (18). La loi REEN porte sur 5 volets : « **Faire prendre conscience** de l’impact environnemental du numérique », « **limiter le renouvellement** des appareils numériques », « favoriser des **usages numériques écologiquement vertueux** », « promouvoir des **datacenters et des réseaux moins énergivores** », « promouvoir une **stratégie numérique responsable** dans les territoires ». A retenir de la loi REEN :

- L’introduction de la sobriété numérique dans les enseignements de la primaire au supérieur est prévue par la loi. L’écoconception des services numériques sont envisagés pour les étudiants ingénieurs au travers de modules spécifiques.
- Un référentiel général de l’écoconception des services numériques doit être défini par l’ARCEP (Autorité de Régulation des Communications Électroniques, des Postes et de la distribution de la presse) et l’ARCOM (ex-CSA). Prévu pour le 1^{er} janvier 2024, celui-ci définira des critères de conception durable des services numériques, afin d’en réduire l’empreinte environnementale.
- Un « observatoire des impacts environnementaux du numérique » destiné à l’analyse et la quantification des impacts directs et indirects du numérique sur l’environnement est mentionné à l’article 4 de la REEN. A noter qu’une mention concerne aussi la contribution apportée par le numérique, notamment **l’intelligence artificielle**, à la transition écologique et solidaire. Il élabore une définition de la sobriété numérique.

Loi Pacte

La loi PACTE (Plan d'Action pour la Croissance et la Transformation des Entreprises) du 22 mai 2019 a défini un cadre sur les enjeux sociaux et environnementaux (ESG) des entreprises. Elle introduit 3 niveaux de contribution. Le **niveau 1**, obligatoire pour toutes les entreprises, concerne la prise en compte des enjeux sociaux et environnementaux dans leurs activités. Le **niveau 2** pour les entreprises volontaires permet de reconnaître une « **raison d'être** » pouvant être inscrite dans les statuts. Cela implique ainsi de porter un projet d'entreprise sur le long terme qui puisse répondre à un intérêt collectif et donner du sens à l'action de l'ensemble des salariés (exemple de Danone : « apporter la santé par l'alimentation au plus grand nombre »). Enfin le **niveau 3** valorise les « **entreprises à missions** » dont l'objectif est d'être au service de la résolution des défis sociaux et environnementaux sans pour autant renoncer à leur modèle lucratif. Une instance de suivi où les salariés sont représentés doit vérifier l'adéquation entre la raison d'être de l'entreprise et les pratiques de l'entreprise. Enfin, un organisme tiers indépendant est chargé de vérifier au moins tous les 2 ans l'atteinte des objectifs retenus. Depuis la promulgation de la loi PACTE, le nombre d'entreprises à missions connaît une progression à deux chiffres chaque trimestre pour atteindre 505 fin 2021. 530000 salariés travaillent maintenant dans ce modèle d'entreprise, ce nombre a été multiplié par 2,5 en un an. 79% des entreprises à missions sont des entreprises de moins de 50 salariés alors que les grandes entreprises et les ETI (0,2% des entreprises en France) représentent 9%.

3. Tendances sociétales

a. Génération climat

Friday of the future

Initié en aout 2018 par la jeune suédoise Greta Thunberg, le mouvement « Friday of the future » (les vendredis pour le futur) a pour objectif de faire pression sur les dirigeants afin qu'ils respectent leurs engagements en matière de lutte contre les changements climatiques. Les militants organisent des grèves scolaires sur tous les continents pour participer à des manifestations et à des rencontres afin de sensibiliser et discuter des actions à mener. Le mouvement revendique 14 millions de personnes, issus de 7500 villes sur tous les continents. La dernière manifestation a eu vendredi 25 mars 2022 et a rassemblé des milliers de manifestants dans plus de 80 pays sous le slogan « PeopleNotProfit » (« Les gens, pas le profit »). En France, c'est notamment la quasi-absence des thématiques climat dans la campagne pour l'élection présidentielle qui a été dénoncée par les participants.

Militantisme étudiant

Le 16 février 2022, des étudiants des prestigieuses universités américaines Yale, Princeton, Stanford, Vanderbilt et MIT ont porté plainte pour obliger leur institution à se désengager des industries polluantes (19). La plainte vise les allocations annuelles des cinq universités privées qui s'élèveraient à plus de 150 milliards de dollars au total. Réunis au sein de **l'ONG Climate Defense Project**, les étudiants font valoir que les dommages causés sur le climat, l'écologie et la santé humaine par les industries des combustibles fossiles sont en totale contradiction avec les objectifs éducatifs et les missions de leurs écoles.

En France, le collectif « **Pour un réveil écologique** », créé en septembre 2018, a lancé un manifeste Étudiant pour un Réveil Écologique pour mobiliser leur établissement sur les questions environnementales, notamment en termes de formation, et pour aider les diplômés à choisir une entreprise suffisamment engagée dans la transition écologique. Plus de 33000 étudiants sont signataires du manifeste et 400 établissements sont mobilisés en Europe. En août 2019, deux membres du collectif sont intervenus à l'université d'été du Medef pour alerter les entreprises sur les difficultés à venir pour recruter des jeunes diplômés si elles ne s'engageaient pas réellement pour l'environnement.

En janvier 2020, les étudiants de l'école polytechnique ont voté contre l'implantation de l'entreprise Total sur le Campus de Saclay. Face à la fronde des élèves et des associations, TotalEnergies a annoncé renoncer à implanter son centre de recherche qui devait accueillir 400 personnes à proximité de l'école.

Collectifs d'employés

Le 4 novembre 2019, plus de 1000 salariés de **Google** ont adressé une lettre ouverte à leur entreprise pour qu'elle mette en place des mesures concrètes et plus ambitieuses en faveur de l'environnement afin d'atteindre zéro émission carbone à l'horizon 2030. La lettre réclamait aussi l'arrêt de tout financement de groupements climato-sceptiques. Auparavant, en avril 2019, plus de 8000 salariés avaient de la même manière interpellé Jeff Bezos, fondateur et PDG d'**Amazon** pour dénoncer l'impact des activités de leur entreprise sur l'environnement. Dans l'esprit des grèves initiées par le mouvement « Friday of the future », des employés d'Amazon, Google et **Microsoft** se sont joints aux manifestations écologistes pour demander à leurs employeurs de réduire leur impact sur l'environnement. Des investissements importants avaient alors été annoncés la veille de l'événement par le PDG de Google et Jeff Bezos promettait le jour même de passer à 100 % d'énergie renouvelable dans les installations et les opérations d'Amazon d'ici 2030.

b. Paradoxe des digital natives

Une étude du CREDOC (20) réalisée pour l'ADEME en 2019 a révélé que l'environnement était devenu un enjeu majeur pour les 18-30 ans puisqu'ils le classent en tête de leurs préoccupations devant l'immigration et le chômage. Leur engagement dans des associations de défense pour l'environnement a connu une forte progression depuis 2016. 12% des 18-24 ans ont déclaré faire partie d'une association environnementale en 2019 (contre 3% en 2015), à comparer aux 4% pour l'ensemble de la population. Cependant cette génération « digital native » qui est née et a grandi dans un environnement numérique contribue sans réellement s'en rendre compte à la pollution « invisible » du numérique. Plus de 4h par jour, c'est le temps moyen passé devant un écran pour visionner essentiellement des vidéos en ligne. Un rapport de The Shift Project (21) indique que le visionnage de vidéos en 2018 a généré plus de 300 MtCO₂ soit autant que les émissions annuelles d'un pays comme l'Espagne. Plus largement le rapport pose la question de l'évolution des usages dans la transition écologique.

c. Des jeunes davantage formés

Sous l'impulsion de la loi REEN qui prévoit l'enseignement de la sobriété numérique de la primaire au supérieur, la formation des jeunes, déjà sensibilisés aux enjeux climatiques, devrait donc se développer ces prochaines années. Le rapport du climatologue Jean Jouzel (22), publié en février 2022, préconise de former 100% des étudiants du supérieur d'ici cinq ans aux enjeux de la transition écologique et du développement durable. The Shift Project a également publié le 10 mars 2022 un rapport final pour transformer les formations d'ingénieur afin d'intégrer dans des cours obligatoires les enjeux socio-écologiques (23). Pas moins de 200h de formation sur 5 ans permettraient de répondre à ce nouveau référentiel de connaissances et de compétences dont les entreprises sont demandeuses pour honorer leurs engagements écologiques et sociétaux.



“ Plusieurs jeunes embauchés chez Talan, soucieux des enjeux environnementaux, ont mis en place une communauté climat avec laquelle notre centre de recherche travaille. C'est venu d'initiatives personnelles, les jeunes diplômés sont plus sensibles à ce type de démarche. ”

Laurent Cervoni - Directeur de la Recherche et de l'Innovation – Talan

Partie 2 - Les impacts environnementaux du numérique

"Il y a assez de tout dans le monde pour satisfaire aux besoins de l'Homme, mais pas assez pour assouvir son avidité"
Mahatma Gandhi

Le « numérique » englobe tous les équipements électroniques qui utilisent des données binaires (24). On y retrouve bien sûr les ordinateurs, les imprimantes, les smartphones, les consoles de jeu, les télévisions connectées à une box. Mais c'est sans oublier des milliers de centres de données, de millions de kilomètres de câbles (cuivre et fibre), de millions d'antennes-relais pour le transport et le traitement des données.

1. L'évaluation de l'impact environnemental

Une étude du GreenIT (25) analyse la quantification de l'empreinte numérique mondiale et son évolution entre 2010 et 2025. Le sujet est vaste et repose sur la mise en place d'indicateurs représentatifs, de méthodes rigoureuses permettant de tenir compte de l'ensemble des équipements nécessaires au fonctionnement de nos activités numériques sur toute leur durée de vie. Comme l'indique cette étude, le numérique est loin d'être immatériel et sans impact sur l'environnement. Si le numérique était un pays, son empreinte serait environ 2 à 3 fois celle de la France, soit 2,5 % de l'empreinte de l'humanité. Les projections à 2025 amèneraient cette contribution à l'empreinte mondiale à un peu moins de 6% et un quintuplement du poids du numérique depuis 2010. Cela pose donc la question de la soutenabilité du numérique vis-à-vis de l'épuisement des matières premières et de la consommation énergétique. Pour mieux comprendre les facteurs d'impacts, cette partie vise à revenir sur les indicateurs associés, les méthodes d'évaluation et l'impact des différents équipements.

a. Les indicateurs environnementaux

Les indicateurs environnementaux qui permettent de bien qualifier les impacts du numérique portent généralement sur l'épuisement des ressources naturelles non renouvelables (ressources abiotiques), les émissions de gaz à effet de serre, la consommation d'eau, la consommation d'énergie primaire et la consommation électrique (énergie finale). Afin de faciliter l'appropriation de ces indicateurs et de qualifier leur impact, il est essentiel de revenir sur chacun d'entre eux.

Épuisement des ressources abiotiques (ADP)

Cet indicateur évalue l'impact du numérique vis-à-vis de l'épuisement des stocks de minerais nécessaires à la fabrication des composants électroniques. Il s'exprime en kg équivalent antimoine (kg eq. SB). L'antimoine, élément chimique de symbole atomique Sb, a été choisi

comme ressource rare de référence. Une ressource plus rare que l'antimoine présentera donc un coefficient de rareté supérieur à 1. De nombreux métaux sont nécessaires pour fabriquer des équipements numériques. Pas moins de 40 métaux sont présents dans un smartphone dont certains sont rares voire critiques compte tenu des réserves actuelles. Le numérique repose surtout sur des métaux tels que le tantale, le gallium, le ruthénium, le germanium, l'indium dont le recyclage est très faible (inférieur à 1%). Leur extraction des terres rares est extrêmement génératrice de GES et consommatrice d'eau.

La production de tantale, minerai extrait du coltan, est utilisée entre 60% et 80% par le secteur de l'électronique pour la fabrication de composants électroniques de faible puissance (condensateurs et filtres à ondes de surface indispensables pour les smartphones) (24). Un impact qui va malheureusement bien au-delà de l'épuisement des ressources. En effet, la région de Kivu en République démocratique du Congo concentre 60% à 80% des réserves mondiales de ce minerai ce qui entraîne en raison de la très forte demande de l'industrie électronique des conflits armés, de nombreux cas d'esclavage, de travail d'enfants et des menaces sur les espèces locales. Entre 2016 et 2018, le prix du kg de tantale est passé de 20 000€ à 95 000€. L'Annexe 3 présente les principaux pays producteurs de minerais rares.

Émission de gaz à effet de serre (GES)

Les gaz à effet de serre contribuent au réchauffement de la planète en retenant une partie de la chaleur reçue par le soleil dans l'atmosphère. Les GES issus de l'activité humaine sont principalement le CO₂ (dioxyde de carbone), CH₄ (méthane), N₂O (protoxyde d'azote) et les gaz fluorés. Dans ces derniers, on retrouve d'ailleurs les NF₃ (trifluorure d'azote) issus de la fabrication de composants électroniques. Cet indicateur d'émission de GES est exprimé en kg équivalent CO₂ (kg eq. CO₂) car tous les gaz n'ont pas le même pouvoir de réchauffement suivant leur durée de vie et leur capacité à réfléchir les rayonnements solaires. Le coefficient d'équivalence est donc de 1 pour le CO₂, 25 pour le CH₄ et atteint la valeur record de 17200 pour le NF₃. La Figure 1 présente la part des émissions mondiales du CO₂ dues au numérique en comparaison avec une sélection de pays. Le numérique contribue davantage aux émissions qu'un pays comme le Japon, lui-même en 5^{ème} position des pays les plus pollueurs du monde derrière la Chine (30,7%), les États-Unis (13,8%), l'Inde (7,1%), la Russie (4,6%). (26)

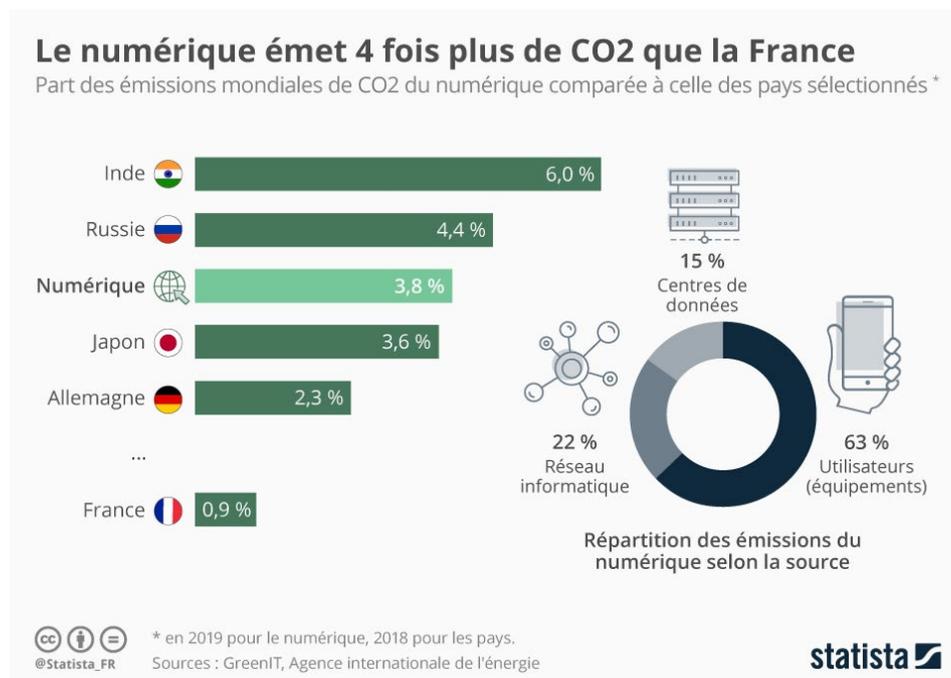


Figure 1 - Part des émissions mondiales de CO₂ du numérique

Tout aussi étonnant, c'est plus de quatre fois les émissions de CO₂ produites par la France. Les géants américains du secteur de la technologie prennent des engagements pour réduire voire annuler leur empreinte carbone. Les objectifs visés concernent la « neutralité carbone », le « zéro émission nette », le « Carbone négatif ». De manière stricte, la neutralité carbone s'entend pour le CO₂ mais elle est souvent utilisée pour l'ensemble des GES.

Le Tableau 1 permet de retrouver de manière synthétique les engagements associés au différentes cibles carbone (neutralité, zéro émission nette et carbone négatif).

Cibles carbone	Engagements
Neutralité carbone	Équilibrer les émissions de CO ₂ et leur retrait dans l'atmosphère
Zéro émission nette (net zero)	Réduire les émissions de CO ₂ aussi proche que possible de zéro.
Carbone négatif (carbon negative)	Éliminer plus de CO ₂ que ce qui est émis.

Tableau 1 - Cibles carbone et engagements

Ces dernières années, les GAFAM ont beaucoup intensifié leur communication vis-à-vis de leur engagement environnemental, multipliant les annonces sur leurs objectifs. Microsoft prévoit d'ailleurs d'effacer son empreinte carbone d'ici 2050 (cf. Annexe 4). Pour parvenir à devenir carbone négatif, les entreprises devront s'appuyer sur les Technologies à Émissions Négatives (TEN) dont le principe est de prélever les GES dans l'atmosphère et de les stocker sur terre, dans le sous-sol ou dans les océans.

Les processus associés à cette approche peuvent être biologiques, technologiques ou géochimiques (cf. Annexe 5). Le processus biologique, même s'il présente quelques

inconvenients, reste le plus simple à mettre en place par rapport aux deux autres dont les impacts ne sont pas encore suffisamment bien évalués.

Bien que controversées, les TEN font partie des priorités politiques de l'Union Européenne pour atteindre la neutralité carbone au plus tard en 2050. Un projet de certification lié au retrait du carbone est actuellement en cours au sein de la Commission Européenne afin de garantir l'intégrité environnementale mais également de permettre la montée en puissance d'absorption de carbone par l'octroi de financements (publics, privés) supplémentaires.

Mais selon l'ADEME, **la neutralité carbone ne s'entend qu'à l'échelle de la planète et des Etats** (27). Plusieurs arguments sont invoqués pour cela :

- Le *périmètre* car la part prépondérante des émissions d'une entreprise concerne généralement le scope 3 (émissions indirectes)
- L'*équité* car certaines entreprises devraient devenir carbone négatif et pas seulement neutre en carbone compte tenu des possibilités de séquestration sur leur territoire.
- L'*efficacité* car le rachat de crédits carbone bon marché pour compenser les émissions n'engage pas à la mise en place d'une véritable stratégie de réduction de ses émissions.

L'ADEME rappelle que la réduction des émissions de GES (prioritaire) et la séquestration de CO₂ dans des puits biologiques ou technologies sont les deux leviers pour atteindre la neutralité carbone. L'agence propose également des recommandations sur l'utilisation de l'argument « neutralité carbone » pour adopter une démarche de communication responsable (28).

Consommation d'eau (EAU)

L'eau est indispensable à la vie. Généralement, on distingue l'eau bleue, directement mobilisable par l'Homme, de l'eau verte mobilisable uniquement par les végétaux. Cet indicateur est exprimé en litre d'eau bleue (l ou m³). Comme l'agriculture et d'autres industries, le numérique est un grand consommateur d'eau bleue et peut contribuer à des périodes de stress hydrique. L'« empreinte eau » du numérique dans le monde représente plus de deux fois la consommation d'eau des Français, soit 8 milliards de mètres cubes d'eau (24). Dans la prochaine décennie, la croissance démographique combinée au réchauffement climatique devrait entraîner une demande en eau douce 40% supérieure à l'offre selon les Nations Unies. Quelques données chiffrées sont intéressantes à considérer. La consommation en eau d'un datacenter de taille moyenne équivaldrait à trois hôpitaux de taille moyenne et dépasserait celle de deux terrains de golf de 18 trous. (29). Si les 86% de l'eau consommée sont recyclés, le fabricant de puces taiwanais TSMC utiliserait tout de même plus de 150 000 tonnes d'eau par jour (30). Enfin, la fabrication d'un ordinateur portable de 2 kg ne nécessite pas moins de 1,5 tonne d'eau.

Énergie primaire

L'énergie primaire permet de fabriquer l'énergie finale (énergie mécanique, électrique, thermique). Cet indicateur est exprimé en Wattheure (Wh) ou kiloWattheure (kWh) par unité de temps. L'extraction des minerais et leur transformation en composants électroniques représentent l'essentiel de l'empreinte du numérique. L'énergie grise nécessaire à leur fabrication est largement supérieur à l'électricité consommée durant toute leur utilisation. On considère donc pour un équipement ou pour un service numérique le bilan énergétique global sur le cycle de vie complet. A titre d'exemple, la fabrication d'un smartphone mobilise 80% de la dépense énergétique et seulement 20% pour la consommation électrique. (24)

Consommation électrique

L'électricité (énergie finale) ne représente qu'une partie de l'énergie consommée par le numérique. Cet indicateur s'exprime également en kiloWattheure (kWh) par unité de temps. D'après un rapport du Shift Project (31), le numérique représenterait près de 10% de la consommation mondiale d'électricité. Cette consommation augmente d'ailleurs de 5% à 7% par an et pourrait ainsi mobiliser 20% de la production mondiale à horizon 2025. Cette consommation se répartie à hauteur de 30% sur les terminaux (ordinateurs, smartphones), 30% sur les data centers et le reste (40%) sur les réseaux. (32)

b. Les approches d'évaluation

Principalement, deux approches permettent d'évaluer l'impact environnemental des produits et des activités d'une entreprise, le **Bilan Carbone** et l'**Analyse du Cycle de Vie** (ACV). Complémentaires, ces outils sont très utiles pour mettre en place une politique environnementale dans une organisation. Il existe également un certain nombre de guides fournissant de bonnes pratiques pour réduire l'impact environnemental d'un service numérique.

Bilan Carbone

Conçu par l'ADEME, le bilan carbone permet d'évaluer les émissions directes et indirectes de gaz à effet de serre qu'une entreprise génère sur l'ensemble de ces activités. Cette analyse lui permet ensuite de mettre en place un plan de réduction de ses émissions. Par simplification, les données sont exprimées en équivalent CO₂ pour les autres gaz à effet de serre. Il existe 3 différentes catégories d'émissions (cf. Annexe 6) :

Scope 1 : émissions directes produites par les sources détenues et contrôlées par l'entreprise

Scope 2 : émissions indirectes liées à la consommation d'énergie (électricité, chauffage, climatisation)

Scope 3 : émissions indirectes (amont et aval) liées à tout le reste

En France, les entreprises de plus de 500 salariés (ou de plus de 250 salariés dans les DOM), les collectivités territoriales de plus de 50 000 habitants, les établissements publics et services d'État de plus de 250 agents ont l'obligation de réaliser un bilan Carbone tous les trois ans (Article L 229-25 du code l'environnement). Au-delà de l'aspect réglementaire, la réalisation d'un bilan carbone permet à l'entreprise qui s'y soumet d'anticiper des évolutions réglementaires en matière de GES, de diminuer sa consommation énergétique et donc de réaliser des économies financières, d'impliquer ses salariées dans une démarche environnementale et de communiquer son engagement à l'externe.

→ **A noter que la loi devrait très vraisemblablement s'élargir aux entreprises de toute taille.**

Analyse de cycle de vie du numérique

Utilisée depuis les années 1990, l'analyse du cycle de vie permet évaluer l'impact d'un produit ou d'un service sur l'environnement. Cette méthode, qui a fait l'objet de 2 normes ISO (14040:2006 : Principes et cadre et 14044 :2006 : Exigences et lignes directrices), repose sur la prise en compte de plusieurs indicateurs environnementaux (**approche multicritère**) et la prise en compte des impacts générés sur toute la vie du produit ou du service (**cycle de vie**).

Pour chiffrer les impacts d'un produit ou d'un service numérique, les indicateurs environnementaux retenus vont généralement caractériser (24) : la régression de la biodiversité, l'épuisement des ressources abiotiques, la consommation d'énergie primaire, la consommation d'eau et d'énergie finale, la production de déchets électroniques, de papier.

Les quatre étapes du cycle de vie d'un produit concernent la **fabrication** (extraction des matières premières, transports amont, fabrication), le **transport** (du site de fabrication au site d'utilisation), l'**utilisation** (consommation électrique) et la **fin de vie** (collecte, recyclage, traitement des déchets). Dans le domaine du numérique, 3 niveaux d'équipements sont concernés : les **terminaux**, les **infrastructures de réseaux** et les **centres de données**. La norme ISO 14040 :2006 définit 4 phases indépendantes pour la réalisation d'une AVC (Définition des objectifs et de la portée, Analyse de l'inventaire, Évaluation de l'impact, Interprétations). L'analyse du cycle de vie est une méthode très aboutie en matière d'écoconception. Elle permet aux entreprises qui entreprennent ce type de démarche de faire des choix éclairés en ayant une meilleure connaissance de leurs impacts environnementaux et d'envisager ainsi des axes d'amélioration. Cette démarche est toutefois exigeante pour l'entreprise, demande d'y consacrer du temps et disposer de l'ensemble des données nécessaires. Ce qui n'est pas toujours possible.

Guides

Parmi les guides et référentiels permettant d'accompagner les bonnes pratiques, on peut citer le **guide de l'ADEME** (33) plutôt orienté grand public, le **référentiel du CIGREF** (34) qui regroupe 100 bonnes pratiques de sobriété numérique à destination des organisations. Ce référentiel comporte des points d'attention sur le data management et les solutions d'IA. Enfin dernièrement l'**AFNOR** a publié en avril 2022 un guide qui fournit aux organisations des bonnes pratiques pour réduire l'impact environnemental de leurs services numériques (35). Il invite notamment en matière d'intelligence artificielle à se poser la question du bien-fondé du service envisagé au regard des enjeux environnementaux.

2. L'univers du numérique

En 2019, le rapport du GreenIT (25) précise que 34 milliards d'équipements pour 4,1 milliards d'utilisateurs (soit 8 équipements par utilisateur) composent cet univers numérique. Entre 2010 et 2018, le poids des données est passé de 0,5 à 2 milliards de téraoctets par an, soit 4 fois plus et le nombre de serveurs a quasiment doublé, passant de 35 millions à 61 millions (24). La règle des trois tiers (36) régit généralement le numérique avec les data centers, les réseaux et les terminaux à destination des utilisateurs. Dans cette partie, un focus un peu plus important a été réalisé sur les data centers. Bien il y ait eu de nombreux progrès en termes d'efficacité énergétique, les data centers cristallisent souvent de nombreuses critiques. Ils font cependant partie intégrante des choix à réaliser par les praticiens de l'IA, il convient donc d'avoir une vision un peu plus complète des paramètres qui les caractérisent.

a. Les data centers

Véritables infrastructures pour conserver, organiser et traiter des données, les data centers accueillent un réseau, des espaces de stockage et des serveurs de calcul. Dans la mesure où les data centers génèrent beaucoup de chaleur, ils nécessitent le plus souvent des systèmes de refroidissement pour fonctionner de manière optimale (à des températures entre 24 et 27°C). Plusieurs techniques de refroidissement peuvent être mises en œuvre ; à l'air (free cooling), à l'eau (free-chilling), adiabatique/évaporatif, refroidissement liquide.

En France, un data center consomme en moyenne 5,15 MWh/m²/an, soit pour une surface de 10000 m² autant qu'une ville de 50000 habitants (37). Aujourd'hui, les data centers représentent 17% de l'empreinte carbone du secteur du numérique mondial. Suivant les études, ceux-ci pourraient réclamer 10% de la production mondiale d'électricité d'ici 2030.

Situation dans le monde

La Figure 2 (37) propose une classification des data centers suivant leur taille et leur implantation associée sur les territoires.

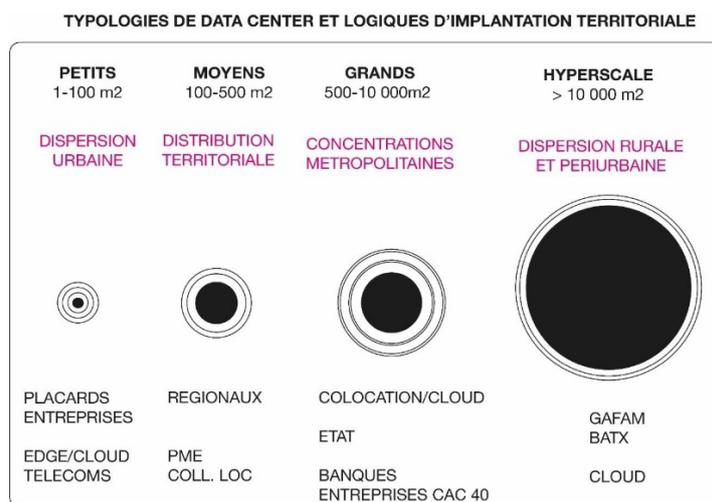


Figure 2 - Typologies de Data Center et logiques d'implantation territoriale (37)

On peut retrouver sur le Tableau 2 une estimation de 2017 (37) du nombre de sites dans le monde en fonction de leur taille et du nombre de serveur hébergés.

Taille	Nombre de serveurs	Nombre de sites
Petit / Moyen	Salle / Armoire	2,8 millions
Moyen / Grand	50 à 100	85 000
Grand / Hyperscale	1000 à 1 millions	8000

Tableau 2 – Estimation par taille et par nombre de serveurs du nombre de sites dans le monde en 2017

Selon le site internet www.cloudscene.com, on dénombre aujourd'hui environ 8400 de data centers (grands et hyperscale) dans le monde. Rien que les Etats-Unis en totalisent 2751, arrivent ensuite loin derrière l'Allemagne (484), l'Angleterre (458), la Chine (447). La France en compte 263 (38). A la fin de l'année 2021, on comptait 700 data centers hyperscale dans le monde. Une progression qui a plus que doublé depuis 2015, puisque l'on n'en dénombrait alors « que » 259. Les dix plus gros datacenters hyperscale du monde combinant des milliers de serveurs, se trouvent en Chine et aux États-Unis dans des bâtiments dont les surfaces varient de 200000 à 1 millions de m².

Et Synergy (39) prévoit une croissance annuelle de l'ordre de 20 à 30% du nombre des data centers hyperscale avec un dépassement de la barre des 1000 en 2025. Les États-Unis représentent 40% des installations et la moitié de la capacité mondiale, viennent ensuite l'Asie-Pacifique (28%) et l'Europe en troisième place (19%). Si en termes de capacité des data centers les GAFAM sont les leaders, les hyperscalers chinois (Tencent, Alibaba ByteDance) connaissent une croissance plus rapide.

Les enjeux des data centers

Pour assurer leur fonctionnement et la qualité de service 365 jours par an, 24h/24h (continuité numérique), les data centers posent des questions en termes d'énergie consommée et de ressources mobilisées. Les enjeux des data center concernent :

- Le cycle de vie des serveurs et l'utilisation de métaux rares,
- Le refroidissement des installations qui peut représenter jusqu'à 50% de la consommation en électricité d'un data center,
- La redondance des réseaux de distribution d'énergie pour assurer l'alimentation électrique de secours en cas de problème. Ce sont le plus souvent des générateurs de secours fonctionnant au fuel.
- Le surdimensionnement des infrastructures pour pallier des pics de trafic. Jusqu'à 40% des serveurs peuvent ainsi être alimentés en électricité mais ne sont pas actifs.
- La duplication des data centers eux-mêmes pour assurer la sécurisation des données en cas de défaillance d'un site.

Plusieurs indicateurs peuvent être mis en place pour analyser la performance des data centers (40), on distingue généralement les indicateurs de consommation d'énergie et les indicateurs environnementaux. Ces indicateurs sont détaillés en Annexe 8.



“ Suivant les clients et leurs priorités, les critères de choix d'un data center reposent sur la sécurité, le coût et l'aspect environnemental. Nos data centers sont certifiées ISO50001. Nous avons développé depuis une dizaine d'années des techniques d'optimisation énergétique allant de la distribution de la chaleur fatale pour chauffer des bâtiments voisins à l'utilisation de groupes froids pour générer et stocker des frigorifiques pendant les heures creuses. ”

Serge Cousin - Président/Fondateur - CIV France

b. Les réseaux

Pour les entreprises et le grand public, le transport des données s'effectue par l'intermédiaire d'infrastructures de réseaux numériques. Il existe 2 types de réseaux : fixes (ADSL, fibre, ...) et mobile (3G, 4G, 5G,...). Pour faire circuler les données, les équipements d'infrastructures (routeurs, antennes-relai, ...) fonctionnent en continu et sont responsables de 40% de la consommation électrique mondiale du numérique. Comparés aux réseaux fixes, les réseaux mobiles sont beaucoup plus gourmands en énergie puisqu'ils consomment le double. En augmentation de 6% chaque année depuis 2016, la consommation énergétique des réseaux fixes et mobiles en France a atteint 3800 GWh en 2020 (41). Elle pourrait même doubler d'ici 2035. Dans la mesure où l'usage de la fibre entraîne quatre fois moins de consommation qu'avec l'ADSL, le remplacement de l'ensemble des câbles en cuivre pourrait apporter des gains significatifs à condition bien sûr que les usages du réseau n'augmentent pas. Ce qui ne semble pas être la tendance.

c. Les terminaux

Les terminaux et les objets connectés sont les premiers responsables des impacts environnementaux du numérique. Ils comprennent les ordinateurs portables, les tablettes, les

smartphones, les télévisions et tous les objets qui intègrent des capteurs capables de se connecter à d'autres systèmes pour échanger des données. La fabrication de ces terminaux représente 80% de l'empreinte carbone du numérique et nécessitent beaucoup de ressources abiotiques. Pas moins de 850 kg de matières premières sont nécessaires pour la fabrication d'un PC portable de 2,5kg.

Sac à dos écologique

Développé en Allemagne dans les années 1990, le concept économique MIPS (Material Input Per Unit of Service) considère le poids total des matières premières rapporté au poids total du produit fini (ou sac à dos écologique). On peut comparer sur le Tableau 3(24) (30) le MIPS d'un ordinateur portable avec celui d'une barrette mémoire de 2g. Ce dernier atteint la valeur impressionnante de 16000 !

Équipement	Matières premières (kg)	Produit fini (kg)	MIPS
Ordinateur portable	850	2,5	350
Smartphone	183	0,150	1220
Barrette mémoire	32	0,002	16000

Tableau 3 – Comparaison des MIPS par équipement (24) (30)

Production de déchets

75 millions de tonnes de DEEE (Déchets d'équipements électriques et électronique) sont produites chaque année dans le monde. La moyenne européenne est de 16,6kg par habitant et par an alors qu'en France ce chiffre atteint 21,5kg. Seul 30% sont collectés à l'échelle de la planète. En France 52% ont été collectés en 2017, soit 10kg sur les 21,5kg de déchets. (24) En Europe, les techniques de recyclage progressent pour les métaux précieux tels que l'or, l'argent et le cuivre mais reste encore inefficaces pour les autres métaux présents en plus petites quantités mais beaucoup plus polluants comme le tantale.

3. L'effet rebond

Introduit en 1865 par l'économiste britannique William Stanley Jevons, l'effet rebond aussi appelé « paradoxe Jevons » caractérise le fait que plus les progrès technologiques améliorent l'efficacité énergétique d'une ressource, plus celle-ci est utilisée, entraînant ainsi une augmentation de la consommation totale de cette ressource. Le Tableau 4 illustre ce phénomène. L'amélioration de la consommation moyenne d'un véhicule n'a pas permis de diminuer les émissions de CO₂ puisque dans le même temps le nombre de voitures vendues a augmenté (effet rebond « directs »). L'augmentation du pouvoir d'achat résultant des baisses de consommation de carburant peut également entraîner une hausse des émissions de CO₂ sur d'autres postes (l'aviation par exemple). C'est l'effet rebond « indirect ».

Automobile entre 2005 et 2018	Consommation moyenne d'un véhicule à essence 8,8 l/100km → 7,2 l/100kms Gain 22%	Vente annuelle de véhicules neufs dans le monde 66 millions → 95 millions Hausse de 44%
Aviation entre 2013 et 2019	Réduction des émissions de CO ₂ par passager par kilomètre : 12%	Hausse des émissions carbone du secteur aérien civil : 29%
Internet mobile entre 2021 et 2025	Consommation moyenne de data en 4G par utilisateur 11,6 gigaoctets	Chaque mois, 20% des internautes consommeront 200 gigaoctets d'Internet mobile grâce à la 5G, soit 10 à 20 fois plus que lorsqu'ils utilisaient la 4G

Tableau 4 – Illustration de l'effet rebond par secteur

Bien que les performances des matériels et des applications informatiques augmentent à un rythme bien plus élevé que dans d'autres secteurs, il est illusoire de penser que la réduction des impacts du numérique se fera sans intégrer davantage de frugalité dans les usages.



“ Dans les usages, les applications ne consomment pas directement de ressources, mais elles mettent en œuvre des composants d'infrastructures, qui eux en consomment. Tout dépend de la valeur apportée par les usages. Si cela génère de la valeur alors cela justifie l'utilisation de ressources. ”

Louis Naugès - Chief Strategy Officer - Wizy.io

4. Les labels du numérique

Fin 2021, l'Alliance Green IT a publié un livre blanc (42) pour aider les organisations à valoriser leurs engagements pour un numérique plus durable et plus responsable. Ce guide propose 7 labels (label **Numérique responsable**, label **engagé RSE**, **Certification B-Corp**, label **Lucie 26000**, **Positive Workplace**, Label **Rexcelys**, label **Planet RSE**) ainsi que des éléments d'aide à la décision. L'intérêt de ces labels est généralement d'initier ou de concrétiser la démarche RSE d'une entreprise. En cela, ils peuvent constituer de bons leviers pour faire évoluer les pratiques. Le label Numérique Responsable est un label sectoriel qu'il est judicieux de retenir pour les métiers du numérique (dont ceux de l'IA). Il bénéficie du retour d'expériences de plusieurs années du label Lucie 26000 qui est un label associé aux entreprises engagées sur la RSE. Plus généraliste, le label Lucie est porté par le référentiel de la norme ISO 26000 (1^{er} standard international de management de la RSE publié en 2010). Cette norme, non certifiante, propose néanmoins un guide pratique pour mettre en place une démarche de développement durable dans une logique d'amélioration continue.

Partie 3 - L'Intelligence Artificielle et sa soutenabilité

"Notre avenir est une course entre la puissance croissante de notre technologie et la sagesse avec laquelle nous l'utiliserons"
Stephen Hawking

1. Qu'est-ce que l'IA ?

Qualifiée de nouvelle électricité du 21^{ème} siècle par Andrew Ng, professeur expert à Stanford, l'intelligence artificielle (IA) est annoncée comme la 4^{ème} révolution industrielle. Elle représente un tournant incontournable et riche d'opportunités pour les entreprises.

a. Définition

L'intelligence artificielle relève de technologies informatiques permettant à des machines d'apprendre et de réaliser des tâches traditionnellement dévolues à l'intelligence humaine. Cette discipline scientifique est « jeune » de soixante-six ans car c'est en 1956 qu'elle a été officialisée lors de la conférence organisée par John McCarthy (Dartmouth) et Marvin Minsky (Princeton) au Dartmouth College – États-Unis. Le mathématicien et cryptologue britannique, Alan Turing, également considéré comme l'un des fondateurs de l'informatique moderne, explorait déjà en 1950 la définition de la conscience d'une machine (test de Turing). En 1974, en raison d'un manque d'avancées concrètes arrive le premier « hivers de l'IA » avec la réduction des financements des gouvernements britannique et américain. Les années quatre-vingt permettent de relancer les projets de recherche avec le succès des systèmes experts permettant à un ordinateur de se comporter comme un humain dans un domaine bien précis. Difficiles et coûteux à maintenir, les systèmes experts entraînent un deuxième hiver de l'IA entre 1987 et 1993 au profit de matériels informatiques plus récents et moins chers mais sans IA. L'émergence du Machine Learning (apprentissage automatique), un sous domaine de l'intelligence artificielle, a lieu dans les années quatre-vingt-dix et entraîne le développement d'algorithmes qui permettent de réaliser des prédictions à partir de données tout en s'améliorant au fil du temps.

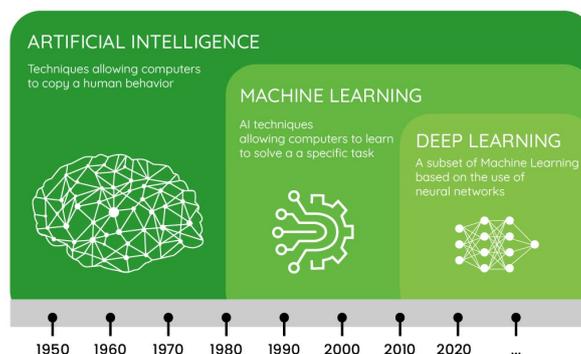


Figure 3 - IA et Sciences des données (Source dilepax)

Portée par le big data, l'explosion de la puissance de calcul (en lien avec la loi de Moore) et le deep learning (apprentissage profond), l'intelligence artificielle connaît depuis le début des années 2000 un intérêt croissant qui ne faiblit pas. Toutes les applications d'intelligence d'artificielle relèvent à ce jour de ce que l'on appelle l'IA faible (ou IA étroite) dédiée à la résolution de tâches spécifiques comme la reconnaissance d'images ou le traitement du langage naturel. Une IA forte (ou IA générale), dotée d'esprit, de sensibilité, de conscience relève encore de la science-fiction.

b. IA connexionniste et IA symbolique

Les rivalités entre ces deux courants de l'IA existent depuis le début de l'histoire de l'intelligence artificielle. Basée sur des règles et des connaissances, l'IA symbolique permet l'explicabilité des algorithmes ce qui est beaucoup moins évident pour l'IA connexionniste basée sur des réseaux de neurones artificiels. Alors que les modèles de deep learning nécessitent de grand volume de données pour pouvoir effectuer des prédictions, l'IA symbolique est très frugale en matière de données. En réalité, chaque approche présente des atouts et des limites et l'avenir semble s'orienter vers l'IA neuro-symbolique combinant ainsi le meilleur des deux mondes.

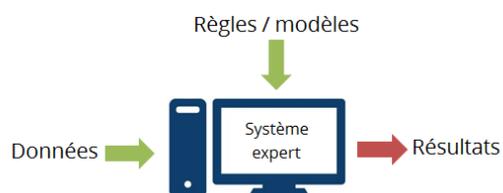


Figure 4 - Principe de fonctionnement de l'IA symbolique (Source : Conseil de l'Europe portail)

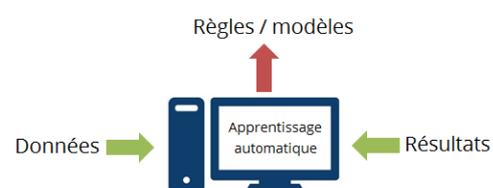


Figure 5 - Principe de fonctionnement de l'IA connexionniste (Source : Conseil de l'Europe portail)

IA symbolique

Prédominante au début des années 1990, l'approche symbolique de l'IA (encore appelée GOFAI – Good Old Fashion AI) a contribué à la mise en place de nombreux systèmes experts (basés sur des moteurs d'inférence). Comme indiqué sur la Figure 4, l'IA symbolique repose sur des règles ou des modèles définis explicitement au préalable et intégrés ensuite dans la machine. A partir des données, la machine agit en fonction des instructions reçues pour produire un résultat.

Parmi les avantages de l'IA symbolique, on retrouve : des jeux de données d'entraînement pouvant être modestes, la possibilité de généraliser le système à d'autres domaines à partir des règles définies, des résultats ou des prédictions aisément interprétables (explicabilité).

Les limitations de l'IA symbolique concernent la reconnaissance d'images et du langage naturel.



“ 90% des communications autour de l'IA relèvent du machine learning et du deep learning et laissent peu de place à d'autres solutions innovantes. Mais justement cette approche de l'IA souveraine et frugale remet à l'honneur des innovations. L'apprentissage et la configuration de notre solution sont quasiment instantanés. Chez Golem.ai, nous avons également l'ambition de fournir à nos clients leur consommation en équivalent CO₂ ”

Jean-Denis Garo - Head of Marketing – Golem.ai

IA connexionniste

Contrairement à l'approche symbolique qui nécessite de programmer a priori des règles logiques de traitement de données pour imiter un raisonnement, l'IA connexionniste s'appuie sur la fourniture de volumes importants de données d'entrée et de résultats attendus pour définir de manière plus ou moins automatisé les règles pouvant les lier (cf. Figure 5).

La précision des modèles dépend du fait de disposer de grandes quantités de données et d'avoir recours une puissance de calcul importante.

IA neuro-symbolique

L'IA neuro-symbolique (ou IA hybride) repose sur l'idée d'intégrer des règles dans un réseau de neurones (43). Cela constitue le défi d'aujourd'hui pour combiner les propriétés du deep learning (apprentissage, reconnaissance de formes) et de l'IA symbolique (règles logiques, transparence, frugalité des données). Cela pourrait conduire à des avancées significatives dans les systèmes d'IA traitant des problèmes complexes en induisant du raisonnement dans l'apprentissage tout en nécessitant beaucoup moins de données pour la formation.



“ La combinaison des approches constitue une voie prometteuse pour consommer moins de données, moins de CPU, de GPU et donc d'énergie. Cette articulation permet d'apporter de l'explicabilité au modèle et combinée à des outils de visualisation, elle permet de déterminer si l'algorithme est pertinent. L'IA symbolique ne consomme rien. Mais il y a aussi un enjeu de formation pour les praticiens de l'IA moins sensibilisés à cette approche. ”

Lauren Cervoni - Directeur de la Recherche et de l'Innovation – Talan

c. Le machine learning

Sous domaine de l'intelligence artificielle, le Machine Learning (Apprentissage Automatique) comprend de nombreux algorithmes pour créer à partir des données des modèles dont les performances s'améliorent au fil de leur entraînement. Le Machine Learning est divisé en trois grandes catégories : **l'apprentissage supervisé**, **l'apprentissage non supervisé**, **l'apprentissage par renforcement**. Toutes les formes de Machine Learning s'appuient sur des statistiques et des probabilités qu'un événement ou une action ne produise ou non.

Apprentissage supervisé

Dans le cas de l'apprentissage supervisé, la formation du modèle nécessite de disposer de données annotées ou étiquetées. Le modèle peut donc apprendre sur le jeu de données d'entraînement puisque les résultats attendus sont connus. Il est ensuite capable de réaliser des prédictions sur un nouveau jeu de données (cette fois-ci inconnues). On distingue 2 types de problèmes à traiter :

- **Classification** : la variable de sortie est une classe (ou catégorie) prédite
- **Régression** : la variable de sortie est une valeur spécifique.

Parmi les algorithmes d'apprentissage supervisé, on retrouve la régression linéaire (régression), la régression logistique (classification), les arbres de classification et de régression, les forêts aléatoires, les K plus proches voisins (KNN K-Nearest Neighbors), les classifieurs de type Naïves Bayes, les machines à vecteurs de supports (SVM) et le boosting de gradient.

Apprentissage non supervisé

A la différence de l'apprentissage supervisé, aucune étiquette n'est fournie à l'algorithme pour l'apprentissage non supervisé. Les données sont alors regroupées de manière automatique en fonction de leur similitude. Il existe 2 types de techniques :

- **Clustering** : rassemblement sous forme de groupes (cluster) des objets similaires présents dans les données ou ayant des liens entre eux (Association)
- **Réduction dimensionnelle** : réduction du nombre de variables analysées pour simplifier le modèle.

Pour ce type d'apprentissage, les principaux algorithmes sont : la méthode des K-moyennes (K-Means), le clustering hiérarchique, l'algorithme Apriori, l'analyse en composantes principales (PCA Principal Component Analysis), la décomposition en valeurs singulières (SVD Single Value Decomposition)

Apprentissage par renforcement

De plus en plus utilisé, l'apprentissage par renforcement (Reinforcement Learning) est une méthode qui consiste pour la machine à apprendre de ses expériences à partir d'un système de récompenses ou de punitions. Étant donné qu'elle est programmée pour maximiser sa récompense, la machine essaiera de trouver, par un mécanisme d'essais-erreurs les tactiques les plus efficaces pour atteindre son but. On trouve 2 types d'apprentissage par renforcement :

- **Monte Carlo** : les récompenses sont obtenues à l'état final
- **Différence temporelle** : les récompenses sont évaluées et accordées à chaque étape

Les algorithmes utilisés sont : Q-learning, Deep Q Network (DQN) et SARSA (State-Action-Reward-State-Action)

d. Le Deep Learning

Au début le perceptron

Inspiré du neurone biologique, le perceptron est un algorithme d'apprentissage supervisé parmi les plus simples pour la classification binaire (0 ou 1). Les données d'entrée (x_1, x_2, \dots, x_n) sont multipliées par les poids (w_1, w_2, \dots, w_n) et la somme est ensuite soumise à une fonction d'activation dont le seuil définit l'activation ou non du neurone. Le résultat prédit est alors comparé au résultat connu et cas d'erreur, les poids sont ajustés en conséquence.

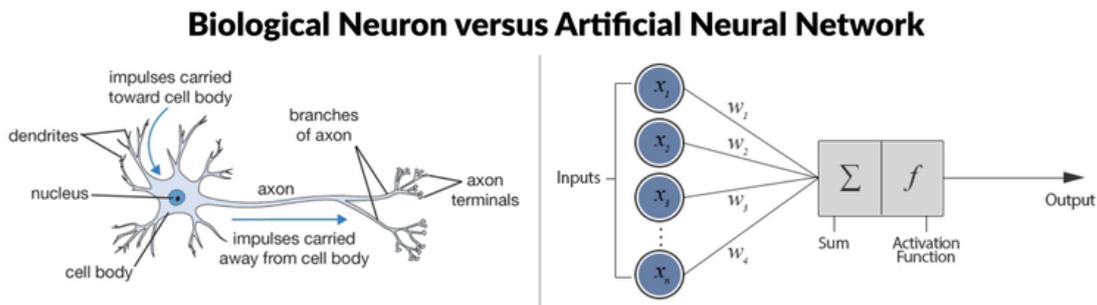


Figure 6 - Représentation d'un neurone biologique et d'un neurone artificiel (Source : Brian Mwandau)

Le tout premier modèle mathématique d'un neurone artificiel fut proposé par les américains Warren S. McCulloch et Walter H. Pitts en 1943 mais c'est le psychologue américain Frank Rosenblatt qui a mis au point le perceptron en 1957.

Les réseaux de neurones

On retrouve des réseaux de neurones perceptron simple couche qui comprennent une couche d'entrée (Input layer) qui capte les données d'entrée et les transmet après pondération et traitement à la seconde couche cachée (Hidden layer) qui en fait de même vers la couche de sortie (Output layer) pour formuler une réponse de sortie. Il s'agit d'un réseau à propagation directe ou avant (feedforward). La limitation du perceptron simple couche vient du fait qu'il ne peut que séparer des groupes séparables linéairement. L'ajout de couches intermédiaires pour former un perceptron multicouche permet de surmonter ce problème.

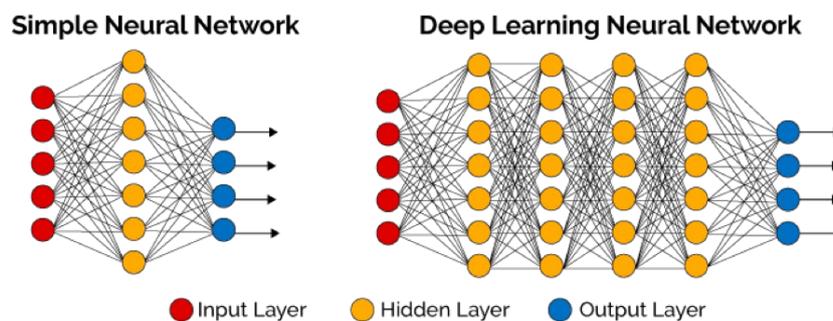


Figure 7 - Réseaux de neurones simple couche et multicouche (Source : Sufyan T. Faraj Al-Janabi)

Cependant, pour apprendre le réseau de neurones doit minimiser ses erreurs et modifier ainsi les poids synaptiques qui entraînent les erreurs les plus importantes. On parle alors de rétropropagation (backpropagation) puisque que les corrections d'erreurs se réalisent de la dernière couche à la première couche du réseau. L'optimisation du modèle s'appuie sur l'utilisation de l'algorithme de la descente de gradient (Gradient descent) qui permet de trouver le minimum d'une fonction de coût associée à l'évaluation de l'erreur. Il s'agit d'un processus itératif, plus la fonction de coût est faible, meilleur est le modèle. Il convient alors de bien choisir le pas d'itération ou le taux d'apprentissage (Learning rate) pour que l'algorithme puisse converger. Pour ce type de modèle, le nombre de paramètres devient très vite très important : nombre de couches, nombre de neurones associés, nombre de poids synaptiques et autres paramètres associés à chaque neurone, ... De plus, comme l'entraînement des réseaux de neurones nécessite de nombreuses données, l'apprentissage d'un modèle de Deep Learning peut nécessiter plusieurs jours voire plusieurs semaines sur du matériel dédié spécifiquement.

De modèles de plus en plus complexes

- **Réseaux de neurones convolutionnels – CNN (Convolutional Neural Networks)**

Les réseaux de neurones convolutionnels sont très efficaces pour classer des images et identifier des objets. Ils comportent généralement des couches de « **convolution** » et des couches de « **pooling** ». La couche de convolution permet par l'application de filtres (ou noyaux de convolution) d'extraire les caractéristiques de l'image. Les filtres de la couche de pooling réduisent la résolution de l'image. Après ces opérations, l'ensemble des informations sont ensuite concaténées sous forme d'un vecteur (Flattening). Enfin la couche « **Fully connected** » correspond à un perceptron multicouche dont le rôle est de classer l'image.

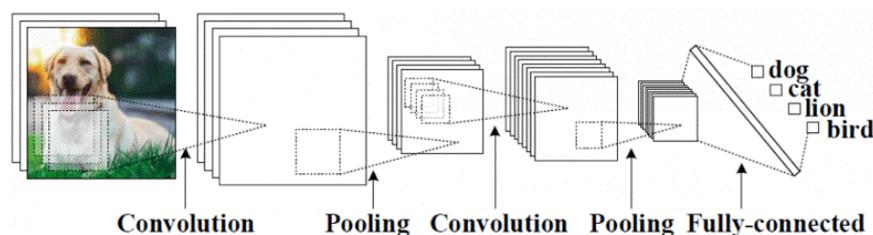


Figure 8 - Architecture d'un CNN (Source : juripredis.com)

Plusieurs centaines de milliers d'images annotées sont nécessaires pour entraîner ce type de réseau, ce qui est assez coûteux en termes de matériel, d'expertise et de temps de calcul (plusieurs semaines).

- **Réseaux de neurones récurrents – RNN (Recurrent Neural Networks)**

Les réseaux de neurones récurrents sont surtout utilisés pour le traitement du langage naturel et la reconnaissance vocale. Ce type réseau utilise les résultats précédents comme entrées

supplémentaires et peuvent traiter des données séquentielles. Ils disposent ainsi d'une mémoire à court terme qui leur permet de retenir les mots précédents dans une phrase mais il ne leur est pas possible d'accéder à des informations d'un passé lointain. Le temps de calcul est long et ils peuvent être difficiles à entraîner en raison de la disparition du gradient (Gradient Vanishing).

- **Réseaux LSTM, GRU et Transformer**

En Annexe 9, sont présentés les réseaux LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit) et les Transformer qui sont des modèles à séquences plus complexes que les réseaux RNN. Ils ont permis d'obtenir de très bons résultats dans la reconnaissance de texte, de la parole, la traduction linguistique mais également la génération de texte ou de musique. Si la précision s'améliore, l'entraînement de ces modèles est souvent difficile pour des raisons de capacité mémoire liée au très grand nombre de paramètres et d'opérations nécessaires. Il devient alors nécessaire de chercher de voies d'optimisation des algorithmes et des matériels.

e. L'IA distribuée

Parmi les différentes approches de l'intelligence artificielle, l'IA distribuée s'appuie sur la coopération de nombreux agents ou nœuds d'apprentissage autonomes qui agissent collectivement mais de manière décentralisée (systèmes multi-agents). L'idée est ici de prendre en compte une intelligence collective plutôt qu'individuelle. On parle également d'intelligence en essaim comme dans les sociétés d'insectes (les fourmis par exemple) où chaque insecte (agent) ne fait pas appel à un mécanisme de raisonnement sophistiqué mais ce sont les interactions entre les individus qui permettent de produire des comportements à plus haut niveau d'intelligence. L'intelligence artificielle distribuée peut rassembler plusieurs réseaux de neurones interconnectés et permet ainsi de réaliser des tâches complexes (gestion de véhicules autonomes ou semi-autonomes par exemple) pouvant nécessiter de très gros volumes de données. C'est aussi une voie utilisée pour résoudre les problèmes d'entraînement de modèles, de planification, d'ordonnancement de flux de production ou de prise de décision. La sécurité des données ainsi que l'intégrité des agents constituent encore des freins à l'utilisation de cette technologie.

2. L'écosystème Big Data et IA

Les développements fulgurants de l'intelligence artificielle ont été rendus possibles par le Big Data (ou la capacité de stocker et de gérer de très gros volumes de données), la montée exponentielle de puissance de calcul et le développement des algorithmes de machine learning.

a. Le Big data

Le Big Data a été rendu possible par l'« alignement de 4 planètes » (44) : l'explosion du volume de données, la démocratisation de l'accès à l'information, la baisse des coûts matériels, une vague d'innovation sans précédent. Le volume mondial de données a été estimé à 47 zettaoctets (mille milliards de téraoctet) en 2020 et devrait atteindre 175 zettaoctets en 2025 (cf. Annexe 10). Ce volume serait ensuite multiplié par 3,5 tous les 5 ans pour atteindre 2142 zettaoctets (45). Souvent la définition du Big Data s'appuie a minima sur ce que l'on appelle les « 3V » : **Volume**, **Variété** et **Vitesse**. Le Big Data permet de gérer les fortes augmentations des volumes de données générées par les entreprises. Il est également capable de traiter tout type de données (variété). Celles-ci sont regroupées en deux catégories : les « données structurées », généralement des bases de données relationnelles (SQL, NewSQL) regroupant des chiffres et des nombres et les « données non structurées » qui relèvent du texte et du multimédia. Enfin la vitesse qui permet au Big Data, via des processus de distribution du stockage et du calcul (regroupement de machines en grappes, parallélisation des opérations), de traiter rapidement des données massives et d'accomplir des calculs complexes au travers des algorithmes d'IA. D'autres mots commençant par la lettre V sont venus ensuite caractériser le Big Data : Variabilité (sens des données), Véracité (qualité des informations), Visualisation (représentation synthétique), Valeur (pour quel usage ?). Compte tenu de l'explosion du volume de données générées, c'est certainement sur la valeur d'usage que les praticiens de l'IA doivent de plus en plus se pencher.

b. Les processeurs

Il existe différents types de processeurs (CPU, GPU, TPU, FPGA, NPU, ...), leur conception dépend des tâches à réaliser et certains processeurs ont été spécialisés afin améliorer les temps de traitement des applications d'IA.

CPU (Central Processing Unit)

Souvent appelé le « cerveau » de l'ordinateur, le CPU (Central Processing Unit) est composé de millions de transistors gravés lors de l'usinage et peut comporter plusieurs cœurs de traitement. Il est chargé de l'exécution des instructions des programmes. Celles-ci sont généralement stockées dans la mémoire et envoyées au processeur dans un langage binaire n'utilisant que des zéros et des uns. Le premier CPU développé par Intel en 1971 avec une finesse de gravure de 10 micromètres et une fréquence de 750 kHz était le 4004. Il comportait 2300 transistors et effectuait 60000 opérations par seconde.

Les processeurs actuels utilisent plusieurs cœurs pour traiter en parallèle plusieurs instructions. En cinquante ans les progrès ont été fulgurants puisque la douzième génération des processeurs Intel contient plusieurs milliards de transistors pour une gravure à 10

nanomètres, propose jusqu'à 16 cœurs pour des fréquences pouvant atteindre les 5,2 Ghz et peut traiter plus de 700 téraflops (soit 700 mille milliards d'opérations par seconde) ! Les grands constructeurs de processeurs que sont Intel et AMD travaillent à réduire la consommation électrique des CPU en optimisant leur architecture interne. En mai 2021, IBM a annoncé avoir mis au point un processeur de la taille d'un ongle gravé à 2 nanomètres et qui comporterait 50 milliards de transistors. Les performances seraient plus élevées que pour une puce gravée à 7 nanomètres pour une consommation d'énergie inférieure de 75%. Bien que les limites des lois physiques viennent questionner la loi de Moore établie en 1965 qui prévoyait le doublement des transistors présents sur un microprocesseur tous les 18 à 24 mois et qui s'est avérée exacte jusqu'à aujourd'hui, Intel a dévoilé en juillet 2021 une feuille de route ambitieuse en visant des procédés de gravure de 1,8 nanomètres en 2025 (soit 18 angström).

GPU (Graphics Processing Unit)

En 2006, l'université de Stanford découvre que les GPU (Graphics Processing Unit), des accélérateurs graphiques avec un très grand nombre de cœurs mais plus simples que les cœurs des processeurs classiques, offrent de meilleures performances par watt. Alors qu'un CPU est limité à quelques dizaines de cœurs, le GPU qui peut en comporter plusieurs dizaines de milliers s'avère donc adapté au calcul intensif grâce à la parallélisation massive de tâches. Agissant comme des microprocesseurs spécialisés, les GPU ont permis le développement du deep learning. Le marché des GPU est dominé par l'américain Nvidia qui a notamment annoncé la sortie en mars 2022 du GPU H100 destiné à remplacer les précédentes versions (A100 et V100) dans les infrastructures de serveurs. Il ne comporte pas moins de 16896 cœurs (6912 pour le A100) et affiche une puissance de calcul de 60 téraflops (en virgule flottante sur 32 bits). D'après Nvidia, 20 GPU H100 permettraient de gérer la totalité du trafic internet. La consommation d'énergie n'est cependant pas en reste car l'enveloppe thermique ou le TDP (Thermal Design Power) du H100, qui caractérise la dissipation d'énergie nécessaire pour fonctionner, est évaluée à 700 watts. Le TDP d'un ordinateur portable est généralement compris entre 15 et 25 watts. Pour le Machine Learning, il est généralement préférable de mobiliser le processeur central (CPU) alors pour le Deep Learning mieux vaut solliciter le processeur graphique (GPU).

TPU (Tensor Processing Unit)

Les TPU (Tensors Processing Unit) ont été conçus entièrement par Google pour des applications spécifiques. Ils relèvent des ASIC (Application Specific Integrated Circuit) des circuits intégrés à application spécifique. Utilisés dans leurs infrastructures depuis 2015, ils n'ont été dévoilés au grand public qu'à partir de 2018. Disponibles en version cloud, les TPU sont très rapides pour effectuer des calculs vectoriels et matriciels denses afin d'accélérer l'entraînement et l'inférence des modèles d'IA. En mai 2021, Google a annoncé que la

quatrième génération de TPU (TPUv4) qui serait deux fois plus performante que la précédente. 4096 TPUv4 combinés dans un « pod » porterait la puissance de calcul à 1 exaflops (1000 téraflops). Dans un article de recherche (46), les auteurs ont comparé les performances de leur modèle NLP (GPT3 de OpenAI et Switch Transformer de Google). Le Tableau 5 montre que le nombre limité de TPU réduit l'enveloppe thermique et donc la consommation d'énergie et les émissions de CO₂.

Modèle	Nombre de paramètres (milliards)	Durée d'entraînement (jours)	Consommation d'énergie (Mwh)	Émission de CO ₂ eq (tonnes)	Processeur	Puissance (TFLOPS)
GPT-3	175	15	1287	552	10000 GPU Nvidia Tesla V100	24,6
Switch transformer	1500 (max)	27	179	59	1024 TPUv3	34,4

Tableau 5 - Performance comparée des modèles NLP GPT-3 et Switch Transformer

Cependant, comme il est rappelé dans un article sur LeMagIT.fr (47), le choix des processeurs n'est pas le seul critère et l'architecture du réseau de neurones est également un critère essentiel pour la réduction des impacts environnementaux

FPGA (Field Programmable Gate Array)

Les FPGA ou réseaux de portes programmables en français constituent une autre option d'accélération matérielle. Contrairement aux CPU, ces circuits sont composés de cellules qui peuvent être reprogrammées après fabrication. Ils peuvent donc être configurés très finement pour différents types de modèles de Machine Learning et permettent d'obtenir de faibles requêtes d'inférence en temps réel. Depuis 2017, Microsoft a intégré ces composants dans ses datacenters Azure.

NPU (Neural Processing Unit)

Les processeurs neuromorphiques ou NPU s'inspirent du fonctionnement du cerveau et pourraient devenir l'avenir en matière d'économie d'énergie (48). C'est d'ailleurs l'enjeu pour intégrer l'IA sur des dispositifs embarqués. L'approche repose sur l'intégration directe de la mémoire aux unités de calcul et à la communication des nombreux cœurs entre eux comme le font les synapses. Le 30 septembre 2021, Intel a dévoilé sa puce Loihi2 qui contient un million de neurones artificiels, soit 8 fois plus que la version précédente (Loihi1) et est aussi dix fois plus rapide. Destinées aux activités de recherche, ces puces sont disponibles en accès

sur le cloud. Cependant les algorithmes et les langages de programmation devront probablement être repensés pour profiter de tout le potentiel qu'elles représentent.



“ Notre approche vise à copier le fonctionnement du cerveau qui travaille un peu comme un système multi agents. En faisant ça, il n'y a pas de raison que notre système consomme plus que le cerveau, soit une vingtaine de watts. Nous sommes en train de simuler informatiquement notre puce et l'on commence à faire le hardware programmable, le FPGA. On vise une consommation pour nos puces de l'ordre de 0,1 à 0,5 watt. ”

Bruno Maisonnier - CEO and Founder – AnotherBrain

Puces photoniques

Encore du domaine de la recherche, les puces photoniques utilisent des photons pour transporter de l'information. Ces dernières ont l'avantage de ne pas créer d'interférence magnétique, ni de créer de la chaleur. L'intérêt est d'accéder à des performances potentiellement cinquante fois supérieures à celle des circuits électriques actuels (49). Le processeur photonique PACE lancé par la société chinoise Xizhi Technology en décembre 2021 promet même d'être cent fois plus rapide que les GPU actuels.

c. L'ordinateur quantique

Alors que les données de l'informatique classique sont codées en « bit » (des zéros ou des uns), l'ordinateur quantique utilise le « qubit » (bit quantique) qui est une combinaison linéaire où le qubit peut prendre la valeur zéro ou un mais aussi les deux ensemble. Ces propriétés permettraient de réaliser des calculs en parallèle de manière simultanée et non plus de manière séquentielle, des quantités phénoménales de données pourraient ainsi être traitées dans des temps extrêmement faibles. Cependant pour y parvenir, il est nécessaire de maintenir les qubits dans un état suffisamment stable (d'autant plus lorsqu'ils sont nombreux) car ils peuvent perdre leurs propriétés quantiques à cause d'interaction avec leur environnement. Cela relève encore du défi pour les équipes de recherche. Les technologies quantiques représentent néanmoins un avenir prometteur pour effectuer des calculs complexes irréalisables sur les machines actuelles. Les applications pourraient concerner le développement de médicaments, la modélisation financière, les prévisions météorologiques et notamment la cybersécurité. En 2021, Google a créé un laboratoire de recherche sur la conception de puces quantiques et prévoit la création d'un ordinateur quantique avec 1 million de qubits pour 2029. Les ordinateurs quantiques actuels ne disposent aujourd'hui que d'une centaine de qubits. Le nouveau processeur Eagle annoncé par IBM fin 2021 disposent de 127 qubits (65 qubits pour le précédent en 2020). La feuille de route d'IBM prévoit 433 qubits en 2023 et 1000 au-delà de 2025.

d. Le edge computing

Alternative au cloud computing, l'edge computing est une architecture informatique permettant le traitement des données à la périphérie du réseau. Les données sont alors traitées là où elles sont générées ce qui permet d'en réduire considérablement le volume avant de les transférer vers le cloud (ou un data center). Le traitement des données peut relever d'un simple filtre mais les recherches visent surtout à embarquer de l'intelligence artificielle au sein même des objets connectés (capteurs, terminaux). Et l'enjeu est de taille car d'après les prévisions (50) le nombre d'appareils IoT dans le monde devrait presque tripler en passant de 8,74 milliards en 2020 à plus de 25,4 milliards d'appareils IoT en 2030. De nombreux secteurs comme l'industrie, les transports, la logistique mais aussi la santé, la finance, la smart city, ou la réalité augmentée doivent pouvoir disposer d'une grande vitesse d'analyse de données. L'edge computing permet de résoudre les problèmes du temps de traitement et d'analyse des données (latence) grâce au traitement local, de l'énergie consommée pour transmettre les données des IoT dans le cloud, de la bande passante (quantité de données transportée par le réseau) et de la sécurité des données par le traitement à proximité de la source. Les constructeurs de matériels informatiques industriels (Schneider Electric, Adventech, Axiomtek, Nexcom) occupent le marché mais l'on retrouve aussi des acteurs comme IBM avec Lenovo, Dell avec Cisco et même Google, Amazon (AWS Greengrass), Microsoft (IoT Hub d'Azure).

3. L'émergence de l'IA pour l'environnement

L'intelligence artificielle peut constituer un véritable levier pour la transition écologique. Suivant une étude de Capgemini (51), l'usage de l'IA dans le monde aurait permis depuis 2017 d'aider les organisations à réduire leurs émissions de GES de 12,9%, à améliorer leur efficacité énergétique de 10,9%, à réduire leurs déchets de 11,7% (cf. Figure 9). Sur les 400 organisations interrogées en 2017, seulement 13% ont néanmoins mis en place des actions en faveur du climat en intégrant des solutions d'IA. Sur ces 13%, 42% ont situées en Europe, 30% en Amériques et 28% en Asie-Pacifique.

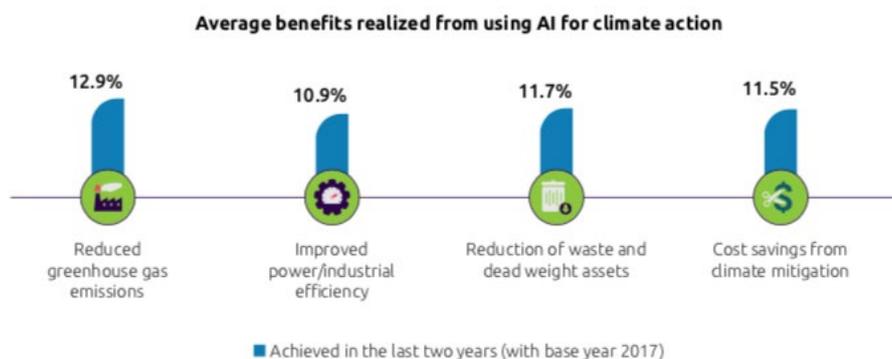


Figure 9 - Bénéfices moyens réalisés grâce à l'utilisation de l'IA pour l'action climatique (Source: Capgemini)

Il convient cependant de bien qualifier les gains environnementaux par rapport à ses coûts de mise en œuvre. Dans un de ses dossiers (52), ImpactAI aborde la définition d'un cadre d'analyse pour évaluer sur des critères objectifs les solutions d'IA vis-à-vis de l'environnement. Ce cadre (cf. Annexe 11) regroupe sept enjeux environnementaux dont les KPI (Key Performance Indicator) ont été définis par les autorités françaises, européennes ou internationales. Le secteur de l'IA (IA for Green) pour l'environnement est très dynamique et de nombreuses applications voient le jour notamment dans les secteurs de l'agriculture, de l'énergie, du transport, des villes intelligentes, de l'efficacité énergétique des bâtiments et des datacenters (53). On pourra retrouver en Annexe 12 un article sur le sujet qui a été co-écrit avec des collègues du MBA MIA et publié sur LinkedIn.



“ A la Cité de l'IA, notre 1^{er} axe est d'orienter les entreprises sur des cas d'usage qui apportent véritablement de la valeur. Le 2^{ème} axe est de dire à une entreprise de réaliser le bilan carbone de son projet avant de se lancer. L'IA peut être consommatrice mais peut aussi générer beaucoup d'impacts positifs sur l'environnement. ”

Manuel Davy - PhD, Founder and CEO – VEKIA / Leader – Cité de l'IA Hauts-de-France

4. Les impacts environnementaux de l'IA

a. Consommation énergétique

En 2019, dans un article qui a fait grand bruit, des chercheurs de l'université du Massachusetts ont évalué l'empreinte carbone des grands modèles de NLP comme les Transformers, BERT, ELMo ou GPT2. Le plus gourmand des modèles affichait une consommation énergétique de 656 MWh, une empreinte carbone de 626 200 livres de CO₂eq (284 tonnes de CO₂eq), soit cinq fois l'empreinte carbone sur la vie totale d'une voiture, carburant compris (54). En décembre 2021, la MIT Technology Review publiait un article intitulé « 2021 was the year of monster AI models » (2021 est l'année des modèles d'IA monstrueux) (55) qui revient sur la prolifération des grands modèles d'IA. C'est en juin 2020 que le modèle GPT-3 publié par OpenAI basé sur un réseau de neurones montre des capacités étonnantes sur la compréhension du langage. Il peut même générer des phrases cohérentes, converser avec des humains, voire de compléter des lignes de code à partir d'un besoin exprimé par écrit. C'est alors le plus gros réseau de neurones jamais construit. GPT-3 compte 175 milliards de paramètres, 10 fois plus que son prédécesseur GPT-2. Aujourd'hui beaucoup de grands modèles d'IA dépassent GPT-3 et cette tendance ne concerne pas que les États-Unis (cf. Annexe 13).

Une étude de 2021 de l'université de Berkeley en Californie a montré que l'entraînement du modèle GPT-3 (175 milliards de paramètres) avait produit 552 tonnes de CO₂eq pour une consommation d'énergie de 1287 MWh. Il a fallu une infrastructure dix fois plus puissante que le supercalculateur français Jean-Zay. Pas loin de 10000 cartes graphiques ont été

nécessaires et un seul GPU aurait demandé 405 ans pour effectuer l'ensemble des calculs sans compter la mémoire nécessaire 10 fois supérieure à celle d'un GPU pour stocker les 700Gb de l'ensemble des paramètres. 8 GPU avaient été utilisés pour le modèle GPT-1 basé uniquement sur 110 millions de paramètres. Pour tenter de palier au problème de la puissance de calcul, Deepmind a lancé RETRO, un modèle de langage qui ne comporte « que » 7 milliards de paramètres. En s'appuyant sur une base de données de 2000 milliards d'extraits de textes, il parvient à rivaliser pour la génération de texte avec des modèles 25 fois plus grands. On peut observer sur la Figure 10 que la quantité de calculs nécessaire pour entraîner les grands modèles d'IA (modèles SOTA : State Of The Art) augmente de manière exponentielle. Elle est multipliée par 15 tous les 2 ans pour les domaines de la vision par ordinateur (VC : Computer Vision), du traitement du langage naturel (NLP : Natural Language Processing) et de l'apprentissage de la parole (Speech). Pour les modèles Transformer, ce facteur atteint 750 tous les 2 ans ! A comparer avec le facteur 2 tous les 2 ans de la loi de Moore.

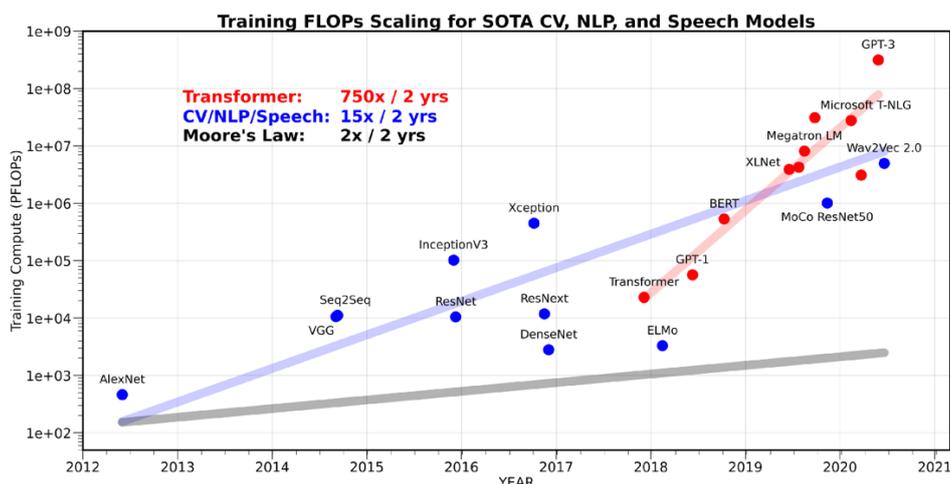


Figure 10 - Évolution de la quantité de calcul en Péta FLOPs pour différents modèles CV, NLP et Speech (Source : Amir Gholami) (56)

b. Les outils d'évaluation des émissions carbone

La mesure des émissions carbone liées aux projets d'IA est indispensable pour envisager par la suite des actions d'optimisation et de réduction. Depuis peu, les fournisseurs de cloud tels que Google, Microsoft et AWS mettent à la disposition des applications qui permettent à leurs clients de mieux évaluer l'impact de leurs activités. D'autres outils, accessibles en ligne ou en téléchargement sont développés par des chercheurs et sont accessibles gratuitement. L'ensemble de ces outils est présenté en Annexe 14. A titre d'exemple, la Figure 11(51) issue du rapport Climate AI de Capgemini Research Institute (51) donne une estimation de l'empreinte carbone de deux cas d'usage courants en IA : Contrôle qualité par reconnaissance d'images dans une usine et reconnaissance optique de caractères pour une grande société d'énergie. En se basant notamment sur les travaux de recherche de Facebook AI Research

et de l'Université Mc Gill, les auteurs ont estimé que les émissions de GES produites lors de la formation et de l'exécution d'un système d'IA variaient de quelques grammes à quelques kilogrammes de CO₂eq. Bien que ces chiffres puissent paraître faibles par rapport à l'empreinte globale d'une grande organisation, il faut se rappeler qu'un projet d'IA s'étend au-delà des phases d'entraînement et d'exécution. Toutes les étapes (collecte de données, stockage, formation, tests, exploitation, achat de matériel et de service cloud,...) ainsi que les différents cas d'utilisation doivent être pris en compte dans l'analyse d'impact.

AI use case	GHG emissions produced in...		Average organization emissions
	Build/training phase	Run/execution phase	
Image recognition system for quality control at a plant	10 kg of CO ₂ eq.	0.3 kg of CO ₂ eq.	6 million ton of CO ₂ eq. – the average annual Scope 1 emissions for a top 30 consumer products manufacturer.
AI-based optical character recognition for a large energy company	0.78 kg of CO ₂ eq.	0.96 kg of CO ₂	40 million ton of CO ₂ eq. – Scope 1 emissions for a large oil & gas major in Europe.

Figure 11 – Exemples d'empreinte GES d'applications d'IA (51)

Ces outils issus d'un domaine de recherche encore émergent permettent surtout de sensibiliser les praticiens de l'IA sur les impacts de leur choix et de leurs usages. Avant de déployer une solution d'IA à grande échelle, une analyse d'impact s'avère donc indispensable.

5. Amélioration des performances énergétiques de l'IA

Le développement des modèles de Deep Learning a augmenté significativement le nombre de paramètres, les ressources nécessaires à leur apprentissage et la latence des prédictions. Des chercheurs de Google Research ont proposé une approche permettant de rendre les modèles d'apprentissage profonds plus petits, plus rapides et plus efficaces (57). Celle-ci est structurée en cinq domaines, les quatre premiers sont axés sur la modélisation (techniques de compression, technique d'apprentissage, automatisation, architecture efficace) et le dernier sur l'infrastructure informatique et les outils. À partir de ce cadre, seront donc présentées des pistes d'améliorations des algorithmes ainsi que des infrastructures et outils. Il semble toutefois important de compléter ces deux points par l'amélioration des données et l'amélioration des pratiques.

a. Amélioration des algorithmes

Techniques de compression

Les techniques de compression s'attachent à optimiser l'architecture du modèle pour le rendre plus petit. Au final, l'idée est de pouvoir assurer son déploiement sur des dispositifs périphériques. On retrouve principalement les techniques dénommées « Quantization » et « Pruning »

Quantization (Quantification)

La quantification consiste à réduire le poids des couches en réduisant leur précision ce qui réduit considérablement les besoins en mémoire et le coût de calcul. Le processus convertit les données en virgule flottante 32 bits (FP32 – Floating Point 32 bits) à une précision plus petite comme pour un entier 8 bits (INT8 – Integer 8 bits). Un algorithme est toutefois appliqué pour s'assurer qu'il n'y ait pas trop de perte de qualité.

Pruning (Élagage)

Le pruning ou l'élagage est un processus qui consiste à supprimer les connexions entre neurones (et donc les poids associés les moins importants) dans un réseau afin d'augmenter la vitesse d'inférence et de réduire la taille de stockage. Le modèle ainsi obtenu est dit « sparse » (épars) avec une proportion des paramètres non nuls beaucoup plus faible ce qui permet de le compresser.

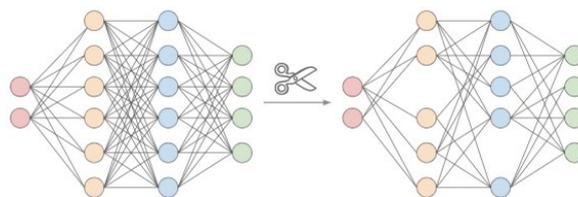


Figure 12 - Illustration de l'élagage d'un réseau de neurones (Source : Google Research)

Techniques d'apprentissage

Distillation

La distillation repose sur le principe de transférer les connaissances d'un grand modèle vers un modèle plus petit tout en maintenant sa validité. Les modèles plus petits peuvent ainsi être déployés sur des matériels moins puissants. Un petit modèle « élève » apprend donc d'un modèle « professeur » plus grand par un processus de formation différent appelé « distillation ».

Avec cette technique, DistillBERT, la version distillée de l'algorithme BERT (Google AI) créé par Hugging Face utilise 40% de paramètres en moins, son entraînement est 60% plus rapide et les performances atteignent 95% des performances de l'algorithme original.

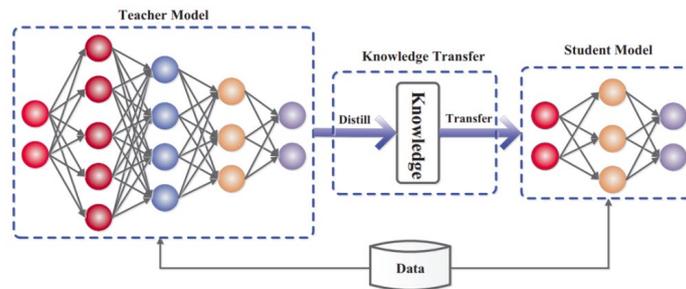


Figure 13 - Modèle Enseignant-Etudiant pour la distillation (Source : analyticsindiamag.com)

Apprentissage par transfert (Tranfert learning)

Ce concept du Transfert learning repose sur l'idée de réutiliser un modèle déjà entraîné pour un problème donné (problème source) et de capitaliser sur les connaissances déjà acquises pour traiter un autre problème (problème cible). Par exemple, un modèle formé pour reconnaître des images de voitures peut être réutilisé pour reconnaître d'autres véhicules comme des camions. Ainsi, comme le modèle est déjà entraîné, très peu de données sont nécessaires pour construire le nouveau modèle et le temps de formation est très faible. Comme indiqué sur la Figure 14, le réseau pré-entraîné est réutilisé sans sa couche finale et les paramètres des couches précédentes sont conservés. Des réglages spécifiques permettent de ne « geler » que certains poids du réseau et de faire des ajustements progressifs. Pour adapter le réseau au problème, il faut ajouter un nouveau classifieur qui correspond aux couches de neurones entièrement connectés (fully connected layer).

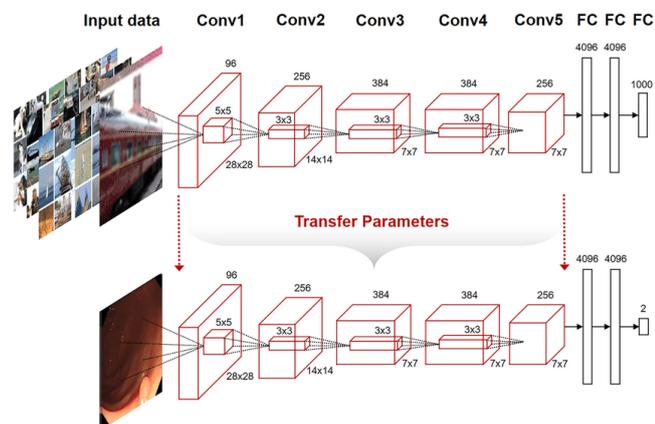


Figure 14 - Schéma de principe du transfert learning (Source : Young Jae Kim et al, 2021)

Apprentissage continu

En machine learning, l'entraînement du modèle s'effectue traditionnellement sur un ensemble de données d'apprentissage puis, après une phase de test, il est ensuite déployé pour des applications réelles. Le modèle est alors fixe et reste statique. La prise en compte de nouvelles données nécessitera donc un nouvel entraînement du modèle pour être actualisé. L'apprentissage continu ou apprentissage adaptatif permet d'adapter le modèle au flux continu

de nouvelles données sans nécessiter un nouvel entraînement complet. Il permet de s'affranchir de ressources matérielles limitées notamment dans le cas de très grands volumes de données.

Automatisation

Optimisation hyperparamétrique

La recherche manuelle des valeurs des hyperparamètres tels que le taux d'apprentissage, le nombre de couches, la décroissance des poids, etc... devient très vite fastidieuse pour optimiser les modèles. Les chercheurs de Google Research présentent des algorithmes permettant d'optimiser de manière automatique ces hyperparamètres pour former des modèles plus efficaces. Cependant, ces méthodes évoquées telles que les « grid search », « random search », « Bayesian Optimization » nécessitent de lancer de nombreux calculs pour évaluer l'influence des hyperparamètres et demandent à être utilisées avec beaucoup de prudence compte tenu des ressources informatiques nécessaires.

Architecture efficace

L'optimisation des architectures d'algorithme ayant apportées de véritables avancées en matière de performances (CNN, Transformer) font actuellement l'objet de nombreuses recherches. A titre d'exemple, on peut citer la convolution séparable en profondeur.

Convolution séparable en profondeur

MobileNet est un réseau de neurones convolutifs simple, peu gourmand en ressources informatiques pour les applications de vision mobile. Il est très utilisé par la classification d'images, la détection de visages et d'images. En 2017, Google a ainsi optimisé l'algorithme de convolution en développant une approche dite de convolution séparable en profondeur (Depth Separate Convolution) qui permet de conserver une bonne précision tout en nécessitant moins de mémoire et d'énergie. Sur le Tableau 6, on constate que le MobileNet n'a perdu que 1% de sa précision mais que le nombre de paramètres a été réduit par 7 et le nombre d'opérations d'un facteur de l'ordre de 8 à 9.

Table 4. Depthwise Separable vs Full Convolution MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
Conv MobileNet	71.7%	4866	29.3
MobileNet	70.6%	569	4.2

Tableau 6 - Comparatif d'architecture : Convolution standard vs Convolution séparable en profondeur (Source: Quantmetry)

b. Amélioration de l'infrastructure matérielle et logicielle

A l'opposé du développement de matériels informatiques toujours plus puissants le courant du TinyML (Minuscule Machine Learning) vise à s'accommoder de fortes contraintes de

ressources (mémoire, puissance de calcul) pour la réalisation d'activités de traitement sur des dispositifs périphériques (cf. Annexe 15).

c. Amélioration des données

L'approche « **model-centric** » consacrée à l'optimisation des modèles d'IA, largement développée dans les travaux de recherche fait maintenant de plus en plus place à une approche centrée sur les données. Cette approche dite « **data-centric** », promue notamment par le Professeur Andrew Ng, consiste à travailler sur l'ensemble des données afin d'améliorer les performances du modèle. Il peut s'agir de vérifier la cohérence des étiquettes, de travailler l'échantillonnage des données d'apprentissage et de rester vigilant sur la volumétrie de l'ensemble. Plutôt que de collecter davantage de données, on investit plutôt sur l'amélioration de leur qualité.

d. Amélioration des pratiques

Selon un article (46) publié par Google et OpenAI et repris dans un article du site LeMagIT (58), quatre facteurs sont à prendre en compte lors du calcul de l'empreinte de l'IA : **l'algorithme, les processeurs, le centre de données, le mix énergétique**. Le directeur général associé et responsable de l'éthique de l'IA au sein de Boston Consulting Group (BCG) GAMMA évoque aussi la nécessité d'une réflexion sur l'empreinte carbone de la chaîne d'approvisionnement de l'IA (acquisition d'algorithmes et de matériel de calcul). Le rapport du ShiftProject relatif à la sobriété numérique (59) propose d'ailleurs un cadre pour un système d'information (SI) soutenable dans les organisations.

Avec les améliorations continues des modèles d'apprentissage profond, le nombre de paramètres, la latence, les ressources nécessaires à l'apprentissage, etc. ont tous augmenté de manière significative. En conséquence, il est devenu essentiel de prêter attention aux métriques du modèle et pas seulement à sa qualité. Plusieurs articles (60) (61) (62) (63) et documents (34) (64) proposent des mesures pour réduire l'empreinte carbone des applications de Machine Learning. En synthèse, on retrouve ci-dessous les mesures les plus efficaces classées par ordre d'impact :

- **Pour les praticiens du machine learning :**
 - Se poser la question de l'intérêt de déployer du machine learning (expertise métiers)
 - Réduire les entrées/sorties de données, les calculs, les copies de données (stockages redondants)
 - Sélectionner des architectures de modèles efficaces par rapport au problème à traiter.
 - Utiliser des techniques d'amélioration d'algorithme (réduction des calculs de 3 à 10 fois)

- Entraîner le modèle sur le cloud plutôt que sur des serveurs sur site (réduction des émissions de 1,4 à 2 fois)
 - Utiliser des processeurs et des systèmes optimisés pour la formation des modèles par rapport aux processeurs généraux (amélioration des performances et de l'efficacité énergétique de 2 à 5 fois)
 - Lutter contre le gaspillage de ressources en évitant la recherche sur grille (grid search) pour l'optimisation des hyperparamètres et en réutilisant lorsque cela est possible des modèles déjà formés.
 - Quantifier et communiquer les émissions en utilisant des progiciels tels que CodeCarbon, Carbon tracker, Experiment impact tracker (inclus dans le code au moment de l'exécution) ou des outils en ligne tels que Green algorithms et ML CO2. Ces chiffres peuvent être partagés avec la communauté des praticiens pour aider à bâtir des points de repères et suivre les améliorations.
- **Pour les institutions :**
 - Privilégier les prestataires dont les solutions d'IA sont soutenables.
 - Déployer un data management dédié à la sobriété numérique.
 - Déployer si possible les calculs dans des régions à faible émission de carbone.
 - Proposer une grille de comparaison énergétique des principaux algorithmes (cf. grille pouvant servir de base de réflexion en Annexe 16).
 - Mettre à disposition des outils institutionnels pour le suivi des émissions et les activer sur l'infrastructure informatique.
 - Limiter l'utilisation des ordinateurs à 72 heures maximum par processus par exemple afin de réduire le gaspillage des ressources.
 - Mener des campagnes de sensibilisation sur l'impact environnemental du Machine Learning.
 - Faciliter les compensations institutionnelles pour les émissions ne pouvant être évitées.



“ Sur les modèles de deep learning, on essaie toujours de trouver le bon « trade-off » entre la complexité du modèle et sa précision. Et dans nos appels d'offres, lorsque nous pouvons accepter un temps de latence « élevé » (plusieurs secondes par inférence), nous ajoutons une clause sur l'utilisation de dispositifs à très faible consommation énergétique. ”

Pierre-Louis Bescon - Head of Data & Advanced Analytics - Roquette

Partie 4 - Vers une culture de la frugalité

"On ne se débarrasse pas d'une habitude en la flanquant par la fenêtre ; il faut lui faire descendre l'escalier marche par marche"
Mark Twain

Les défis pour notre monde ne sont pas malheureusement pas récents : consommation exponentielle des ressources, émissions de gaz à effet de serre, recyclage peu efficace, surconsommation. Ils demandent maintenant à repenser profondément notre modèle de société pour une utilisation plus efficiente des ressources au bénéfice de tous. Les entreprises doivent donc mettre en place un système d'innovation permettant d'articuler les bénéfices environnementaux (et sociétaux) et les opportunités de croissance.

1. L'innovation frugale

Inspirée du *Jugaad* indien (une approche permettant de répondre à un besoin bien identifié avec des ressources très limitées), l'innovation frugale s'est d'abord pratiquée dans des pays à faibles ressources pour ensuite s'étendre aux pays développés. Les contraintes amenées par la frugalité permettent d'engager un processus créatif dont le but est de faire émerger de véritables innovations. L'innovation frugale ne vient pas remplacer des approches d'innovation plus classiques mais peut apporter des outils complémentaires pertinents suivant les situations et les contextes.

a. La frugalité

A la différence de la **sobriété** qui s'inscrit davantage dans la remise en cause de la société de consommation par l'économie des ressources utilisées, la **frugalité** repose davantage sur la capacité à les faire fructifier sans en abuser. Le principe de la frugalité est donc plutôt contre-intuitif car il s'agit de faire mieux mais avec moins. Pour une entreprise, la frugalité représente un véritable défi collectif puisqu'elle ne peut s'envisager que sur l'ensemble de ces activités.



“ Notre solution basée sur de l'IA symbolique a toujours été « frugal by design » mais nous mettons beaucoup d'énergie à communiquer sur la souveraineté et l'explicabilité des décisions. C'était le gros sujet du machine learning. Le sujet de la consommation énergétique est arrivé bien plus tard. Nous travaillons à rendre plus visible cette frugalité qui garantit performance et responsabilité. ”

Guillaume Navarre - CCO & Cofondateur – Golem.ai

b. Les 6 principes clés

La situation climatique mais également l'économie post-covid auront besoin d'entrepreneurs innovants socialement et écologiquement engagés. Faire plus avec moins, tel est donc le concept de l'innovation frugale qui repose principalement sur 6 principes clés : Transformer les contraintes en opportunité, Apprendre à faire mieux avec moins, Penser et agir de manière

flexible, Opter pour la simplicité, Intégrer les fragiles et les exclus, Suivre son cœur (cf. Annexe 17).

2. Le cadre du changement

Inspiré du Golden Circle (cf. Figure 15), un outil de communication découvert par Simon Sinek, les auteurs du Guide de l'Innovation Frugale (65) posent trois questions interdépendantes et essentielles pour la définition d'un cadre de changement :

- Le « Quoi » : Quel est l'objectif final ?
- Le « Comment » : Comment l'atteindre de manière rentable ?
- Le « Pourquoi » : Pourquoi chacun devrait-il changer ?

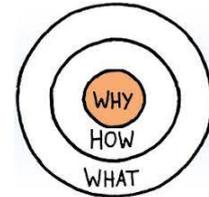


Figure 15 - The Golden Circle

a. Le « Quoi »

Le « Quoi » suppose de définir clairement les objectifs à atteindre. Pour engager les parties prenantes, les objectifs se doivent d'être ambitieux mais crédibles. Par exemple, Microsoft a annoncé en 2020 son ambition de devenir Carbone négatif d'ici 2030 et de compenser d'ici 2050 tout le carbone émis par ses opérations depuis sa fondation en 1975. Cela s'est traduit par un engagement de l'entreprise au plus haut niveau. Il est ensuite essentiel que ces objectifs soient déclinés de manière cohérente dans toute l'entreprise pour qu'ils puissent provoquer un changement d'échelle.

b. Le « Comment »

Pour parvenir à atteindre des objectifs audacieux, l'entreprise se doit d'y consacrer des moyens. Cela passe par la mise à disposition des bons outils aux salariés, ainsi qu'aux fournisseurs et clients pour lui permettre de traiter ses externalités. Cela peut aussi lui demander de changer de modèle économique travaillant sur des systèmes innovants qui peuvent constituer des solutions de rupture. En management, il est courant de dire que la stratégie d'une entreprise ne fait pas le poids face à sa culture. Il s'agit donc de changer l'état d'esprit de l'ensemble du personnel et à tous les niveaux de la hiérarchie. Pour maximiser l'engagement des employés sur des objectifs environnementaux, la culture de l'entreprise joue un rôle important et peut aussi devoir évoluer pour amener des modifications de comportement et favoriser ainsi l'adhésion au changement.

c. Le « Pourquoi »

La question du « Pourquoi » est certainement la plus importante des trois. C'est elle qui donnera le sens et la justification des modifications à apporter sur le modèle économique de

l'entreprise, sur la nécessité de se transformer collectivement et individuellement. Finalement le « Pourquoi » doit constituer la cause partagée défendue autant pour le salarié que pour son entreprise. Cela peut concerner la pérennité des activités de l'entreprise par exemple. Les appels d'offre publics intègrent maintenant de plus en plus des critères environnementaux pour les candidats (réalisation d'un bilan carbone, suivi d'indicateur de performances environnementales, ...). Cela peut également être un enjeu de fidélisation des clients plus soucieux de l'écologie dans leurs actes d'achat mais également des personnels trouvant ainsi davantage d'alignement avec leurs propres valeurs. Mais l'entreprise peut aussi choisir de s'engager dans une démarche à plus long terme pour sécuriser son avenir car les investisseurs sont de plus en plus attentifs aux critères ESG (Environnementaux, Sociaux, Gouvernance).



“ La cohérence de la politique environnementale d'une entreprise est fondamentale. C'est une notion d'histoire qu'elle raconte à ses collaborateurs. Il ne peut y avoir d'injonctions contradictoires. Dans tous les cas, je pars toujours du principe qu'il faut développer une démarche qui est de comprendre, mesurer, décider et agir. ”

Vincent Courboulay - Maître de conférences HDR - La Rochelle Université / Directeur scientifique - INR

3. Le pilotage et l'accompagnement au changement

Pour assurer le changement comme dans tout projet de transformation, celui-ci doit être soutenu au plus haut niveau. Les chefs d'entreprises ne doivent pas se contenter de définir des objectifs ambitieux mais réellement s'engager dans leur mise en œuvre. Cela suppose également d'y consacrer des moyens et des ressources humaines au travers d'une équipe dédiée à la transformation environnementale de l'entreprise. Les missions couvrent généralement **l'accompagnement des projets**, le **développement humain**, le **pilotage de la transformation**, **l'ouverture et la prospective** (66).

a. Accompagnement des projets

Pour atteindre les objectifs environnementaux, une équipe dédiée à l'accompagnement au changement doit être constituée pour formaliser et maîtriser une méthodologie constituée de processus et d'outils. Il peut s'agir de personnes volontaires fédérées autour de la diminution de l'impact environnemental de l'entreprise, mais sur des objectifs ambitieux ces actions relèvent généralement d'une direction de la transformation.

Les OKR (Objectives Key Results)

Difficile cependant d'améliorer ce que l'on ne mesure pas. C'est le principe de la démarche des OKR, Objectifs et Résultats Clés, qui a été développée par Andy Grove, CEO d'Intel dans les années 70. Google a également appliqué cette méthode dès 1999 et elle est encore déployée sur tous leurs projets. Afin d'accompagner l'ambition de l'entreprise sur les enjeux

environnementaux, la méthode OKR semble particulièrement bien adaptée. En plaçant durablement l'entreprise dans la recherche de la performance, cette démarche permet d'atteindre des objectifs ambitieux. Elle permet également un alignement des collaborateurs qui contribuent ainsi collectivement à l'atteinte d'objectifs connectés à la stratégie. Suivant la maturité de l'entreprise, les exemples d'OKR appliqués à une équipe de praticiens de l'IA pourraient être :

Objectifs	Contribuer à la diminution de l'empreinte environnementale de l'entreprise
Résultat clé	Être en capacité d'évaluer l'empreinte environnementale de nos applications d'IA
Tâches	<ul style="list-style-type: none"> - Sensibiliser les membres de l'équipe à l'impact environnemental du numérique - Tester des outils d'évaluation carbone et en choisir un - Évaluer l'empreinte carbone des applications IA de l'année

Objectifs	Diminuer de 10% l'empreinte carbone de nos usages d'IA
Résultat clé	Faire des choix argumentés d'algorithmes
Tâches	<ul style="list-style-type: none"> - Mettre en place un tableau de bord de suivi des GES générés par les usages de l'IA - Justifier l'algorithme retenu en intégrant les paramètres des émissions de GES - Mettre en place des techniques d'optimisation d'algorithme.

Tableau 7 - Exemple d'objectifs - Résultats clés



“ A court terme, il sera difficile de convaincre les équipes métiers de sélectionner une IA plus frugale et moins énergivore tant que cela ne sera pas mis dans leurs objectifs. Mais c'est une méthode très « top-down ». Or la base la conduite du changement est de démontrer à une personne que ce qu'il a gagné et plus important que ce qu'il a à perdre. Les premiers à convaincre vont être les personnes du top management pour que cela soit intégré dans les plans stratégiques. ”

Côme Chatagnon - Partner Associé – AI Builders

Labels et certifications

Encore faut-il savoir d'où l'on part. Aussi, décider de s'engager dans l'obtention de labels et de certifications constitue un bon levier pour commencer à mettre en place les dispositifs de mesure et de structurer la démarche. Le guide des labels RSE et numériques (42) présenté dans la partie 2 donne les éléments de réflexion permettant de choisir le label le plus adapté à la maturité de l'entreprise qui souhaite s'engager dans cette voie. Au-delà de la dimension numérique, la réalisation du bilan carbone (qui devrait bientôt s'imposer à toute entreprise) permet d'avoir une idée précise de son empreinte carbone.

Communauté de pratique

Tout ne peut reposer uniquement sur une équipe dédiée ou sur une direction de la transformation. Il convient de mettre en place une communauté de pratique constituée d'ambassadeurs du changement dont le rôle est d'assurer la cohérence du discours et de mettre en pratique les méthodes du changement. Son rôle est de favoriser le partage de connaissances afin d'impliquer davantage de personnes au travers de préparations

d'événements, d'activités de veille, de co-constructions de nouvelles connaissances et de modules de formation. Les membres de la communauté peuvent ainsi améliorer leurs compétences professionnelles, renforcer leur sentiment d'identité professionnelle ainsi que leur confiance pour capter plus efficacement des pratiques novatrices.

b. Développement humain

Suivant la maturité de l'entreprise sur les enjeux du développement durable et l'activité des personnes, la mobilisation des acteurs s'appuie généralement sur trois étapes :

- Les actions de **sensibilisation** (passage d'un niveau d'inconscient à un premier niveau de conscience)
- Les actions de **formation** (développement de nouvelles connaissances, de pratiques professionnelles, de comportements)
- Les actions d'**ancrage** dans les pratiques de management.

Depuis ces dix dernières années, les modalités de formation se sont beaucoup diversifiées. Plusieurs formules peuvent être déployées : le présentiel, l'apprentissage en ligne (e-learning), le blended learning (e-learning et présentiel), les actions de coaching.

Sensibiliser aux enjeux climatiques et impacts du numérique

Les actions de sensibilisation dont le but est de créer une prise de conscience peuvent revêtir plusieurs formes :

- **Évènements** (conférence, exposition, convention, débats ...)
- **Communication interne** (affiche, newsletter, intranet, note de service, forum en ligne ...)
- **Interpellation** (stickers, bilan carbone individuel, quizz, jeux, challenge, serious game, ...)



“ Plus de 300 collaborateurs ont participé à la fresque du climat chez Talan, nous préparons la fresque du numérique et nous avons proposé le MOOC de l'INR. Les actions de sensibilisation sont nécessaires mais pas suffisantes. Elles sont utiles pour créer un terrain plus propice avant de lancer des projets de transformation ”

Arnaud Soumet - Consultant Senior – Talan Consulting

Il s'agit surtout de donner les éléments de compréhension du problème, d'identifier les pistes pour le réduire et d'évaluer l'impact relatif des actions menées dans ses propres activités. On retrouvera ci-dessous plusieurs actions de sensibilisation assez rapides à mettre en place :

- Informer régulièrement les décideurs par des articles de presse et des publications pour faire progresser leur niveau de compréhension,
- Organiser des rencontres avec des acteurs engagés,
- Organiser des journées dédiées (fresque du climat, fresque numérique, journée de l'innovation frugale),
- Proposer des parcours de sensibilisation en ligne (MOOC INRIA Impacts environnementaux du numérique (67), MOOC INR Numérique Responsable (68) (69))

- Lancer des actions ludiques au travers de challenges ou défis.

Former

Les actions de formation doivent faire l'objet d'un plan de formation qui précisera le cahier des charges des modules de formation, les personnes concernées, les ressources associées, le planning prévisionnel et le coût (généralement assez important en gestion du changement). Les formations peuvent être collectives sur une population ciblée ou individuelles (de type coaching). Un levier intéressant en termes d'ancrage est de réaliser les formations par des personnes relais (ambassadeurs) de l'entreprise. Ces relais, formés préalablement pourront ainsi devenir des experts en interne. Plus simple à mettre en place, des parcours de formation peuvent être intégrés au processus d'intégration des nouveaux embauchés et proposés sur la base du volontariat aux salariés déjà présents. Il convient néanmoins d'effectuer au préalable l'analyse des besoins pédagogiques en fonction des objectifs visés. On distingue généralement des besoins de formation en termes de savoir, savoir-faire et savoir être. Dans le domaine du numérique, l'Institut du Numérique propose un parcours original de montée compétences en 3 niveaux : connaissances, compétences, ambassadeurs (cf. Annexe 18). Celui-ci peut constituer un parcours pertinent à la formation des personnes relais sur les **impacts du numérique** de l'entreprise comme des formations sur la façon de mener un **bilan carbone** ou une **démarche d'écoconception**. D'autres formations sont également à prévoir sur le volet managérial afin d'aider les manager à co-construire le changement et à créer les conditions d'acceptation chez les collaborateurs.

Ancrer le changement (Évolution et innovation managériale)

Aussi, quel que soit leur niveau dans l'entreprise (manager de terrain, manager intermédiaire et top management) les manager doivent disposer des compétences nécessaires pour déployer et ancrer le changement. Celles-ci sont identifiées dans la boucle Manager/ Leader (cf. Annexe 19) qui regroupe 7 compétences clés : l'organisation du travail, l'animation d'équipe, la mise en place de dispositifs de contrôle, la **gestion du changement**, la prise de décision, la capacité de délégation et le pilotage de la performance. Un diagnostic initial au travers d'une grille d'auto-évaluation ou d'une méthode d'évaluation à 360° doit permettre à chaque manager de s'interroger sur ses pratiques managériales.

Il convient également de mettre en place des actions **d'innovations managériales** pour faire évoluer les pratiques de management en vue de garantir l'atteinte des objectifs de changement mais également de pérenniser les comportements. Ces pratiques d'innovations managériales ont pour but de revoir les modes de management basés sur le modèle hiérarchique afin de libérer la parole, favoriser l'écoute et positionner le collaboratif au service du changement. Ces nouveaux modes de coordination des acteurs permettent de favoriser l'innovation et de

développer l'autonomie des personnes. Différentes activités permettent aux acteurs de s'interroger et de faire évoluer leurs pratiques managériales, on peut citer le tutorat, le peer coaching, le design thinking, le co-design, les ateliers participatifs, les méthodes agiles, les hackathons, les incubateurs, ...

c. Pilotage de la transformation

Comme tout projet, la conduite du changement s'organise, se planifie et se pilote. La mise en place de tableaux de bord est alors nécessaire pour s'assurer de la bonne réalisation des activités prévues (communication, sensibilisation, formation, accompagnement), suivre l'évolution de la perception du changement par les équipes, l'efficacité des actions de formation et les avancées dans les différents services. La réussite d'un changement repose donc sur des mesures précises effectuées régulièrement.

Définir une feuille de route du changement

Établie en phase initiale, la feuille de route présente une vue générale du projet de changement et explique à toutes les parties prenantes ce qui sera fait. Elle doit préciser la stratégie retenue, l'équipe de pilotage, le budget alloué, les ressources associées, les livrables et jalons retenus ainsi que le planning prévisionnel. Il s'agit également d'identifier tout risque potentiel à prendre en considération vis-à-vis de la progression du projet. La feuille de route permet ensuite au travers de plusieurs indicateurs de vérifier régulièrement si les efforts portent leurs fruits et que le projet est en bonne voie. On retrouvera un exemple de feuille de route générique en Annexe 20 ainsi qu'un exemple de tableau de suivi des actions Annexe 21.

Évaluer le changement

L'appréciation du changement en cours fait partie des actions de pilotage. Cette appréciation peut se mesurer, pour une population cible, aux travers de 4 indicateurs réunis sous l'intitulé de baromètre ICAP (70) :

- le **taux d'information (I)** : il permet de connaître la proportion de personnes correctement informées. A l'annonce du projet de changement, ce taux doit être a minima de 25% pour en assurer une visibilité suffisante (voire 50% pour un changement d'organisation ou de culture).
- le **taux de compréhension (C)** : il vise à évaluer la proportion de personnes qui ont compris les enjeux du changement au regard de leurs activités quotidiennes. Alors que le taux d'information évolue assez linéairement, le taux de compréhension progresse plutôt par palier. La compréhension reste indispensable pour « faire » et envisager un déploiement plus large d'actions efficaces. Un taux de 60% constitue un point de bascule.
- le **taux d'adhésion (A)** : Il représente la proportion de personnes qui adhèrent au projet de changement et qui le jugent positivement. Après une phase d'excitation lors du lancement où l'adhésion doit être élevée, le taux d'adhésion peut prendre la forme d'une courbe en U avec

une baisse liée aux efforts à mettre en œuvre pour l'apprentissage. Il est donc important de communiquer sur les succès pour faire remonter ce taux au-delà de 50% et éviter ainsi le désengagement.

Le **taux de participation (P)** : cet indicateur caractérise le pourcentage de personnes qui participent concrètement au changement par des actions ponctuelles ou récurrentes. Ce taux évolue généralement manière de linéairement croissante mais avec des oscillations liées aux obligations de production. Une réactivation régulière est donc nécessaire pour maintenir la progression régulière de la participation.

Ces quatre indicateurs sont habituellement établis par l'intermédiaire de questionnaires trimestriels en ligne comportant entre 8 et 20 questions maximum. Un taux de retour d'au moins 40% est conseillé. Un exemple de questionnaire est présenté en Annexe 22.

Mesurer la performance

Les indicateurs Métiers doivent permettre de mesurer l'écart de performance entre le début et la fin de la conduite du changement. Si le changement concerne la réduction des émissions de gaz à effet de serre par exemple, on pourra ainsi s'appuyer utilement sur les OKR dont les objectifs et résultats clés auront été déclinés de manière chiffrée sur chaque service. Leur évolution permet de matérialiser l'impact de la conduite du changement sur la dimension opérationnelle. Le rapport du Shift Project (59) organise différents indicateurs sur les infrastructures et le réseau, les logiciels et application ainsi que les usages métiers pour remonter au bilan carbone d'un système d'information.

Évaluer la capacité à changer

Afin d'orienter les actions à mener en termes d'accompagnement, de communication et de formation, il est préférable d'évaluer la perception du changement par les acteurs. On retrouve sur la Figure 16 les sept catégories de profils réparties en fonction de leur perception du changement et de leur niveau d'implication. Les « avocats » sont les inconditionnels du changement avec un niveau d'implication maximum et une perception très positive. A l'inverse, les « opposants » sont ouvertement contre le projet de changement. Il est préférable de consacrer du temps auprès des alliés et de les valoriser plutôt que de convaincre les opposants. Ce se sont les alliés qui feront la promotion du projet. Les déchirées sont à traiter avec attention car ils portent généralement des responsabilités importantes. Généralement leur discours peut aller de l'intérêt à la critique suivant leurs interlocuteurs.

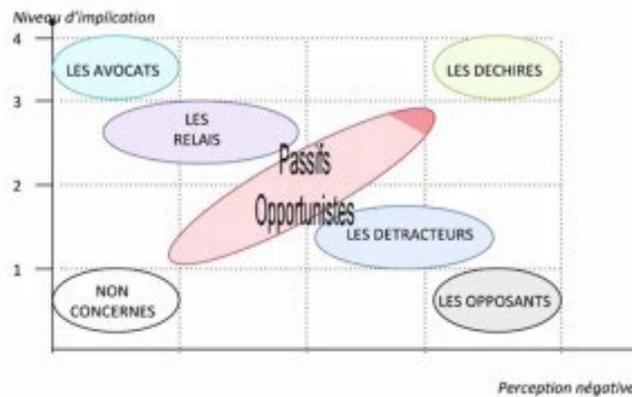


Figure 16 - Cartographie des alliés (Source : outilspourdiriger.fr)

On retrouvera Annexe 23 des exemples d'actions type à mettre en place pour chaque catégorie d'acteurs.

Initier et suivre des programmes de transformation culturelle

Pour mener à bien un projet de transformation, il est indispensable de tenir compte de la culture de l'entreprise. La description de la culture d'entreprise s'appuie généralement sur la marguerite sociologique (cf. Annexe 24) qui reprend les sept éléments d'un système culturel : les rites, les routines, les mythes, les symboles, les systèmes de contrôle, les structures organisationnelles, les structures de pouvoir. Atteindre collectivement des objectifs ambitieux repose nécessairement sur le développement d'un état d'esprit favorisant la coopération, le travail en équipe, l'engagement, l'initiative et la créativité. La culture d'entreprise doit être en cohérence avec les valeurs portées par l'entreprise et doit faire sens pour les collaborateurs. On retrouvera en Annexe 25 plusieurs propositions d'actions organisées en 4 axes pour changer et renforcer la culture d'entreprise dans cette voie :

- Actions de la gouvernance et des dirigeants
- Récompenses culturellement cohérentes
- Sélection et socialisation des employés
- Alignement des artefacts « observables »

Les changements attendus, liés aux enjeux environnementaux, revêtent davantage une dimension sociologique plutôt qu'organisationnelle. La conduite du changement est donc à orienter sur la collaboration et l'éducation des acteurs plutôt que sur des pratiques directives et interventionnistes.

d. Ouvertures et perspectives

L'équipe en charge d'accompagner le changement se doit également de favoriser l'accès à des organisations extérieures pour nourrir la réflexion des managers et des collaborateurs. Il peut s'agir d'entreprises, de think tank, d'organismes de recherche, d'institution académiques, d'incubateurs de start-up. Au-delà des possibilités d'échanges entre pairs, ces opportunités

de rencontres contribuent largement aux activités de veille et de prospective indispensables pour imaginer et préparer l'entreprise aux enjeux du futur. Les « learning expedition » ou voyages apprenants permettent également d'explorer des écosystèmes innovants et constituent des leviers puissants du changement. Pour développer la culture d'innovation, de plus en plus d'entreprises mettent en place un incubateur interne pour accueillir et accompagner les porteurs de projets en mode start up. Le collaborateur a ainsi un accès immédiat à des ressources matérielles, financières et humaines pour développer son idée. La nature de l'incubateur est très stratégique puisque que le portefeuille de projets doit servir la conduite du changement. Il importe donc qu'il soit géré par la direction de la transformation ou la direction de l'innovation.

4. Du changement à la transformation

La question de la responsabilité des entreprises monte fortement dans l'opinion publique et auprès des dirigeants d'entreprise avec pour objectif principal de ne plus uniquement maximiser les profits mais également d'agir pour le bien commun. Accélééré par le réchauffement climatique et par la crise sanitaire de 2020, ce phénomène amène les entreprises à s'interroger sur leur modèle pour concilier une performance économique durable et un impact social et environnemental positif. Pour parvenir à ce nouveau modèle, cela suppose de placer la responsabilité au cœur des activités de l'entreprise. Dans leur ouvrage (71), les auteurs expliquent en quoi investir dans des initiatives sociétales peut être favorable à la performance économique de l'entreprise : **l'amélioration de la réputation**, la **fidélisation des consommateurs** conduisant à une hausse des marges, la **motivation des employés** se traduisant par une meilleure productivité, des arguments **d'attractivité pour le recrutement des jeunes générations**, une **meilleure gestion des risques** par la sécurisation de la chaîne de valeur, une **meilleure capacité à innover** et enfin la **confiance des marchés financiers** entraînant une baisse des primes de risque. Une démarche de responsabilité sociétale est destinée à intégrer l'ensemble de l'écosystème de l'entreprise au sein de sa raison d'être. En France, la loi PACTE de 2019 permet maintenant de renforcer la responsabilité sociale et environnementale des entreprises en leur confiant la définition de leur raison d'être et le choix de devenir une société à mission.

a. La raison d'être

Pour une entreprise, identifier sa raison d'être consiste à définir quelle est son utilité. Pour cela, comme indiqué dans la loi PACTE, il lui convient de mener une réflexion de fond au travers d'un « doute existentiel fécond » pour déterminer ses responsabilités et le sens profond de ses engagements. Bâtir un avenir sur le long terme ne se résume pas uniquement à faire du profit et à satisfaire ses actionnaires mais aussi à générer de la valeur pour

l'ensemble de ses parties prenantes. La raison d'être d'une entreprise constitue donc sa contribution d'aujourd'hui et de demain aux grands défis. La raison d'être va au-delà d'une valeur, elle exprime la singularité de l'entreprise, ce qu'elle apporte à ses parties prenantes et à la société ainsi que sa vision d'avenir. On retrouve ici le « pourquoi » du Golden Circle. Dans leur guide sur la raison d'être de l'entreprise (72), les auteurs soulignent le rôle clé des dirigeants et recommandent d'associer le plus en amont possible le conseil d'administration à la démarche. La définition d'une raison d'être demande un temps long (de 3 à 6 mois minimum) pour permettre la réflexion collective. Une fois définie, il est raisonnable de laisser « vivre » la raison d'être a minima avant son inscription dans les statuts. L'avantage de la raison d'être réside dans l'engagement de moyens que l'entreprise consacre pour y être alignée mais il n'y a pas d'engagement de résultats.



“ Une raison d'être bien faite a tellement d'avantages que ça vaut la peine d'y passer du temps. Ça amène un alignement et une cohérence générale entre la raison d'être, la stratégie et les actions au quotidien. C'est un alignement qui est très aidant pour le management, d'autant plus si la raison d'être a été élaborée de manière participative et collaborative. ”

Caroline Truc - DRH et accompagnement du changement, Coach certifiée HEC - Indépendante

b. L'entreprise à mission

Adopter la qualité d'entreprise à mission est plus ambitieux mais aussi plus engageant. En effet la loi prévoit des modalités de suivi de l'exécution des missions. L'entreprise a ainsi la possibilité de définir des indicateurs extra financiers comme la réduction des émissions de CO₂ par exemple. Un comité de mission et un organisme tiers indépendant (OTI) devront garantir ensuite la validité des engagements pris. Le contrôle d'un OTI a lieu tous les 2 ans (3 ans pour les entreprises de moins de 50 salariés). On retrouvera en Annexe 26 une grille d'autoévaluation de la mission qui s'appuie sur 5 critères : singularité de la mission, mobilisation des parties prenantes, alignement avec le business model, contribution sociétale et ambition d'innovation. La « mission » représente ainsi pour une entreprise l'opportunité de pivoter l'ensemble de ses activités pour être alignée avec sa raison d'être. Elle constitue un levier puissant dans la conduite du changement pour repenser le modèle managérial, développer les initiatives et la responsabilisation des collaborateurs, favoriser l'innovation, renforcer tout l'écosystème et porter **les questions RSE au niveau stratégique**.

5. IA frugale, un audit des pratiques nécessaire

En synthèse, adopter une approche frugale dans le cadre d'un projet d'IA revêt plusieurs dimensions et il est donc nécessaire de se poser ou de poser un certain nombre de questions au préalable. On retrouvera ci-dessous par thématique quelques questions essentielles :

	Questions	Réponses
RSE	L'entreprise a-t-elle mis en place une démarche RSE ?	
	Dispose-t-on d'un bilan carbone sur l'ensemble des activités ?	
	L'entreprise a-t-elle déjà mené des analyses de cycle de vie pour ses produits, ses services ?	
Labels	L'entreprise dispose-t-elle de labels environnementaux, de labels sur le numérique ?	
	Une démarche de labélisation est-elle envisagée ?	
Données	Les données pour le projet sont-elles facilement accessibles ?	
	Dispose-t-on de grands volumes de données ?	
	Les données sont-elles spécifiques au secteur d'activité ?	
	La qualité des données a-t-elle été évaluée ?	
	Les données sont-elles traitées avant stockage ?	
	La problématique nécessite-t-elle du traitement d'images, du langage naturel ?	
Centre de données	Les données et les calculs sont-ils gérés sur un serveur local, sur un cloud privé, public ?	
	Connaît-on les indicateurs de consommation et environnementaux du centre de données ?	
	Peut-on paramétrer le type de processeurs utilisés ?	
	Connait-on le mix énergétique du centre de données ?	
	Un faible temps de latence est-il indispensable pour l'usage envisagé ?	
Choix de modèles	Faut-il nécessairement déployer une solution d'IA ?	
	Existe-t-il des règles métiers sur la problématique à traiter ?	
	Les métiers ont-ils été consultés ?	
	Des solutions d'IA symbolique peuvent-elles répondre à la problématique ?	
	Existe-t-il des modèles déjà existants pouvant répondre à la problématique ?	
	Peut-on mettre en place du transfer learning ?	
	Des démarches d'optimisation de modèles sont-elles menées ?	

	Des outils d'évaluation d'empreinte carbone sont-ils à disposition ?	
	L'entreprise dispose-t-elle de références en termes de consommation énergétique et d'empreinte carbone de modèle d'IA ?	
Sensibilisation	Les salariés ont-ils été sensibilisés à la problématique environnementale et aux impacts du numérique (fresque du climat, du numérique, journée de l'innovation frugale) ?	
	Des ambassadeurs ou des personnes relais ont-ils été formés ?	
	Des journées dédiées à l'environnement, au numérique responsable sont-elles organisées ?	
Formation	Y a-t-il un cadre d'apprentissage (coaching, mentoring, communauté de pratique) ?	
	Quel est le plan de formation de l'entreprise ?	
	Des certifications sont-elles proposées, demandées à des salariés (numérique responsable)	
Management	Les objectifs environnementaux collectifs sont-ils déclinés en objectifs individuels ?	
	Comment sont valorisés les efforts individuels en termes de réduction de GES ?	
	Les managers créent-ils un cadre propice aux initiatives ?	
Changement	L'entreprise est-elle dans un processus de changement ? Si oui, lequel ?	
	Une équipe est-elle dédiée à la conduite du changement ?	
	Quels sont les messages portés par les dirigeants ?	
Culture d'entreprise	Quelle est la culture de l'entreprise ? Comment s'incarne-t-elle ?	
	Le salarié peut-il facilement faire des propositions d'amélioration ou d'innovation ? Est-ce valorisé ?	
	L'entreprise dispose-t-elle d'une raison d'être en lien avec la RSE ?	
	Les salariés connaissent-ils la raison d'être de l'entreprise ?	
	L'organisation a-t-elle la qualité d'entreprise à mission ?	

Tableau 8 - IA Frugale - Trame d'audit

Cette trame non exhaustive constitue un premier outil destiné à évaluer rapidement le niveau de maturité d'une entreprise pour aborder un projet d'IA de manière plus vertueuse. L'approche technologique ne peut suffire et c'est bien l'intelligence des humains qui pourra relever ce défi.

Conclusion

La nécessité d'agir concrètement et rapidement pour l'environnement ne fait maintenant plus de doute. Il y a urgence. La prise de conscience progressive, les réglementations en cours d'élaboration, l'évolution des mentalités doivent maintenant concourir à des pratiques plus vertueuses pour l'avenir de la planète. Le numérique fait désormais l'objet d'une plus grande attention car son développement pose assurément la question de sa soutenabilité. Il importe cependant que chacun puisse agir à son niveau et en connaissance de cause afin d'en limiter l'impact environnemental.

Tout utilisateur de ressources informatiques, et notamment le praticien de l'intelligence artificielle, se doit donc de s'interroger sur ses pratiques. La dimension environnementale devra de plus en plus être intégrée dans le management d'un projet d'intelligence artificielle en considérant les impacts globaux (positifs et négatifs) de la solution développée. Les métriques environnementales sont encore trop peu présentes dans l'évaluation d'un projet d'IA. Cela demande d'être outillé mais également d'appréhender des concepts qui vont bien au-delà de la maîtrise de tel ou tel algorithme. La facilité d'accès à l'ensemble de ces outils ne doit pas empêcher d'adopter une démarche plus raisonnée, plus frugale.

Mais il ne suffit pas de décréter une approche frugale. Celle-ci passe nécessairement par la compréhension du contexte environnemental, par la capacité à mieux évaluer l'impact de ses activités, par son intégration dans l'ensemble des métiers et notamment ceux de l'IA. Cette inflexion vers davantage de frugalité constitue un changement de paradigme pour les organisations habituellement plus concentrées sur les performances brutes et l'atteinte des objectifs financiers. Ce changement se doit donc d'être accompagné dans le temps. Plus qu'une dimension stratégique, la frugalité doit s'intégrer dans l'état d'esprit de l'entreprise et se vivre pleinement au quotidien. En cela, la frugalité relève véritablement de la culture d'entreprise.

Pour le collaborateur, c'est avant tout la question du sens donné à ses activités. Ainsi, de plus en plus d'entreprises perçoivent maintenant l'intérêt de conjuguer profitabilité et responsabilité en se dotant d'une raison d'être. Intégrée dans les statuts, celle-ci fournit alors un cadre pour les décisions stratégiques. Enfin la qualité de société à mission de la loi Pacte renforce les engagements de l'entreprise avec le suivi d'indicateurs extra financiers valorisant ainsi ses contributions à la société et à l'environnement.

Faire aussi bien voire mieux avec moins fera inmanquablement partie des nouvelles approches des entreprises. Il est temps de s'y préparer.

Bibliographie

1. **Pôle interministériel de prospective et d'anticipation des mutations économiques (PIPAME).** *Prospective - Intelligence Artificielle - Etat de l'art et perspectives pour la France*. 2019. ISBN : 978-2-11-152634-1 / ISSN : 2491-0058.
2. **Julia, Luc.** *On va droit dans le mur ?* Paris : First Editions, 2022. ISBN : 9782412078679.
3. **World Economic Forum.** *The Global Risks Report 2022 - 17th Edition*. 2022. ISBN: 978-2-940631-09-4.
4. **Dennis Meadows, Donella Meadows, Jorgen Randers.** *Les limites à la croissance (dans un monde fini)*. Paris : l'écopoche, 2022. ISBN : 978-2-37425-332-9.
5. **ADEME.** *Transitions(s) 2050 - Choisir maintenant, agir pour le climat. ADEME*. [En ligne] 30 Novembre 2021. <https://librairie.ademe.fr/cadic/6531/transitions2050-rapport-compresse.pdf?modal=false>.
6. **The Shift Project.** *Climat, crises : Le plan de transformation de l'économie française*. Paris : Odile Jacob, 2022. ISBN : 978-2-7381-5427-9.
7. **Grandi, Michel De.** ONU : les objectifs de développement mis en péril par la crise du coronavirus. *LesEchos*. [En ligne] 21 Septembre 2020. <https://www.lesechos.fr/monde/enjeux-internationaux/onu-les-objectifs-de-developpement-mis-en-peril-par-la-crise-du-coronavirus-1244269>.
8. **Villani, Cédric.** *Donner un sens à l'intelligence artificielle - Pour une stratégie nationale et européenne*. 2018. ISBN : 978-2-11-145700-3.
9. **Gouvernement Français.** *France 2030 - Stratégie Nationale pour l'Intelligence Artificielle - 2ème phase*. [En ligne] 8 Novembre 2021. <https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2021-11/dossier-de-presse---strat-gie-nationale-pour-l-intelligence-artificielle-2e-phase-14920.pdf>.
10. **UNESCO.** Les États membres de l'UNESCO adoptent le tout premier accord sur l'éthique de l'intelligence artificielle. *UNESCO*. [En ligne] 25 Novembre 2021. <https://fr.unesco.org/news/etats-membres-lunesco-adoptent-premier-accord-lethique-lintelligence-artificielle>.
11. —. Recommandation sur l'éthique de l'intelligence artificielle. *UNESDOC - Bibliothèque Numérique*. [En ligne] 2021. https://unesdoc.unesco.org/ark:/48223/pf0000380455_fre.
12. **CNN (Conseil National du Numérique).** Feuille de route sur l'environnement et le numérique. *ccnumerique.fr*. [En ligne] 08 Juillet 2020.

<https://cnnumerique.fr/files/uploads/2020/CNNum%20-%20Feuille%20de%20route%20environnement%20%26%20numerique.pdf>.

13. **Ministère de la Transition écologique.** Feuille de route "Numérique & environnement". [En ligne] 23 02 2021.

https://www.ecologie.gouv.fr/sites/default/files/Feuille_de_route_Numerique_Environnement.pdf.

14. —. Mission interministérielle - Numérique écoresponsable. *Réglementation - Numérique écoresponsable*. [En ligne] 15 02 2022.

<https://ecoresponsable.numerique.gouv.fr/réglementations/>.

15. **Guillaume Chevrollier, Jean-Michel Houllégatte.** Rapport d'information, fait au nom de la commission de l'aménagement du territoire et du développement durable par la mission d'information sur l'empreinte environnementale du numérique. *Senat*. [En ligne] 24 Juin 2020. <https://www.senat.fr/rap/r19-555/r19-5551.pdf>.

16. **Sénat.** 25 propositions pour une transition numérique écologique. *www.senat.fr*. [En ligne] Juin 2020.

http://www.senat.fr/fileadmin/Fichiers/Images/redaction_multimedia/2020/2020-Documents_pdf/20200624_Conf_presse_Dev_Dur/20200624_Conf_Dev_Dur_Infographie.pdf.

17. **AFNOR.** Feuille de route stratégique - La normalisation de l'Intelligence Artificielle. *AFNOR Normalisation*. [En ligne] 29 Mars 2022. <https://info.promotion-afnor.org/file/Wb/zzzzz/Feuille-de-route-normalisation-IA.pdf>.

18. **Bordage, Frédéric.** Une loi pour réduire l'empreinte environnementale du numérique en France. *GreenIT*. [En ligne] 14 12 2021. <https://www.greenit.fr/2021/12/14/une-loi-pour-reduire-lempreinte-environnementale-du-numerique-en-france/>.

19. **Courrier international.** États-Unis. Cinq grandes universités américaines assignées en justice par leurs étudiants au nom du climat. *Courrier international*. [En ligne] 17 Février 2022. <https://www.courrierinternational.com/article/etats-unis-cinq-grandes-universites-americales-assignees-en-justice-par-leurs-etudiants-au>.

20. **CREDOC, Alina Koschmieder, Lucie Brice-Masencal, Sandra Hoibian.**

Environnement: les jeunes ont de fortes inquiétudes mais leurs comportements restent consommateurs. *CREDOC*. [En ligne] Décembre 2019.

<https://www.credoc.fr/publications/environnement-les-jeunes-ont-de-fortes-inquietudes-mais-leurs-comportements-restent-consumeristes>. ISSN 0295-9976.

21. **Efoui-Hess, Maxime.** Climat : l'insoutenable usage de la vidéo en ligne. *The Shift Project*. [En ligne] Juillet 2019. <https://theshiftproject.org/wp-content/uploads/2019/07/2019-01.pdf>.
22. **Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation.** *Rapport " Sensibiliser et former aux enjeux de la transition écologique et du développement durable dans l'enseignement supérieur"*. [En ligne] Février 2022. <https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2022-02/rapport-former-aux-enjeux-de-la-transition-ecologique-dans-l-es-pdf-16808.pdf>.
23. **The Shift Project.** Former l'ingénieur du XXI^e siècle. *The Shift Project*. [En ligne] 10 Mars 2022. <https://theshiftproject.org/article/publication-rapport-former-lingenieur-du-21esiecle/>.
24. **Bordage, Frédéric.** *Sobriété numérique*. Paris : Buchet Chastel, 2019. ISBN : 978-2-283-03215-2 / ISSN : 2492-9107.
25. —. Empreinte environnementale du numérique mondial. *GreenIT.fr*. [En ligne] Septembre 2019. https://www.greenit.fr/wp-content/uploads/2019/10/2019-10-GREENIT-etude_EENM-rapport-accessible.VF_.pdf.
26. **Gaudiaut, Tristan.** Les plus gros pollueurs du monde. *statista*. [En ligne] 1 Novembre 2021. <https://fr-statista-com.devinci.idm.oclc.org/infographie/9668/plus-gros-emetteurs-de-co2-dans-le-monde/>.
27. **ADEME.** Avis de l'ADEME - Tous les acteurs doivent agir collectivement pour la neutralité carbone, mais aucun acteur ne devrait se revendiquer neutre en carbone. *ADEME presse*. [En ligne] 1 avril 2021. <https://presse.ademe.fr/2021/04/avis-de-lademe-tous-les-acteurs-doivent-agir-collectivement-pour-la-neutralite-carbone-mais-aucun-acteur-ne-devrait-se-revendiquer-neutre-en-carbone.html>.
28. —. Utilisation de l'argument de « neutralité carbone » dans les communications. *ADEME La Librairie*. [En ligne] Février 2022. <https://librairie.ademe.fr/developpement-durable/5335-utilisation-de-l-argument-de-neutralite-carbone-dans-les-communications.html>.
29. **Donnelly, Caroline.** Pourquoi il devient critique d'évaluer l'usage de l'eau dans le datacenter. *LeMagIT*. [En ligne] 22 12 2021. <https://www.lemagit.fr/conseil/Pourquoi-il-devient-critique-devaluer-lusage-de-leau-dans-le-datacenter>.
30. **Pitron, Guillaume.** *L'enfer numérique*. Paris : LLL - Les Liens qui Libèrent, 2021. ISBN : 979-10-209-0996-1.

31. **The Shift Project.** Lean IT - pour une sobriété numérique. *The Shift Project*. [En ligne] Octobre 2018. <https://theshiftproject.org/wp-content/uploads/2018/11/Rapport-final-v8-WEB.pdf>.
32. **CNRS, Laure Cailloce.** Numérique : le grand gâchis énergétique. *CNRS - Le Journal*. [En ligne] 18 Mai 2018. <https://lejournal.cnrs.fr/articles/numerique-le-grand-gachis-energetique>.
33. **ADEME.** La face cachée du numérique - Réduire les impacts du numérique sur l'environnement. *librairie.ademe.fr*. [En ligne] Novembre 2019. <https://librairie.ademe.fr/cadic/2351/guide-pratique-face-cachee-numerique.pdf?modal=false>.
34. **CIGREF.** Sobriété Numérique - Référentiel de bonnes pratiques pour accompagner les démarches de sobriété numérique dans les organisations. *cigref.fr*. [En ligne] Octobre 2020. <https://www.cigref.fr/wp/wp-content/uploads/2020/10/Cigref-Sobriete-Numerique-Referentiel-100-Bonnes-pratiques-de-sobriete-numerique-October-2020.xlsx>.
35. **AFNOR.** AFNOR SPEC 2201. *www.boutique.afnor.org*. [En ligne] Avril 2022. <https://www.boutique.afnor.org/fr-fr/norme/afnor-spec-2201/ecoconception-des-services-numeriques/fa203506/323315>.
36. **Courboulay, Vincent.** *Vers un numérique responsable - Repensons notre dépendance aux technologies digitales*. Paris : Actes Sud, 2021. ISBN 978-2-330-14375-6.
37. **ADEME, Cécile Diguët, Fanny Lopez.** L'impact spatial et énergétique des data centers sur les territoires. *L'institut Paris Région*. [En ligne] Février 2019. https://www.institutparisregion.fr/fileadmin/NewEtudes/Etude_1780/RAPPORT_ENERNUM_FINAL-_1mars2019.pdf.
38. **statista.** Number of data centers worldwide in 2022, by country. *statista*. [En ligne] 17 Mars 2022. <https://www-statista-com.devinci.idm.oclc.org/statistics/1228433/data-centers-worldwide-by-country/>.
39. **Synergy.** Pipeline of Over 300 New Hyperscale Data Centers Drives Healthy Growth Forecasts. *Synergy*. [En ligne] 23 Mars 2022. <https://www.srgresearch.com/articles/pipeline-of-over-300-new-hyperscale-data-centers-drives-healthy-growth-forecasts>.
40. *Indicateurs pour un datacenter efficient ? Mesurer pour améliorer !* **Gabrielle Feltin, Bernard Bouterin, Xavier Canehan.** Nantes : JRES 2017, 2017.
41. **ADEME - ARCEP.** Evaluation de l'impact environnemental du numérique en France. *ARCEP*. [En ligne] 19 Janvier 2022. https://www.arcep.fr/uploads/tx_gspublication/etude-numerique-environnement-ademe-arcep-note-synthese_janv2022.pdf.

42. **Alliance Green IT.** Guide des labels RSI & Numérique Responsable - Comment valoriser votre engagement ? *www.alliancegreenit.org*. [En ligne] Décembre 2021. <https://alliancegreenit.org/media/ressource-publication/guide-labels-agit.pdf>.
43. **Lemberger, Pirmin.** IA symbolique et Deep Learning – le temps de la réconciliation ? [En ligne] <https://cdn2.hubspot.net/hubfs/2820264/IA%20Pirmim/IA%20symbolique%20et%20DeepLearning.pdf>.
44. **Levy, Mick.** *Sortez vos données du frigo - Une entreprise performante avec la data et l'IA*. Malakoff : Dunod, 2021. ISBN : 987-2-10-081705-4.
45. **ADVAES, Emmanuelle Olivié-Paul.** L'IA, une aide précieuse pour maîtriser l'empreinte environnementale ? *www.advaes.fr*. [En ligne] 8 Aout 2021. <https://advaes.fr/2021/08/08/ia-une-aide-precieuse-pour-maitriser-lempreinte-environnementale/>.
46. **Patterson, David, et al.** Carbon Emissions and Large Neural Network Training. *arxiv*. [En ligne] <https://arxiv.org/ftp/arxiv/papers/2104/2104.10350.pdf>.
47. **Raoul, Gaétan.** Deep learning : les promesses de la technique « Mixture of Experts ». *LeMagIT*. [En ligne] 02 Février 2022. <https://www.lemagit.fr/actualites/252512844/Deep-learning-les-promesses-de-la-technique-Mixture-of-Experts>.
48. **Whitten, Allison.** AI Overcomes Stumbling Block on Brain-Inspired Hardware. *Quantamagazine*. [En ligne] 17 Février 2022. <https://www.quantamagazine.org/ai-overcomes-stumbling-block-on-brain-inspired-hardware-20220217/>.
49. **Ezratty, Olivier.** Les usages de l'intelligence artificielle. *Opinions Libres*. [En ligne] Février 2021. <https://www.oezratty.net/wordpress/2021/usages-intelligence-artificielle-2021/>.
50. **statista.** Number of IoT connected devices worldwide 2019-2030. *statista*. [En ligne] 17 Mars 2022. <https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/>.
51. **Capgemini Research Institute.** Climate AI - How artificial intelligence can power your climate action strategy. [En ligne] 15 Juillet 2020. https://www.capgemini.com/wp-content/uploads/2020/11/Climate-AI_Final.pdf.
52. **Impact AI.** AI for Green : la branche verte de l'intelligence artificielle. *Impact AI*. [En ligne] 28 Mai 2021. <https://www.impact-ai.fr/fr/2021/05/28/ai-for-green-une-serie-speciale-par-impact-ai/>.
53. **Open Studio.** Intelligence artificielle & protection de l'environnement - Le paradoxe d'une technologie énergivore au service des défis écologiques de demain. *Open Studio*. [En ligne]

21 10 2020.

https://www.openstudio.fr/app/uploads/2022/01/DGOpenStudio_LivreIAEnvironnement_15x21_V5_PublicationNumerique.pdf.

54. **Emma Strubell, Ananya Ganesh, Andrew McCallum.** Energy and Policy Considerations for Deep Learning in NLP. *arxiv*. [En ligne] July 2019. <https://arxiv.org/abs/1906.02243>.

55. **Heaven, Will Douglas.** 2021 was the year of monster AI models. *MIT Technology Review*. [En ligne] 21 Décembre 2021. <https://www.technologyreview.com/2021/12/21/1042835/2021-was-the-year-of-monster-ai-models/>.

56. **Gholami, Ami.** AI and Memory Wall. *www.medium.com*. [En ligne] 29 Mars 2021. <https://medium.com/riselab/ai-and-memory-wall-2cb4265cb0b8>.

57. **Google Research, Gaurav Menghani.** Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. *arxiv*. [En ligne] 21 Juin 2021. <https://arxiv.org/abs/2106.08962>.

58. **Morgan, Lison et Raoul, Gaétan.** Mesurer l’empreinte ambivalente de l’IA sur l’environnement. *LeMAGIT*. [En ligne] 17 01 2022. <https://www.lemagit.fr/actualites/252512081/Mesurer-lempreinte-ambivalente-de-lIA-sur-lenvironnement>.

59. **The Shift Project, Hugues Ferreboeuf.** Déployer la sobriété numérique. *The Shift Project*. [En ligne] Janvier 2020. https://theshiftproject.org/wp-content/uploads/2020/01/2020-01-16_Rapport-intermédiaire_Déployer-la-sobriété-numérique_v5.pdf.

60. **Luccioni, Sasha et Ligozat, Anne-Laure.** A Practical Guide to Quantifying Carbon Emissions for Machine Learning researchers and practitioners. *HAL open science*. [En ligne] 13 Octobre 2021. <https://hal.archives-ouvertes.fr/hal-03376391/document>.

61. **Patterson, David.** Good News About the Carbon Footprint of Machine Learning Training. *Google AI Blog*. [En ligne] 15 Février 2022. <https://ai.googleblog.com/2022/02/good-news-about-carbon-footprint-of.html>.

62. **Lasse F. Wolff Anthony, Benjamin Kanding, Raghavendra Selvan.** Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. *www.arxiv.org*. [En ligne] 6 Juillet 2020. <https://arxiv.org/pdf/2007.03051.pdf>.

63. **LeMagIT, Théo Alves Da Costa.** IA frugale : trois règles « simples » pour optimiser les algorithmes. *www.lemagit.fr*. [En ligne] 15 Décembre 2021. <https://www.lemagit.fr/conseil/IA-frugale-trois-regles-simples-pour-optimiser-les-algorithmes>.
64. **Capgemini Research Institute.** Sustainable IT. [En ligne] 05 Mars 2021. <https://www.capgemini.com/fr-fr/wp-content/uploads/sites/2/2021/04/Final-Web-Report-Sustainable-IT.pdf>.
65. **Radjou, Navi et Prabhu, Jaideep.** *Le guide de l'innovation frugale - Les 6 principes clés pour faire mieux avec moins*. Paris : Diateino, 2019. ISBN : 978-2-35456-373-8.
66. **David Autissier, Kevin Johnson, Emily Metais-Wiersch.** *Du changement à la transformation - Stratégie et pilotage de la transformation*. Malakok : Dunod, 2018. ISBN 978-2-10-078089-1.
67. **INRIA.** Impacts environnementaux du numérique. *FUN MOOC*. [En ligne] <https://www.fun-mooc.fr/fr/cours/impacts-environnementaux-du-numerique/>.
68. **Institut du Numérique Responsable.** MOOC Sensibilisation Numérique Responsable . *institutnr.org*. [En ligne] <https://institutnr.org/mooc-sensibilisation-numerique-responsable>.
69. —. MOOC Numérique Responsable (Complet). *institutnr.org*. [En ligne] <https://institutnr.org/mooc-numerique-responsable-complet>.
70. **David Autissier, Jean-Michel Moutot.** *Méthode de conduite du changement - Diagnostic, Accompagnement, Performance*. Malakoff : Dunod, 2016. ISBN 978-2-10-075401-4.
71. **Jean-Noël Felli, Patrick Lenain.** *L'entreprise vraiment responsable - La raison d'être : un levier d'innovation et de performance*. Paris : Magnard-Vuibert, 2021.
72. **Des Enjeux & des Hommes.** La raison d'être de l'entreprise. *desenjeuxetdeshommes.com*. [En ligne] Avril 2020. <https://www.desenjeuxetdeshommes.com/guides/la-raison-detre-de-lentreprise-tirons-les-lecons-de-lexperience/>.
73. **CITEPA.** 6e rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat (Giec) 1er volume : les sciences physiques du changement climatique - Synthèse du Résumé à l'intention des décideurs du Giec. [En ligne] 16 Septembre 2021. https://www.citepa.org/wp-content/uploads/Citepa_2021_09_d01_INT_GIEC_AR6_Vol1_09082021_VF.pdf.

74. **Bon Pote.** Synthèse et analyse du nouveau rapport du GIEC. *Bon Pote*. [En ligne] 14 Février 2022. <https://bonpote.com/synthese-et-analyse-du-nouveau-rapport-du-giec/>.
75. **Naugès, Louis.** Frugalité Numérique : centres de calcul (Deuxième partie). *Louis Naugès*. [En ligne] 15 Janvier 2020. https://nauges.typepad.com/my_weblog/2020/01/frugalite-numerique-centres-de-calcul-deuxieme-partie.html.
76. **Greenpeace.** Clicking clean : who is winning the race to build a green internet? *ClickClean*. [En ligne] Janvier 2017. <http://www.clickclean.org/france/fr/>.
77. **Grandmontagne, Yves.** La question de l'eau dans le datacenter : vers un WUE ? *Datacenter Magazine*. [En ligne] 15 Avril 2020. <https://datacenter-magazine.fr/la-question-de-leau-dans-le-datacenter-vers-un-wue/>.
78. **Lawrence Berkeley National Laboratory, Arman Shehabi, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin.** United States Data Center Energy Usage Report. *Connaissance des énergies*. [En ligne] Juin 2016. https://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf.
79. **Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin.** Attention Is All You Need. *Arxiv*. [En ligne] 6 Décembre 2017. <https://arxiv.org/pdf/1706.03762.pdf>.
80. **Stewart, Matthew.** iny Machine Learning: The Next AI Revolution. *Towards Data Science*. [En ligne] 2 Octobre 2022. <https://towardsdatascience.com/tiny-machine-learning-the-next-ai-revolution-495c26463868>.
81. **Radjou, Navi.** Comment le jugaad peut reconstruire l'économie post Covid-19. *Harvard Business Review*. [En ligne] 25 Février 2021. <https://www.hbrfrance.fr/chroniques-experts/2021/07/37455-comment-le-jugaad-peut-reconstruire-leconomie-post-covid-19/>.
82. **INR (Institut du Numérique Responsable).** Académie NR. *INR*. [En ligne] <https://www.academie-nr.org/>.

Annexes

Annexe 1 - Rapport du GIEC

Le rapport du **groupe 1** se concentre sur 5 scénarios :

- **Scénario SSP1-1.9** : +1,5°C échéance 2100 – très forte baisse des émissions dès 2025 (neutralité carbone en 2050)
- **Scénario SSP1 – 2.6** : +2,0°C échéance 2100 – baisse continue des émissions après 2025 (neutralité carbone entre 2050 et 2100)
- **Scénario SSP2-4.5** : ± 3,0°C – pic des émissions vers 2030 (neutralité carbone non atteinte en 2100)
- **Scénario SSP3-7.0** : + 3,6°C – hausse forte des émissions
- **Scénario SSP5-8.5** : + 4,4°C – hausse très forte des émissions

Scénario	Court terme (2021-2040)		Moyen terme (2041-2060)		Long terme (2081-2100)	
	Meilleure estimation (°C)	Fourchette <i>très probable</i> (°C)	Meilleure estimation (°C)	Fourchette <i>très probable</i> (°C)	Meilleure estimation (°C)	Fourchette <i>très probable</i> (°C)
SSP1-1.9	1,5	1,2 à 1,7	1,6	1,2 à 2,0	1,4	1,0 à 1,8
SSP1-2.6	1,5	1,2 à 1,8	1,7	1,3 à 2,2	1,8	1,3 à 2,4
SSP2-4.5	1,5	1,2 à 1,8	2,0	1,6 à 2,5	2,7	2,1 à 3,5
SSP3-7.0	1,5	1,2 à 1,8	2,1	1,7 à 2,6	3,6	2,8 à 4,6
SSP5-8.5	1,6	1,2 à 1,9	2,4	1,9 à 3,0	4,4	3,3 à 5,7

Tableau 9 – Évolution de la température mondiale à la surface de la Terre pour trois périodes de 20 au regard des cinq scénarios SSP étudiés par le Giec (73)

Sur un scénario intermédiaire d'émission de CO₂, le rapport indique une forte probabilité que le seuil de +2°C de réchauffement soit dépassé au cours du 21^{ème} siècle. Pour stabiliser la hausse de la température mondiale à la surface de la Terre, il faudrait atteindre au niveau mondial, zéro émission nette de CO₂. 1000 Gt d'émissions cumulées de CO₂ conduirait à une hausse de la température moyenne à la surface de la Terre de +0,45°C.

Comme l'indique la Figure 17(74) 2390 Gt de CO₂ ont été produites par l'Homme sur la période 1850-2019. Le budget carbone restant pour limiter le réchauffement à +1,5°C serait de 500 Gt CO₂. Sur le rythme actuel, ce budget serait dépassé en 2033. La limitation du réchauffement à +2°C amènerait ce budget carbone restant à 1350 Gt CO₂ mais, sur le même rythme, celui-ci serait atteint en 2058.

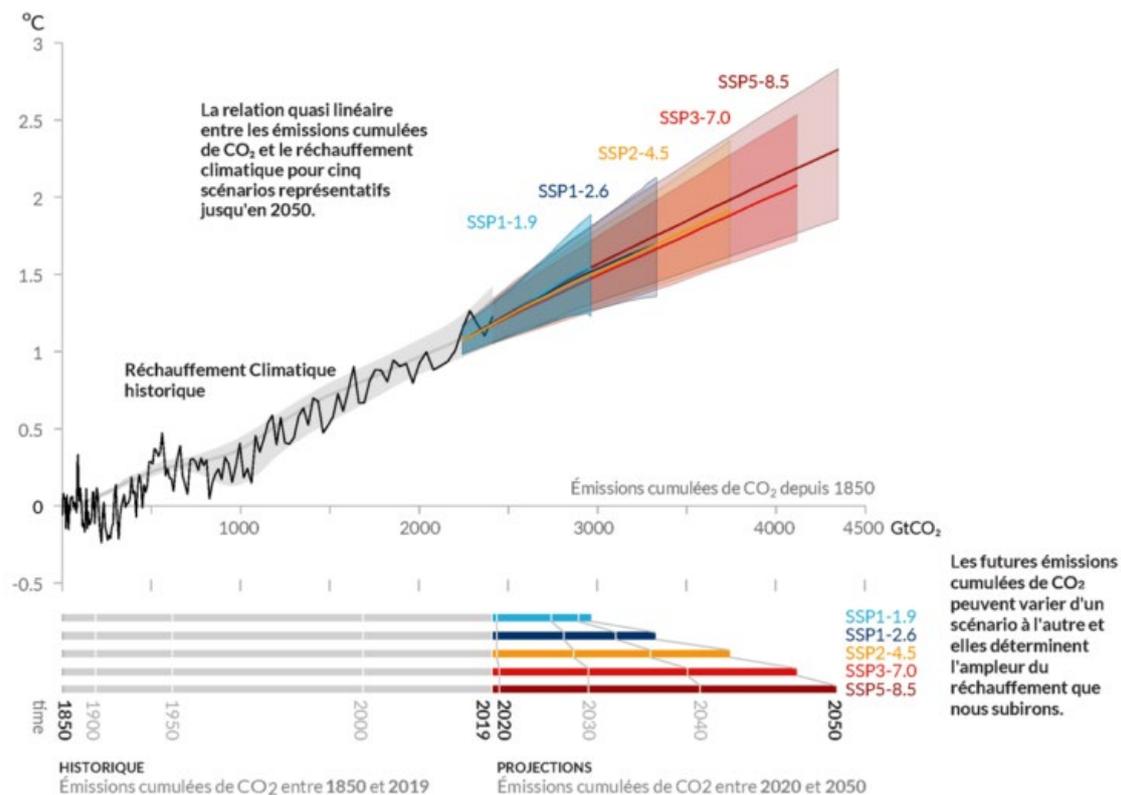


Figure 17 – Augmentation de la température à la surface du globe depuis 1850-1900 en fonction des émissions cumulées de CO₂ (GtCO₂) (74)

A retenir du rapport du groupe 1 du GIEC

- Le seuil de 1,5°C devrait être dépassé entre 2021 et 2040 et le réchauffement se poursuivra au moins jusqu'en 2050.
- Pour éviter un réchauffement de 2°C, il est indispensable de réduire immédiatement et fortement les émissions de GES mais également de façon durable et généralisée. Pour atteindre cet objectif, d'ici à 2100, les pays industrialisés doivent diviser leurs émissions de gaz à effet de serre par quatre.
- Avec le réchauffement climatique, les événements météorologiques verront leur fréquence et leur intensité augmenter.
- La montée du niveau des mers ou la fonte des glaciers témoignent désormais des dégâts irréversibles causés par le réchauffement climatique.
- Avec l'augmentation de CO₂, les océans et les terres seront de moins en moins capables d'en absorber.

Le rapport du **groupe 2** porte sur **les impacts, les vulnérabilités et les capacités d'adaptation à la crise climatique**. Celui-ci met en évidence les impacts déjà observables du réchauffement climatique sur les écosystèmes et les sociétés humaines. Les événements météorologiques extrêmes se font plus fréquents et altèrent certaines régions provoquant des

baisses de rendements agricoles et des pertes de biodiversités pouvant être définitive. L'intégrité de la ressource en eau potable et des ressources alimentaires devrait être de plus en plus menacée, notamment en Afrique, Amérique Latine, et Asie du Sud. Les impacts du changement climatique concernent également le bien-être et la santé des populations. Une partie du rapport est consacré à des mesures d'adaptation pour renforcer la résilience des écosystèmes et diminuer ainsi la vulnérabilité des personnes et des biens aux risques. La végétalisation des zones urbaines permet par exemple de limiter les températures en zone urbaine avec un bénéfice induit sur les émissions de gaz à effet de serre grâce à une plus faible utilisation de climatisation. L'adaptation des écosystèmes et des sociétés humaines réclamera néanmoins un engagement politique fort ainsi que des moyens financiers importants.

A retenir du rapport du groupe 2 du GIEC

- 3,6 milliards de personnes vivent des zones hautement vulnérables au changement climatique
- Dans le cas d'un réchauffement de 2°C, jusqu'à 18% des espèces terrestres auront un risque élevé d'extinction
- 8% des terres cultivables seront inadapté à la culture en cas de hausse de 1,6°C
- 8 à 80 millions de personnes pourraient souffrir de la faim d'ici 2050
- Entre 500 millions à 1 milliard de personnes vivent dans des zones côtières à risque d'ici 2050
- 127 milliards de dollars par an seront nécessaires pour d'adapter au changement climatique d'ici 2030

Le rapport du **groupe 3** traite plus particulièrement de **la lutte contre le réchauffement climatique**. L'accent est mis sur la nécessité d'agir au plus vite sur la réduction des gaz à effet de serre car ceux-ci n'ont cessé d'augmenter ces vingt dernière années (+12% de plus qu'en 2010, +54% de plus qu'en 2000). Si rien n'est fait l'augmentation de la température de la planète devrait atteindre +3,2°C et entraîner des catastrophes écologiques, économiques, sanitaires et sociales. Pour limiter le réchauffement sous des valeurs seuil de +1,5°C à 2°C, les émissions doivent être réduites en valeur absolue dès aujourd'hui et au plus tard dans les 2 à 3 prochaines années.

Les rédacteurs en appellent donc à un changement drastique de nos modes productions et de consommation. Le concept de **sobriété** apparaît alors comme la réponse principale à la problématique d'émission de gaz à effet de serre. Responsables de la plus grande part des émissions de CO₂, les villes mais également l'industrie présentent de fortes potentialités pour la transition climatique.

Profondément ancré dans les énergies fossiles, notre système énergétique devra basculer vers des énergies bas carbone (nucléaire, énergies renouvelables) pour déployer l'électrification. Pour atteindre la neutralité carbone, le rapport du GIEC souligne également la nécessité d'avoir recours à des techniques de stockage comme piste complémentaire.

Bien que les investissements et les efforts à consentir pour lutter contre le réchauffement climatique demeurent importants, les auteurs indiquent que ceux-ci seront bien **inférieurs aux coûts de l'inaction**.

A retenir du rapport du groupe 3 du GIEC

- La situation est plus grave qu'estimée précédemment. Il reste moins de 3 ans pour agir et rester sous les seuils de +1,5°C à 2°C.
- 2010-2019 représente la décennie où les émissions de gaz à effet de serre ont atteint le plus haut niveau jamais atteint (56 GtCO₂eq par an en moyenne)
- Pour limiter le réchauffement climatique à 1,5 à 2°C, il faudra réduire respectivement de 43% à 17% les émissions par rapport à 2019 d'ici à 2030.
- La neutralité carbone devra être atteinte au plus tard en 2050 pour endiguer la dérive climatique.

Annexe 2 - Les Objectifs de Développement Durable (ODD)



Figure 18 - Objectifs de Développement Durable (ODD)

On retrouve sur la Figure 18 et ci-dessous la liste des 17 objectifs :

1. L'éradication de la pauvreté
2. La lutte contre la faim
3. La santé et le bien-être des populations et des travailleurs
4. L'accès à une éducation de qualité
5. L'égalité entre les sexes
6. L'accès à l'eau salubre et l'assainissement
7. L'accès à une énergie propre et d'un coût abordable
8. Le travail décent et la croissance économique
9. La promotion de l'innovation et des infrastructures durables
10. La réduction des inégalités
11. La création de villes et de communautés durables
12. La production et la consommation responsable
13. La lutte contre le changement climatique
14. La protection de la faune et de la flore aquatiques
15. La protection de la faune et de la flore terrestres
16. La paix, la justice et des institutions efficaces
17. Le renforcement des partenariats pour les objectifs mondiaux

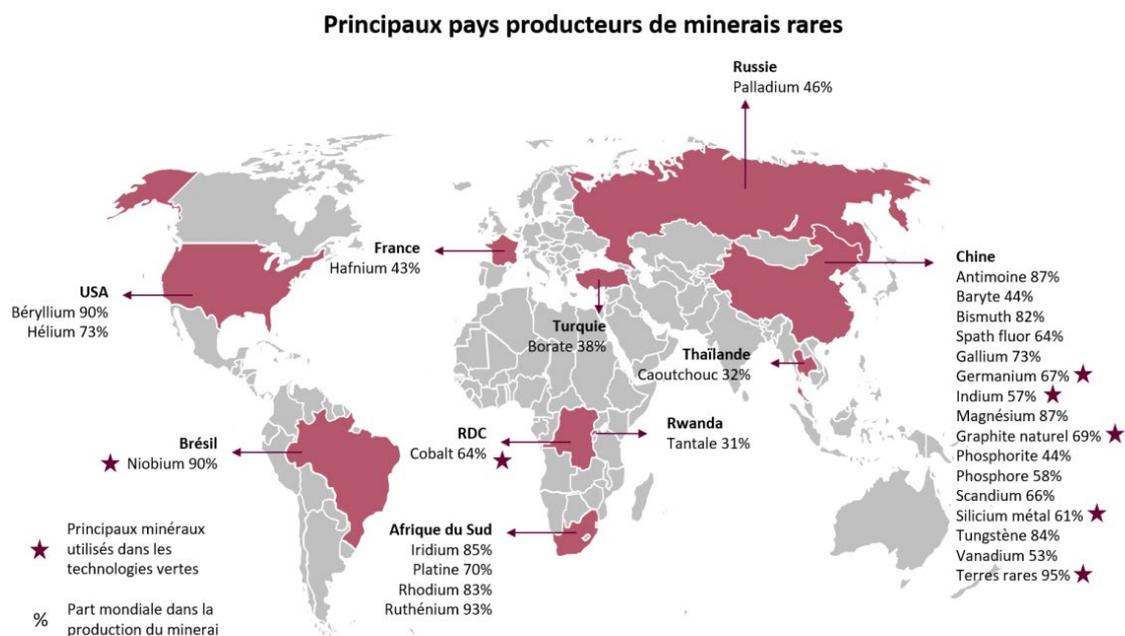
Chaque acteur (États, entreprises, investisseurs, ONG, citoyens, collectivités locales, écoles, ...) sont appelés à se mobiliser sur les ODD dans leurs domaines de compétences.

Bien qu'adoptés en 2015, l'agenda 2030 et les ODD ne sont pas sur la bonne voie pour être atteints. En 2020, la crise sanitaire, sociale et a impacté fortement plusieurs ODD. (7)

A retenir des objectifs de l'ONU sur le développement durable (ODD)

- Le texte de l'Agenda 2030 appelle explicitement les entreprises à contribuer aux ODD.
- Les entreprises sont incitées à intégrer les ODD dans leur stratégie de responsabilités sociétale (RSE) et plus largement dans leur stratégie.
- Les entreprises peuvent avoir un impact négatif sur certains ODD comme la lutte contre le changement climatique (émissions de gaz à effet de serre), le respect de la biodiversité (déforestations, déchets, ...), le travail décent (recours au travail des enfants).

Annexe 3 - Principaux pays producteurs de minerais rares



Source : Analyse Sia Partners d'après « La guerre des métaux rares – la face cachée de la transition énergétique et numérique » de Guillaume Pitron

Figure 19 - Principaux pays producteurs de minerais rares (Source Sia Partners)

Annexe 4 - Engagements des GAFAM pour l'environnement

Entreprise	Cible	Année	Engagements
Google	Zéro émission nette	2030	Zéro émission carbone d'ici 2030 Neutre en carbone depuis 2007 Consommation électrique compensée avec des énergies renouvelables depuis 2017 Héritage carbone intégralement éliminé par l'achat de compensations carbone
Apple	Neutralité carbone	2030	Conception de ses produits pour réduire son empreinte carbone. Gagner en efficacité énergétique (magasins, bureaux, data centers, sites de fabrication). Utiliser des énergies renouvelables. Éviter les émissions directes. Investir dans des solutions naturelles.
Facebook	Neutralité carbone	2030	Neutralité carbone pour toute sa chaîne de valeur, y compris les émissions de ses fournisseurs et les déplacements de ses employés. Activités alimentées à 100% par des énergies renouvelables.
Amazon	Neutralité carbone	2040	Neutralité carbone pour l'ensemble de ses activités d'ici 2040 Utiliser 100% d'énergies renouvelables d'ici 2025 Déployer 100 000 véhicules de livraisons électriques d'ici 2030 Investir 2 milliards de dollars en soutien au développement de technologies et de services de décarbonation (Climate Pledge Fund) Investir 100 millions de dollars dans des projets de reforestation et des solutions d'atténuation du changement climatique
Microsoft	Carbon negative	2030	Technologie d'élimination du carbone (TEN) Compenser d'ici 2050 tout le carbone émis par ses opérations depuis sa fondation en 1975 Compenser sa consommation d'électricité annuelle avec des énergies renouvelables Se passer du diesel pour sa flotte automobile d'ici 2030 Taxe carbone interne étendue au Scope 3

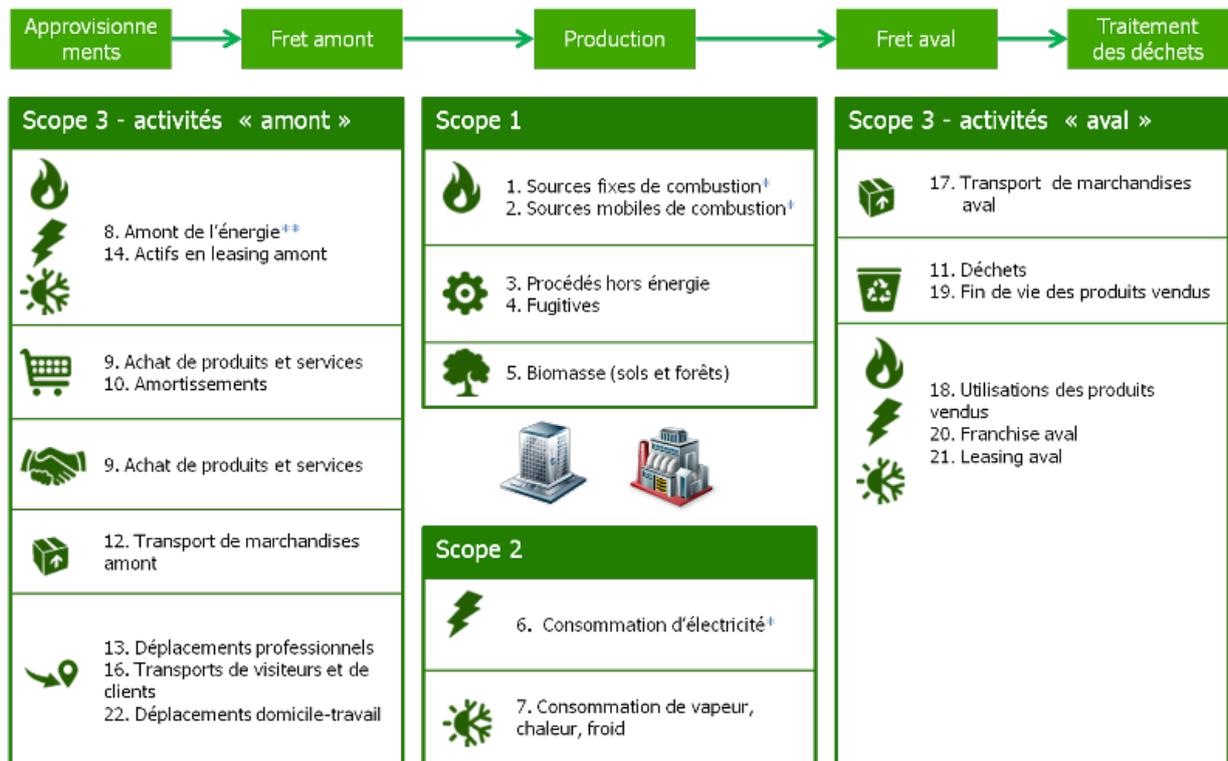
Tableau 10 - Engagements des GAFAM pour l'environnement

Annexe 5 - Technologies d'émissions négatives (TEN)

	Technologie	Principe	Avantage	Inconvénient
Processus Biologique	Afforestation / reforestation	Ajout d'arbres supplémentaires et replantation des forêts disparues	Meilleure qualité des sols Coût peu élevé	Perte de biodiversité Diminution de l'albédo (part des rayonnements solaires renvoyés vers l'atmosphère)
	Séquestration de carbone dans les sols	Restauration de pratiques d'utilisation des terres	Meilleure qualité des sols Préservation de la sécurité alimentaire Coût peu élevé	Émission de protoxyde d'azote (GES)
Processus Technologique	Bioénergie associée au captage et stockage du carbone BECCS (Bio-Energy with Carbon Capture and Storage)	Utilisation de la biomasse pour la production de chaleur. Séparation, captage et stockage du CO ₂	Production d'énergie décarbonée	Perte de biodiversité Menace pour la sécurité alimentaire Pollution de l'air Coût élevé Émission de protoxyde d'azote (GES)
	Captage direct du CO ₂ de l'air avant stockage DACCS (Direct Air Capture with Carbon Storage, «DACCS»).	Procédés d'extraction de CO ₂ présent dans l'air ambiant.		Coût très élevé
Processus Géochimique	Alcalinisation et fertilisation de l'océan	Ensemencer les océans avec du sulfate de fer pour doper le développement du phytoplancton qui capture le CO ₂		Pollution par l'extraction des minerais Pollution des eaux profondes

Tableau 11 – Technologies d'émissions négatives (TEN)

Annexe 6 – Cartographie des scopes 1, 2 et 3



[†] Utiliser uniquement la part combustion des facteurs d'émissions

^{†*} Utiliser uniquement la part amont des facteurs d'émissions

Figure 20 - Cartographie des scopes 1, 2 et 3 (source : ADEME – Bilan GES)

Annexe 7 - Clean Energy Index par opérateur de centres de données

Entreprises	Note globale	 Clean Energy Index	 Gaz naturel	 Charbon	 Nucléaire
		24%	3%	67%	3%
		17%	24%	30%	26%
		83%	4%	5%	5%
		24%	3%	67%	3%
		67%	7%	15%	9%
		56%	14%	15%	10%
		29%	29%	27%	15%
		32%	23%	31%	10%
		24%	3%	67%	3%

Tableau 12 - Clean Energy Index par opérateur de centres de données (Source Greenpeace)

Annexe 8 - Indicateurs de performances des data centers

Indicateurs de consommation d'énergie

PUE (Power Usage Effectiveness)

Très utilisé, l'indicateur PUE (normalisé ISO/IEC 30134-2 en 2016) caractérise l'efficacité énergétique d'un centre de calcul. Il s'exprime de la façon suivante :

$$PUE = \frac{E_{DC}}{E_{IT}} = \frac{\text{Energie consommée par le datacenter}}{\text{Energie consommée par les systèmes informatiques}}$$

Équation 1 - PUE (Power Usage Effectiveness)

Une valeur de proche de 1 est bien sûr optimum car l'essentiel de l'énergie est alors utilisé par les serveurs. Entre 2007 et 2020, le PUE des meilleurs data centers privés dans le monde est passé de 2,5 de **1,59**. Le Tableau 13(75) permet de comparer les performances des data centers suivant leur typologie. L'avantage est ici clairement en faveur des clouds publics qui réduisent significativement leurs émissions carbone pour un même usage.

Type de data center	PUE
Cloud publics	1,11
Cloud privés	1,59
Centres de calcul des entreprises grandes et moyennes	> 2
PME ou TPE	> 3 ou 4

Tableau 13 - PUE moyen par type de data center (75)

DCEM (Data Center Energy Management)

Loin de faire consensus sur sa représentativité en termes d'impact environnemental, le PUE est un indicateur développé en Amérique du Nord. Le DCEM est quant à lui un indicateur français conçu par le CRIP (Club des Responsables d'Infrastructure et de Production), il tient compte de la consommation d'énergie globale, du rendement de l'infrastructure informatique, de la part d'énergie réutilisée pour d'autres usages mais également la part d'énergie renouvelable employée. Plus complexe, il n'est encore peu utilisé et peine à s'imposer par manque de soutien de l'Europe.

ERE (Energy Reuse Effectiveness) et ERF (Energy Reuse Factor)

Lorsque le data center est conçu en conséquence, les indicateurs ERE (Energy Reuse Effectiveness) et ERF (Energy Reuse Factor) permettent de tenir compte de la quantité d'énergie réutilisée sous forme de chaleur ou d'électricité.

$$ERE = \frac{E_{DC} - E_{REUSE}}{E_{IT}} = \frac{\text{Energie consommée par le data center} - \text{Energie réutilisée}}{\text{Energie consommée par les systèmes informatiques}}$$

Équation 2 - ERE (Energy Reuse Effectiveness)

$$ERF = \frac{E_{REUSE}}{E_{DC}} = \frac{\text{Energie réutilisée}}{\text{Energie consommée par le data center}}$$

Équation 3 – ERF (Energy Reuse Factor)

Si aucune énergie est réutilisée, l'indicateur ERE est équivalent au PUE et l'indicateur ERF vaut 0.

Indicateurs environnementaux

REF (Renewable Energy Factor)

L'indicateur REF caractérise la part d'énergie renouvelable sur l'énergie totale consommée par le data center.

$$REF = \frac{E_{RENEW}}{E_{DC}} = \frac{\text{Energie renouvelable consommée}}{\text{Energie consommée par le data center}}$$

Équation 4 - REF (Renewable Energy Factor)

CUE (Carbon Usage Effectiveness)

Le CUE considère les émissions totales de GES (exprimée en kg de CO₂) par rapport à l'énergie consommée par les systèmes informatiques (kWh).

$$CUE = \left[\frac{kgCO_2}{kWh} \right] = \frac{\text{Emissions totales de CO}_2}{\text{Energie consommée par les systèmes informatiques}}$$

Équation 5 - CUE (Carbon Usage Effectiveness)

Idéalement, si le data center est alimenté avec une électricité 100% renouvelable, le CUE vaut 0. Pour calculer le CUE qui caractérise finalement l'impact du data center sur l'environnement, il faut connaître le mix énergétique (charbon, nucléaire, gaz, éolien, ...) de son alimentation en électricité. Un rapport de Greenpeace USA sorti en 2017 (76) définit un indice d'énergie propre (Clean Energy index) et évalue ainsi les principaux opérateurs de centres de données. Le tableau en Annexe 7 en synthétise quelques résultats. On mesure bien ici l'intérêt de s'intéresser au mix énergétique d'un centre de données.

WUE (Water Usage Effectiveness)

Moins utilisé, l'indicateur WUE permet néanmoins de mesurer l'efficacité de l'utilisation de l'eau par un data center et d'évaluer ainsi son impact sur l'environnement.

$$WUE \left[\frac{l}{kWh} \right] = \frac{\text{Eau consommée par le datacenter}}{\text{Energie consommée par les systèmes informatiques}}$$

Équation 6 – WUE (Water Usage Effectiveness)

L'ordre de grandeur pour un data center moyen est de 1,8 litre pour 1 kWh. Celui-ci descendrait à 0,2 l/kWh pour un data center performant (77).

Production de déchets

Généralement les serveurs sont remplacés tous les 5 ans. Il convient donc de prendre en compte la quantité de déchets produits par le data center et d'évaluer les possibilités de réemploi ou de recyclage.

Modèle d'analyse énergétique des centres de calcul

Dans son blog (75), Louis Naugès propose un modèle d'analyse énergétique des centres de calcul reposant sur l'objectif d'une consommation minimum d'énergie pour l'exécution d'une application (principe de frugalité numérique)

Son approche repose sur le Taux d'Utilisation des Serveurs (TUS), le PUE, le type d'énergie utilisée (renouvelable ou carbonée) et l'énergie totale consommée. En s'appuyant sur une étude de Berkeley (78), l'auteur rappelle qu'un Cloud public de type hyperscale (TUS = 0,50) est 3,33 plus performant qu'un serveur d'entreprise (TUS = 0,15).

Afin de caractériser l'efficacité énergétique d'un data center pour activer une application, il introduit le Ratio Efficacité Énergétique (R2E), défini comme suit :

$$R2E = \frac{100 \times PUE}{TUS}$$

Équation 7 - R2E (Ratio Efficacité Énergétique)

Le Tableau 14 présente les valeurs de ce ratio ainsi que le % d'énergie électrique utile suivant le type de data center. Sur la base de cette approche, un data center d'entreprise « gaspillerait » donc environ 92% de l'énergie (pour 55% dans le cas d'un data center hyperscale).

Type de data center	PUE	TUS	R2E	% d'énergie électrique utile	% d'énergie « gaspillée »
Data center « parfait »	1	1	100	100%	0%
Data center hyperscale	1,1	0,50	220	45%	55%
Data center d'entreprise	2	0,15	1333	8%	92%

Tableau 14 - R2E et % d'énergie utile par type de data center

On mesure bien ici l'importance de s'intéresser à la part d'énergie renouvelable utilisée par les fournisseurs de solutions Cloud puisque qu'une bonne partie de l'énergie n'est pas utilisée pour l'exécution d'une application.

Annexe 9 - Modèles LSTM, GRU et Transformer

• Réseaux LSTM (Long Short-Term Memory) et GRU (Gated Recurrent Unit)

Plus complexes que les réseaux RNN, les réseaux LSTM et GRU permettent de gérer la mémoire à court et long terme grâce à un système de portes. Outre la reconnaissance de texte (pouvant être manuscrit) et de la parole, ils permettent aussi la traduction linguistique, la génération de texte et de musique. La cellule (neurone) dans les couches cachées d'un LSTM comporte trois portes : porte de l'oubli (Forget gate), porte d'entrée (Input gate) et porte de sortie (Output gate) et à chaque porte est associé un poids. En fonction des valeurs de ces derniers et de la valeur de la fonction d'activation obtenue, la porte d'oubli décide ou non de conserver l'information déjà présente (un mot par exemple) dans la mémoire. De la même façon, la porte d'entrée décide quelle nouvelle information est à stocker. A partir de ces informations, l'état de la cellule (Cell state) est actualisé ce qui permet enfin à la porte de sortie de définir un état caché (Hidden state) nécessaire aux prochaines prédictions.

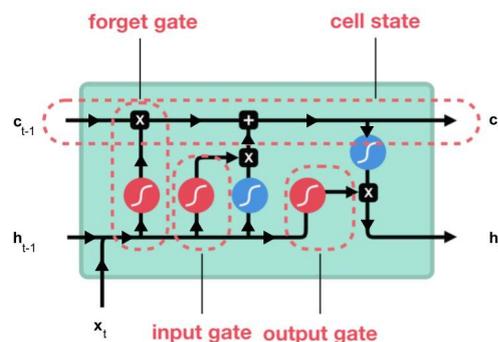


Figure 21 - Cellule LSTM (Source : Pensée artificielle)

Les réseaux GRU sont une variante des LSTM et requièrent moins de paramètres. Une cellule GRU comporte 2 portes : une porte de reset (Reset gate) et une porte de mise à jour (Update gate). La porte de mise à jour rassemble les portes d'entrée et d'oubli des LSTM et la porte de reset constitue la porte de sortie pour définir l'état caché. Devenus très populaires, ces deux modèles ont besoin de beaucoup de ressources (notamment de bande passante mémoire) et de temps pour être entraînés.

• Transformer

Comme les LSTM, le modèle Transformer est un modèle à séquences (seq2seq) qui prend une séquence en entrée et renvoie une séquence en sortie (la traduction d'un texte par exemple). Dans les LSTM, l'interdépendance des mots est prise en compte par une architecture « Encoder-Decoder » traitant la séquence de mots par une représentation vectorielle. Dans le Transformer, l'interdépendance des mots est conservée mais s'appuie sur

un mécanisme d'attention permettant de considérer à la fois les séquences d'entrée et de sortie prédites. Ce mécanisme est décrit dans un article intitulé « Attention Is All You Need » publié en 2017 par Google Brain et Google Research (79). Les performances du Transformer se sont montrées impressionnantes sur plusieurs jeux de traduction. Il existe de nombreux modèles pré-entraînés, parmi les plus populaires, on retrouve BERT (Google, 2019), GTP-3 (OpenAI, 2019), RoBERTa, CamenBERT, FlauBERT, Text-to-Text Transfer Transformer (T5), Turing Natural Language Generation (T-NLG, Microsoft, 2020), Megatron-Turing Natural Language Generation (MT-NLG).

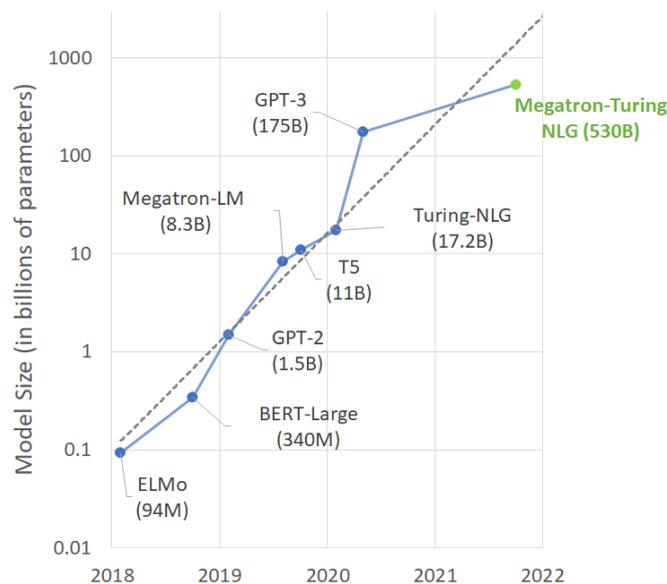
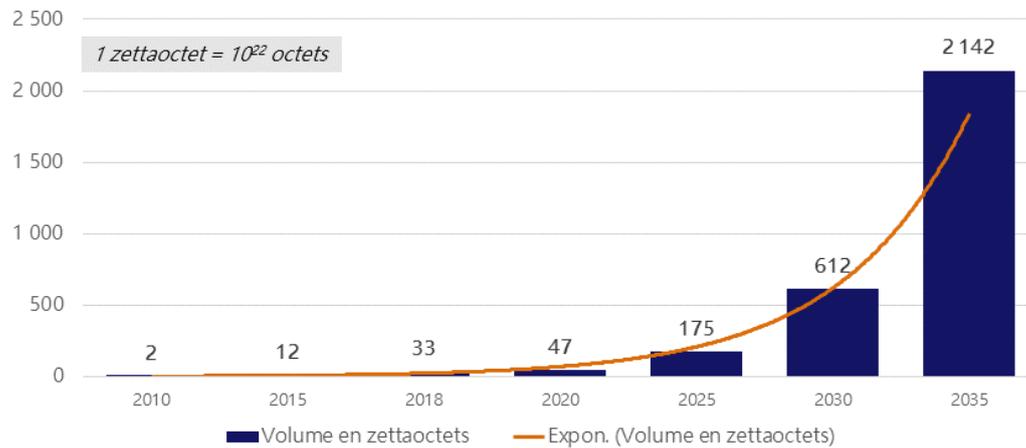


Figure 22 - Evolution du nombre de paramètres pour les modèles Transformer (Source: ysance)

Comme on peut le constater sur la Figure 22, le nombre de paramètres des modèles a augmenté de manière significative ces dernières années. Le modèle MT-NLG avec 530 milliards de paramètres en comporte trois fois plus que le modèle GPT-3 (175 milliards). Si la précision s'améliore, l'entraînement de ces modèles est souvent difficile pour des raisons de capacité mémoire et du très grand nombre d'opérations nécessaires. Il devient alors nécessaire de chercher de voies d'optimisation des algorithmes et des matériels.

Annexe 10 – Évolution du volume de données dans le monde

Evolution du volume de données numériques créés et consommés – Monde, 2010-2035



Source : IDC // Statista // JDN - Le volume de données mondial sera multiplié par 45 entre 2020 et 2035

© COPYRIGHT ADVAES – WWW.ADVAES.COM

Figure 23 - Évolution du volume de données dans le monde, 2010-2035 (Source : ADVAES)

Annexe 11 - Extrait du canevas Impact AI

 Thème	 Enjeux	 KPI emblématiques récurrents
Environnemental	Performance énergétique	Consommation énergétique (KWh)
	Émissions de gaz à effet de serre	Émissions de CO ₂ (TCO ₂) Objectifs et taux de réduction des émissions de Gaz à Effet de Serre (%)
	Adaptation au réchauffement climatique	%habitations en zone inondable, montant des dommages en €
	Consommation d' eau	Taux d'autosuffisance en eau dans les zones (%) Consommation d'eau (m3)
	Biodiversité et écosystèmes	% espèces en voie de disparition
	Maîtrise de la pollution	Émissions de NOx (tonnes), taux d'oxygène des cours d'eau,...
	Gestion des déchets	Taux de déchets recyclés (% des déchets) Quantité de déchets de chaque catégorie (tonnes)

Figure 24 - Extrait du canevas Impact AI

Annexe 12 - Article LinkedIn

L'IA, la botte secrète de l'environnement ?

- Publié le 18 mars 2022
- Vincent Six



Crédit photo : Photo créée par jcomp - freepik.com

Tous les experts s'accordent sur le fait que l'activité humaine influe fortement sur notre belle planète ! Le 6e rapport du GIEC indique une forte probabilité que le seuil de +2°C de réchauffement soit dépassé au cours du XXIe siècle. Et notre activité numérique représente plus de 4% des émissions de gaz à effet de serre mondiales et 10% de la consommation électrique. Ango, Fabien, Thierry et moi-même avons tenté de synthétiser le résultat de nos recherches sur la problématique spécifique de l'intelligence artificielle et de l'environnement. La première partie de cet article sera consacrée aux solutions concrètes apportées par l'IA dans ce domaine. Dans notre deuxième partie, nous nous intéresserons aux ressources nécessaires à l'IA... et qui, donc, impactent l'environnement.

Prêts à nous suivre dans notre analyse ? Allez, c'est parti !

I. L'IA au secours de notre planète

Avant toute chose, définissons l'IA :

On peut comparer l'IA à une forme d'intelligence attribuée aux machines. Il s'agit d'un outil leur permettant d'être autonomes et d'adopter des comportements humains, voire surhumains, grâce à l'utilisation d'algorithmes.

L'IA a la capacité d'apprendre et de comprendre des données ou informations qui lui sont soumises, de les mémoriser et d'utiliser l'expérience issue des actions précédentes pour s'adapter à de nouvelles situations. On entend souvent parler de Machine Learning (ML), qui est le modèle d'apprentissage le plus utilisé.

On la retrouve dans tous les secteurs de l'économie et dans tous les objets du quotidien (robots, voitures autonomes, assistants virtuels, logiciels, smartphones, etc.). Il est donc essentiel d'atteindre un juste équilibre entre éthique, protection de la vie privée et utilisation d'outils qui nous facilitent la vie et améliorent notre quotidien.

A ce titre, l'IA peut être un véritable levier dans la transition écologique en apportant des solutions concrètes aux problèmes générés par l'activité humaine. Nous allons donc nous intéresser aux solutions existantes et aux solutions de demain.

. Le bâtiment, domaine à fort potentiel de lutte contre le réchauffement climatique

Le secteur du bâtiment représente 44 % de l'énergie consommée en France, loin devant le secteur des transports (31,3%). Chaque année, le secteur du bâtiment émet plus de 123 millions de tonnes de CO₂, ce qui en fait l'un des domaines-cibles dans la lutte contre le réchauffement climatique et pour la transition énergétique.

L'IA peut apporter une aide précieuse. Exemple : une entreprise appelée EnerNOC a créé une application qui permet aux gestionnaires de bâtiments de réduire la consommation d'énergie en fournissant des données en temps réel et des recommandations sur la façon dont ils peuvent optimiser la consommation d'énergie d'une structure. Ce pilotage peut entraîner par exemple la baisse de luminosité, l'ajustement de la température, le changement de planification de lignes de production à d'autres moments dans la journée etc.

. La fibre optique et l'IA, une piste pour atteindre la neutralité carbone ?

Avec les réseaux à fibre optique, nous pouvons espérer consommer cinq fois moins d'électricité qu'avec les réseaux cuivre, avec un cycle de vie plus long. Cela est rendu possible par la gestion des capteurs qui permettent la réduction de notre consommation d'énergie, de chauffage ou d'eau. Cette technologie rendra nos logements et bureaux plus « intelligents » et réduira notre empreinte carbone.

. L'agriculture, de meilleures récoltes avec moins de ressources

Des drones dans les champs ! Ceux-ci assurent la surveillance des cultures et des sols. Les images collectées sont analysées par des algorithmes d'apprentissage automatique et des conseils sont alors fournis pour optimiser la lutte antiparasitaire en limitant ainsi l'usage de produits phytosanitaires. En complément des données déjà collectées, les images captées par des satellites et les données météorologiques alimentent des modèles d'IA pour une agriculture de précision en proposant un plan de culture personnalisé (type de culture, gestion de l'eau, plantation optimale). La robotique agricole rend aussi possible l'automatisation de la récolte et le contrôle des mauvaises herbes. Toutes ces applications accompagnent l'agriculteur dans sa prise de décision pour une gestion plus durable des ressources.

Le transport : en route pour l'optimisation des trajets

. Sightness, un système pour décarboner les transports

Avec 10 % des émissions totales de CO₂, le transport de marchandises a un impact sur l'environnement. Le logiciel Sightness automatise le traitement des informations afin d'obtenir

des données fiables et homogènes, et l'algorithme permet d'identifier les anomalies et de suggérer des optimisations efficaces. L'algorithme IA du logiciel permet de suivre l'évolution des émissions, mais aussi d'identifier les pistes d'optimisation et de recommander différents scénarios de plans d'action, en comparant par exemple les différentes routes possibles ou la fréquence des livraisons.

. Eodyn, un système qui optimise le transport maritime

Yann Guichoux (CEO d'Eodyn) s'intéresse aux télécommunications et à l'Automatic Identification System (AIS), qui équipe tous les navires de commerce et de transport depuis 2007. Ce système permet d'envoyer en temps réel la situation de n'importe quel bateau : vitesse, route, identifiants... Grâce à l'Intelligence artificielle développée par la start up bretonne, ils étudient ces informations après avoir ingéré les caractéristiques propres à chaque navire. Grâce aux quelques 100 000 navires qui parcourent simultanément la planète, la société dispose d'autant de points de mesures pour fournir en temps réel ces informations. S'il était installé sur tous les navires de commerce, le système permettrait d'économiser entre 2 à 3% de carburant et d'éviter l'émission de 37 mégatonnes de CO2 chaque année. Et celui-ci est en passe d'être amélioré en étudiant les possibilités d'étendre la technologie à la mesure de la houle et des vents.

. Le jumeau numérique, une technologie à venir au service de l'environnement

Est-ce que vous connaissez E2 ?

Alors non, rien à voir avec R2D2 de Stars Wars ! Il s'agit d'un jumeau numérique. Une belle avancée de NVIDIA* et de ses partenaires !

Surtout utilisé dans l'industrie 4.0, le jumeau numérique permet, avant même la conception d'une usine, de simuler son ergonomie, sa productivité et même sa consommation d'énergie.

Avec Earth 2 (E2), un projet de jumeau numérique de la Terre, nous serons capables de prédire le changement climatique en tirant parti de vitesses de calcul alimentées par l'intelligence artificielle. C'est juste impressionnant !

Autant dire que nous pourrions évaluer des hypothèses de développement émises par des gouvernements et anticiper nos actions sur notre planète. □

Vers 0% émissions de carbone nettes ?

Les émissions de carbone sont une préoccupation majeure en matière d'environnement et de changement climatique. De nombreux pays, industries et entreprises se sont fixé des objectifs ambitieux dans le domaine de la neutralité carbone et de la réduction des émissions, ce qui constitue un engagement fort envers le monde. Et l'industrie de la téléphonie mobile ne fait pas exception avec son avancement vers la 5G. Certaines industries ont reçu l'ordre de réduire

leurs émissions de carbone, comme Royal Dutch Shell, à qui un tribunal de La Haye a ordonné de diminuer ses émissions mondiales de carbone de 45 % d'ici à la fin de 2030 par rapport aux niveaux de 2019.

D'autres entreprises, comme BCG (<https://www.bcg.com/publications/2021/ai-to-reduce-carbon-emissions>), tentent de réduire les émissions de carbone et les coûts grâce à la puissance de l'IA. Elles utilisent une solution alimentée par l'IA pour aider les entreprises à mesurer avec précision, simuler, suivre et optimiser leurs émissions à l'échelle. D'après leurs hypothèses, l'IA peut permettre d'atteindre 5 % à 10 % de la réduction nécessaire, soit entre 2,6 et 5,3 gigatonnes de CO₂.

Mais toute médaille a son revers et l'IA pourrait aussi représenter l'une des activités les plus productrices de CO₂.

Nous avons souhaité étudier ce sujet afin d'y voir plus clair.

II. Le numérique (et donc en partie l'IA) est à l'origine de plus de 4% des émissions de gaz à effet de serre

Qu'en est-il des applications de l'IA ? Plusieurs rapports s'alarment de la progression de la consommation énergétique des applications de l'apprentissage profond.

Une petite définition :

Le deep learning ou apprentissage profond est un type d'intelligence artificielle dérivé du machine learning (apprentissage automatique) où la machine est capable d'apprendre par elle-même, contrairement à la programmation, où elle se contente d'exécuter à la lettre des règles prédéterminées.

Des chercheurs du MIT ont même estimé l'empreinte carbone des quatre modèles à la pointe de l'IA (GPT-2, BERT, ELMo, Transformer). Certaines technologies particulièrement complexes pourraient consommer l'équivalent de 316 allers-retours New York – San Francisco. Notamment dans les processus de Deep Learning, la création d'une IA efficace pourrait être coûteuse pour l'environnement. Près de 300 000 kilogrammes d'émissions équivalentes de dioxyde de carbone sont créés au cours du processus de formation d'un seul modèle.

L'IA consomme de l'énergie, car elle repose sur une infrastructure numérique qui, dans certains cas, impose beaucoup de données pour lui faire apprendre un comportement qu'elle sera en mesure de reproduire automatiquement par la suite. Pour cela, elle doit être "entraînée". Cet exercice peut nécessiter une grande puissance de traitement en fonction du nombre d'opérations à réaliser. Les algorithmes d'apprentissage automatiques basés sur des modèles mathématiques s'appuient sur des données d'échantillon, "données d'entraînement",

pour établir des prédictions ou prendre des décisions sans être explicitement programmés pour le faire. En d'autres termes, c'est grâce à l'entraînement qu'une application d'IA devient « intelligente ».

Vers une prise de conscience pour une IA raisonnée ?

Les chercheurs et ingénieurs du monde entier travaillent dans le domaine de l'optimisation des consommations énergétiques des solutions IA. L'objectif ? S'inspirer de l'intelligence humaine et de sa consommation (20W), et effectuer des millions d'opérations à la seconde, propres à l'intelligence de tous.

Les chercheurs travaillent sur de nombreuses pistes : de nouvelles puces écoconçues pour la forte demande de calcul des réseaux de neurones profonds ; mise à disposition en open source des modèles déjà entraînés. En effet, il ne sert à rien de refaire des millions de calculs pour des cas d'usages dans les entreprises s'ils sont déjà implémentés.

La course à l'IA verte est donc bel et bien lancée !

Conclusion :

Vous l'aurez compris, l'IA a des effets positifs et négatifs sur l'environnement. Car si elle s'annonce comme un véritable levier dans la transition écologique, il n'en demeure pas moins qu'elle consomme de l'énergie.

Avec la pression des accords de Paris, qui imposent aux Etats-membres de réduire leur empreinte carbone de 55% d'ici 2030, l'utilisation de l'intelligence artificielle et la mobilisation de l'intelligence collective des chercheurs du monde entier est indispensable. Puces écoconçues, mise à disposition des modèles... les pistes d'optimisation de cet outil fantastique ne manquent pas ! Certains évènements nous montrent que notre modèle de consommation est fragile et pas éternel. L'innovation de l'IA au service de l'environnement, basée sur l'éthique et les valeurs, peut participer, à n'en pas douter, à la résolution du dilemme séculaire de l'homme : comment vivre en équilibre avec l'environnement.

Il y va de notre responsabilité collective envers les prochaines générations et nous devons travailler sans relâche, chercheurs, industriels, politiques, citoyens pour leur transmettre notre unique maison, la Terre.

[Bwele lobe Belibi](#) (Ango), [Fabien Omhover](#), [Thierry OLIVIER](#) et [Vincent Six](#) | MBA MIA | Insitut Léonard de Vinci

* www.nvidia.com

Annexe 13 - Comparaison des grands modèles d'IA

Entreprise	Pays	Modèle	Nombre de paramètres (en milliards)
OpenAI	USA	GPT-1	0,11
OpenAI	USA	GPT-2	1,5
OpenAI	USA	GPT-3	175
AI21 Labs	USA	Jurassic-1	178
DeepMind	USA	Gopher	280
Microsoft	USA	Megatron-Turing NLG	530
Google	USA	BERT	0,11
Google	USA	Switch Transformer	1000
Google	USA	GLaM	1200
Huawei	Chine	Pangu	200
Inspur	Chine	Yuan 1.0	245
Baidu et Peng Cheng Laboratory	Chine	PCL-BAIDU Wenxin	280
Beijing Academy of AI	Chine	Wu Dao 2.0	1750
Naver	Corée du sud	HyperCLOVA	204

Tableau 15 - Comparaison des grands modèles d'IA en fonction du nombre de paramètres

Annexe 14 - Les outils d'évaluation des émissions carbone

Les outils des fournisseurs cloud

Google

Disponible depuis février 2021, l'outil **Carbon Footprint** permet aux clients de Google de connaître leurs émissions de carbone liées à l'utilisation du cloud. Depuis février 2022, il fait partie de la suite **Carbon Sense** disposant dorénavant d'une fonction (**Active Assist**) basée sur l'IA pour repérer et désactiver les ressources inutilisées par les projets. Google a constaté que 600 000 kCO₂eq étaient émis inutilement par des machines virtuelles inutilisées.

Microsoft

Disponible en téléchargement depuis juillet 2021, l'application **Emissions Impact Dashboard** de Microsoft fournit à ses clients un tableau de bord pour analyser en temps réel leurs émissions de gaz à effet de serre associées à leur utilisation du cloud Azure. La méthode validée par l'Université de Stanford en 2018 s'aligne sur les normes ISO de mesure des émissions GES et inclut les scope 1, 2 et 3 de Microsoft.

AWS

En mars 2022, AWS a lancé un outil gratuit pour ses clients afin de suivre l'empreinte carbone de leur consommation de services en cloud. L'outil permet également la comparaison avec l'empreinte estimée d'un centre de données sur site. Le rapport d'analyse est disponible via la console de facturation AWS et présente les émissions par zone géographique et par service. Conformément au protocole sur les gaz à effet de serre et aux normes ISO, toutes les émissions (directes et indirectes) sont exprimées en tonnes métriques d'équivalents CO₂ (MTCO₂e). Par le passage en cloud, AWS annonce pouvoir réduire jusqu'à 88% l'empreinte carbone par rapport aux centres de données entreprises.

Les outils en ligne ou en téléchargement

En ligne ou en téléchargement, plusieurs outils permettent de fournir des indications sur la consommation d'électricité et l'empreinte carbone associées à l'exécution d'un programme d'IA :

CodeCarbon

Développé par l'institut de recherche canadien Mila dont le fondateur est Yoshua Bengio (Prix Turing 2018), le logiciel open source CodeCarbon (<https://codecarbon.io>) permet d'évaluer l'empreinte carbone des algorithmes d'IA. L'estimation de l'énergie consommée et donc des émissions de CO₂ associées à l'entraînement d'un modèle s'appuie l'énergie consommée par le serveur et le type d'énergie utilisée (énergies fossiles, éoliens, nucléaire). Présentés sous

forme de tableaux de bord, les résultats obtenus permettent de comparer les émissions de l'entraînement d'un modèle à l'équivalent en nombre de kilomètres parcourus en voiture, en journées d'activité d'un ménage moyen ou en heures de télévisions visionnées. L'outil recommande également des régions de calcul à faible intensité de carbone pour les principaux fournisseurs de cloud (AWS, Azure et GCP).

ML CO2 Impact

S'appuyant sur les données d'entrée telles que le type de matériel, la durée d'exécution, le fournisseur cloud et sa localisation, ML CO2 Impact est un outil en ligne (<https://mlco2.github.io/impact/#home>) qui permet d'estimer l'empreinte carbone. Ce calculateur fournit deux chiffres : les émissions de carbone brutes produites et les émissions de carbones compensées estimatives. Ces résultats exprimés en kg de CO₂eq sont comparés au nombre équivalent de kilomètres parcourus par une voiture moyenne, à la masse de équivalent de charbon brûlé et au nombre d'arbres nécessaires pour sequestrer le CO₂ sur une période de 10 ans.

Green Algorithms

Reprenant les principes des précédents outils, Green Algorithms permet de saisir en ligne (<http://www.green-algorithms.org>) également des informations telles que la taille de la mémoire disponible, le nombre de cœurs. Il est également possible de tenir compte du fait que le modèle est exécuté plusieurs fois pour le réglage des hyperparamètres par la sélection d'un « pragmatic scaling factor »

Carbon Tracker

Créé par deux étudiants de l'université de Copenhague (Lasse F. Wolff Anthony et Benjamin Kanging), la librairie python Carbontracker (<https://github.com/lfwa/carbontracker>) prédit le temps total de formation et l'empreinte carbone d'un modèle à partir des informations GPU, CPU et DRAM.

Experiment impact tracker

Disponible sur Github (<https://github.com/Breakend/experiment-impact-tracker>), la librairie python Experiment impact tracker permet de également de suivre la consommation d'énergie, les émissions de carbone et l'utilisation des capacités d'un système de calcul.

Annexe 15 - Tiny ML

Le TinyML désigne une approche pour réduire la taille des modèles et des applications de Machine Learning afin de les faire fonctionner sur des processeurs ou des puces d'IA à faible coût.

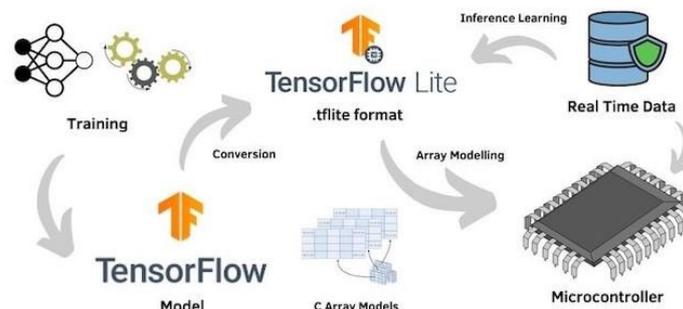


Figure 25 – Flux de travail TensorFlow Lite Lite (Source : makeueof.com)

L'intérêt du TinyML pour les industriels concerne le Edge Computing et les applications informatiques économes en énergie. Initialement, le concept de l'internet des objets repose sur le fait que les données d'un appareil local sont envoyées sur le cloud pour y être traitées. Cela pose néanmoins question vis-à-vis de la confidentialité des données, de la latence, du stockage et de l'efficacité énergétique. Le tableau ci-dessous reprend les arguments en faveur du TinyML développés dans un article publié sur le site Towards Data Science (80)

Efficacité énergétique	La transmission des données (par fil ou sans fil) est plus énergivore que les calculs embarqués d'environ un ordre de grandeur.
Confidentialité	La transmission de données rend possible leur interception par un acteur malveillant
Stockage	La plupart des données envoyées par les IoT ne sont pas utilisées. Un traitement plus intelligent embarqué permet de réduire la quantité de données nécessaire au cloud
Latence	Le temps de réponse d'un IoT intelligent est plus faible car la dépendance à la transmission des données au cloud est réduite.

Tableau 16 – Arguments en faveur du TinyML en fonction des problématiques posées

Après l'entraînement classique du modèle sur ordinateur ou via le cloud, TinyML s'appuie sur un processus de compression profonde. Les principales bibliothèques de logiciels pour le déploiement de modèles de deep learning pré-entraînés sur IoT sont : TensorFlow Lite, Pytorch Mobile, Apache TVM. Les terminaux peuvent alors parfois être équipés de GPU voire de NPU. Google a développé des puces EdgeTPU qui sont des puces ASIC personnalisées pour l'exécution d'inférences sur des dispositifs à la périphérie avec des besoins réduits en énergie (utilisant 2 watts pour 4 Téra FLOPS).

Annexe 16 - Tableaux comparatif des modes d'apprentissage

Tableau comparatif des modes d'apprentissage

	Supervisé	Non supervisé	Renforcé	Profond - Supervisé	Profond - Non Supervisé	Profond - Renforcé
Types de données en entrée	 Structurées	 Structurées	 Structurées	 Non structurées	 Non structurées	 Structurées et non structurées
Apprentissage en continu	 Non	 Non	 Oui	 Non	 Non	 Oui
Transfert d'apprentissage*	 Non	 Non	 Non	 Oui	 Oui	 Oui
Complexité	 Faible à moyenne	 Moyenne	 Élevée	 Très élevée	 Extrêmement élevée	 Extrêmement élevée
Durée de l'apprentissage	 Secondes à heures	 Minutes à jours	 En continu	 Minutes à jours	 Minutes à jours	 En continu
Transparence et explicabilité	 Très bonne	 Bonne	 Faible	 Très faible	 Nulle	 Nulle
Risque de biais	 Fort	 Faible à Moyen	 Faible à Moyen	 Fort	 Moyen	 Moyen
Type de ressources	CPU	CPU	CPU	CPU, TPU ou GPU	CPU, TPU ou GPU	CPU, TPU ou GPU

Comparatif issu du livre blanc "Intelligence Artificielle : restez maître de votre futur" réalisé par Business & Decision (groupe Orange). © Business & Decision

Figure 26 - Tableau comparatif des modes d'apprentissage (Source : Business & Decision)

Annexe 17 - Les Six principes clés de l'innovation frugale

Transformer les contraintes en opportunité

Les entreprises doivent adopter un positionnement optimiste dans un monde de plus en plus VUCA (Volatilité, Incertitude, Complexité, Ambiguïté) et considérer que chaque problème a sa solution. Dans un article du Harvard Business Review (81), cette approche est illustrée avec la startup finlandaise Soletair Power qui a mis en place une technologie capable de capter le CO₂ à partir du système de ventilation d'un immeuble de bureaux. Elle combine ensuite le CO₂ avec de l'hydrogène pour créer des carburants liquides plus durables.

Apprendre à faire mieux avec moins

Nous vivons dans un monde aux ressources finies (cf. Rapport Meadows), aussi les entreprises devront apprendre à innover pour mieux utiliser les ressources pour délivrer de la valeur à ces parties prenantes. Dacia, après avoir lancé la Logan à 5000 euros en 2004, a fait de son dernier modèle Dacia Spring à 12400 euros, la voiture électrique la moins chère d'Europe.

Penser et agir de manière flexible

Le rythme du changement s'est beaucoup accéléré ces dernières années et les dirigeants doivent pouvoir se remettre en question régulièrement pour envisager d'autres voies, sortir des référentiels existants. Cela suppose une certaine agilité d'esprit, une capacité à apprendre voire à désapprendre. Décathlon a mené entre septembre 2019 et avril 2020 un exercice d'intelligence collective à l'échelle mondiale ouvert à tous pour définir ce qui était attendu de Décathlon. Parmi les 5 objectifs définis pour mettre en place sa Vision V21.1 (<https://decathlonvision.com/fr/>) figure une économie régénérative bénéficiant à la fois aux Hommes et à la planète.

Opter pour la simplicité

Les services R&D des entreprises, notamment en France, ont généralement tendance à concevoir des produits aux fonctionnalités complexes dont très peu de personnes en font usage. Les consommateurs sont davantage en attente de solutions simples qui répondent réellement à leurs attentes ou à leur problématique.

Intégrer les fragiles et les exclus

La situation sanitaire a entraîné une augmentation de la précarité économique de bon nombre de citoyens à faible revenu. Le pouvoir d'achat des classes moyennes ne cesse également de reculer. L'enjeu pour l'entreprise est ici de pouvoir proposer des solutions réellement abordables et de créer de la valeur par une plus grande intégration de personnes en difficulté.

Via un réseau de 6000 buralistes, Nickel a par exemple permis à plus de 2 millions de Français d'ouvrir un compte sans banque pour 20 euros de tenue de compte seulement.

Suivre son cœur pour rester fidèle à soi-même et à ses valeurs.

Pour favoriser la prise de décision et l'engagement, les entreprises ont tout à gagner à tenir compte de l'intuition des employés et leur permettre de poursuivre leur passion pour identifier des moyens de contribuer positivement à la société et à la planète. L'engagement de la MAIF, « l'assureur militant » est sur ce point éclairant puisqu'en avril 2020 son DG, Pascal Demuger, reversa 100 millions d'euros d'économie à ses sociétaires pour les accompagner pendant la crise Covid-19 et leur éviter ainsi de devoir licencier. Bien qu'à court terme, cette décision ait entraîné une réduction par quatre de ses résultats par rapport à 2019, la MAIF a défendu ses valeurs de solidarité et séduit de nouveaux adhérents. Elle estime à moyen terme pouvoir être rentable tout en contribuant au bien commun.

Annexe 18 - INR – Parcours global de montée en compétences

« numérique responsable »

L'Institut du Numérique Responsable (INR) propose un parcours global de montée en compétences « numérique responsable » en 3 étapes :

- **Certificat de connaissances « Numérique responsable »**

Ce certificat atteste la maîtrise des connaissances fondamentales de l'approche Numérique Responsable. Il peut être obtenu par le passage d'un examen en ligne, sous forme de QCM.

Le MOOC Numérique Responsable proposé par l'INR (82) permet de s'y préparer en proposant 4h30 de contenus (vidéos, textes et contenus interactifs). Il est gratuit et accessible à tous.

- **Attestation de suivi de formation**

Portée par La Rochelle Université et l'INR, la formation de compétences Numérique Responsable, d'une durée de 2 jours, permet d'acquérir des connaissances et des compétences pratiques nécessaires à la mise en œuvre de la démarche Numérique Responsable.

- **Certificat ambassadeur « Numérique Responsable »**

Cette étape du parcours est en cours d'expérimentation depuis 25 février 2021 avec la Région Nouvelle-Aquitaine sur un premier module de sensibilisation. La formation d'une durée de 6 mois a pour objectif de former des référentes et référents de la démarche Numérique Responsable. Le certificat d'ambassadeur atteste la maîtrise de la démarche numérique responsable.

Annexe 19 - Double boucle Manager / Leader

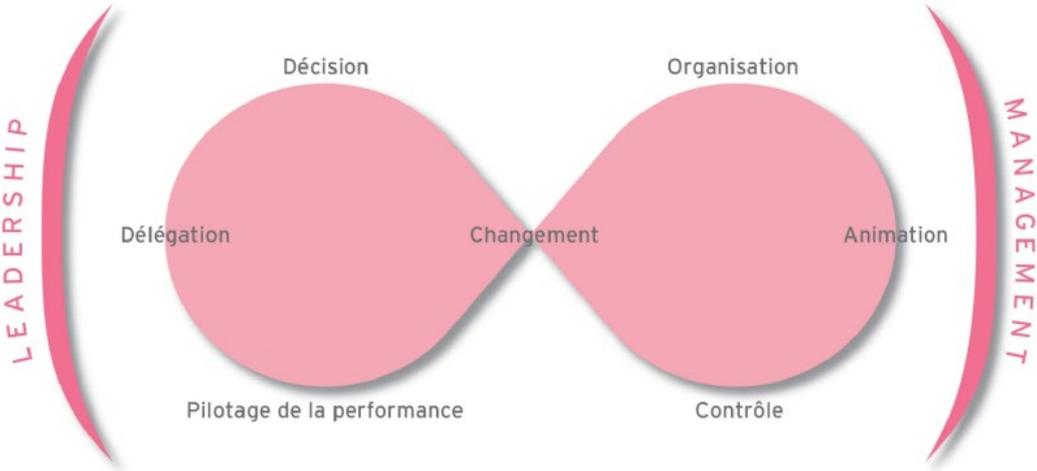


Figure 27 - Double boucle Manager / Leader (Source : emarketing.fr)

Annexe 20 – Feuille de route conduite du changement

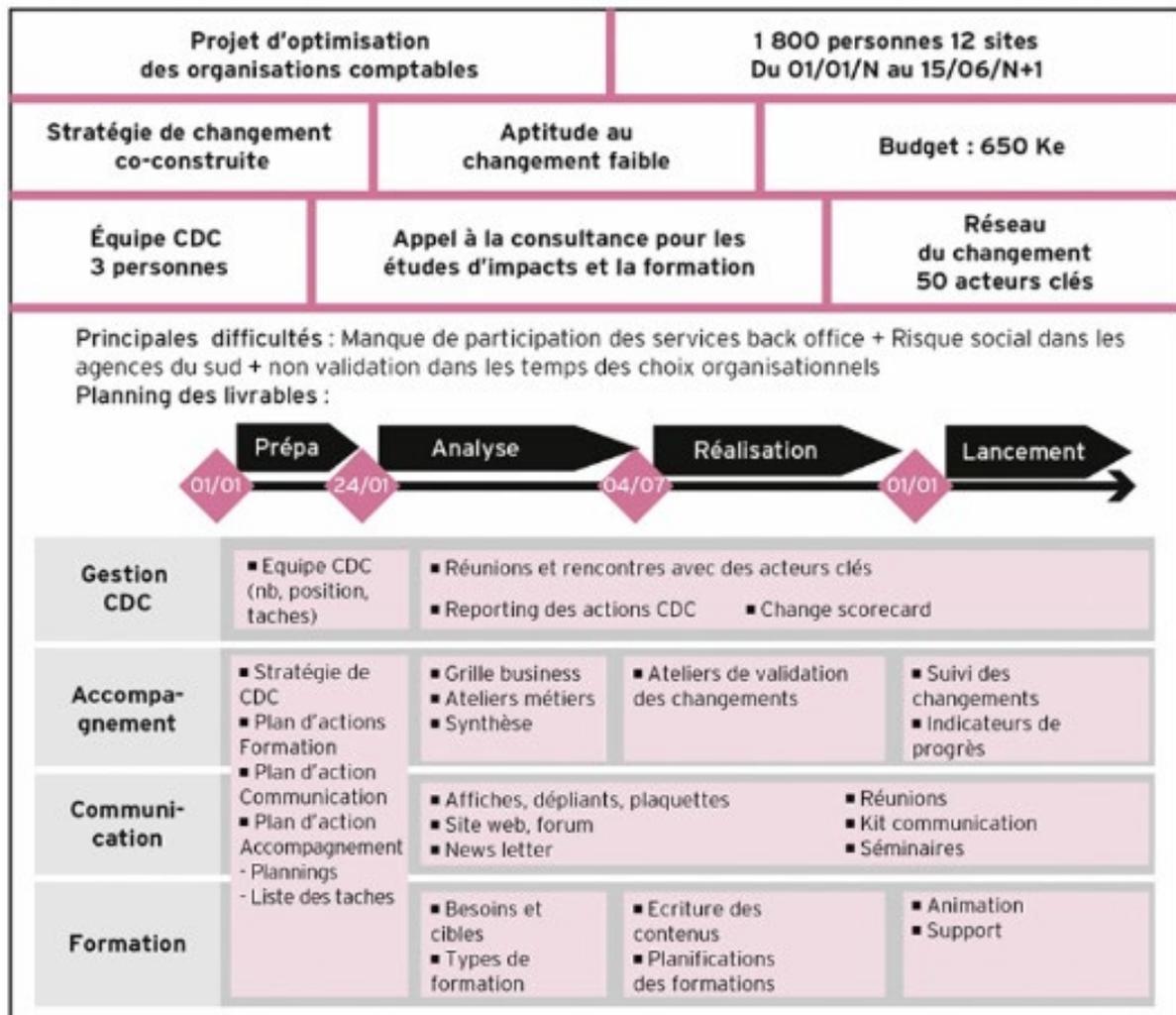


Tableau 17 - Feuille de route conduite du changement (Source : emarketing.fr)

Annexe 21 – Tableau de suivi des actions de conduite du changement

Au...	Date prévisionnelle	Fait/Pasfait/Retard	Dépenses prévues	Dépenses réelles	Graphiques
Études d'impacts & accompagnement					
Repérage quantitatif et qualitatif des impacts					<p>En retard 20 % A faire 60 % Fait 20 %</p>
Analyse des impacts humains					
Etude d'impacts process facturation					
Principaux changements humains par population					
Plan de transition					
Actions d'accompagnement population 1					
Formation					
Définition des besoins de formation					<p>En retard 10 % A faire 40 % Fait 80 %</p>
Définition des bénéficiaires et des contraintes					
Plan de formation					
Logistique de formation					
Ecritures des supports de formation					
Formation vague 1					
Evaluation des formation					
Evaluation des stagiaires					
Communication					
Positionnement Com					<p>En retard 10 % A faire 10 % Fait 80 %</p>
Mix Com					
Plan de communicatuin					
Kit de communication					
Site Web					
Affiche					
Réunions de lancement					
Réunions d'information					
Pilotage					
Tableau de bord de conduite du changement					<p>En retard 40 % A faire 30 % Fait 30 %</p>
Questionnaire ICAP et Verbatim					
ICAP 1 et verbatim 1					
ICAP 2 et verbatim 2					
Gestion des riques					
Indicateurs métiers					
Indicateurs de transformation					
Réunions de pilotage					

Tableau 18 - Tableau de suivi des actions de conduite du changement (Source : emarketing.fr)

Annexe 22 – Exemple de questionnaire ICAP

Taux d'information	Oui, régulièrement	Oui, de temps en temps	Oui, une seule fois	Non, jamais
Avez-vous entendu parler du projet ?				
Votre hiérarchie vous a-t-elle présenté le projet ?				
Avez-vous reçu des informations concernant le projet ?				
Avez-vous recherché des informations concernant le projet ?				
Taux de compréhension	Oui, de manière exhaustive	Oui, de manière partielle	Non, mais j'aimerais	Non, pas du tout
Comprenez-vous les objectifs et les modalités du projet ?				
Avez-vous sollicité des personnes ou des experts pour une meilleure compréhension du projet ?				
Avez-vous une idée claire de ce qui changera pour vous ?				
Vous êtes-vous documenté sur le projet et ses thématiques ?				
Taux d'adhésion	Oui, complètement	Oui, dans l'ensemble	Oui, un peu	Non, pas du tout
Pensez-vous que ce projet soit utile pour l'entreprise ?				
Pensez-vous que ce projet soit utile pour votre activité ?				
Pensez-vous que ce projet soit bien accueilli par les salariés ?				
Pensez-vous que toutes les conditions soient réunies pour la réussite de ce projet ?				
Taux de participation	Oui, c'est une priorité	Oui, de manière partielle	Non, mais j'aimerais	Non, et ça ne m'intéresse pas
Avez-vous participé à des réunions du projet ?				
Avez-vous produit des diagnostics, analyses et propositions pour le projet ?				
Avez-vous envisagé une intégration des éléments du projet dans votre activité ?				
Avez-vous testé certains éléments du projet dans votre activité ?				

Tableau 19 - Exemple de questionnaire ICAP (Source e-marketing.fr)

Annexe 23 - Cartographie des acteurs



Blog Gestion de Projet

Cartographie des acteurs

Catégorie d'acteurs	Points d'appui à exploiter	Points de vigilance à adresser	Actions possibles
C1 : engagés/alignés	Sont le noyau de l'alliance, prêts à prendre des responsabilités, à contribuer avec énergie pour faire aboutir le projet	Peuvent être susceptibles, et veulent être parties prenantes dans toutes les décisions. Veulent s'impliquer partout au risque de perturber le management du projet,	Les écouter et tenir compte leurs propositions et critiques, Les impliquer prudemment en leur donnant des responsabilités, de la reconnaissance et les exploiter pour influencer les hésitants
C4 : Hésitants C7 : indécis	Sont potentiellement faciles à convertir en alliés engagés si l'on fait preuve de persuasion	Peuvent être une proie facile pour les autres groupes qui sont dans l'opposition	Leur consacrer du temps sur le terrain, identifier leurs motivations, et leur faire des propositions concrètes pour les persuader du bien-fondé du projet
C5/ C6 : Passifs	Ils ne présentent pas de nuisance pour le projet	Il est difficile de les mettre en action. En général ce sont des suiveurs	Le fait de leur demander de s'engager est une perte de temps. Il faut simplement leur montrer de l'intérêt par une proximité terrain.
C2 : Opposants	Défendent une position claire contre le projet avec des arguments. Ils sont ouverts au dialogue.	Ont un pouvoir d'influence sur les autres qui sont indécis ou hésitants	Surveiller leurs actions de près Et dialoguer avec eux pour les embarquer le plus tôt possible, car ils deviendraient des alliés très puissants
C3 : Déchirés	En général ils sont peu nombreux dans cette position très inconfortable et finissent par changer de position en début de projet	Peuvent facilement devenir des opposants actifs ou des irréductibles	Les rassurer sur les impacts positifs attendus du changement, et de rester à leur écoute en sondant leur moral pour agir au bon moment
C 8 : Irréductibles	Sont des opposants sans arguments valables. En général ils ont des comptes à régler avec le passé dans l'entreprise ou avec le management	Sont imprévisibles et potentiellement nuisibles	Les surveiller de près pour contrecarrer d'éventuelles actions dangereuses. Il suffit de les ignorer pacifiquement, car ils finiront par embarquer tôt ou tard quand ils verront les autres le faire

Figure 28- Cartographie des acteurs (Source : blog-gestion-de-projet.com)

Annexe 24 - Marguerite sociologique

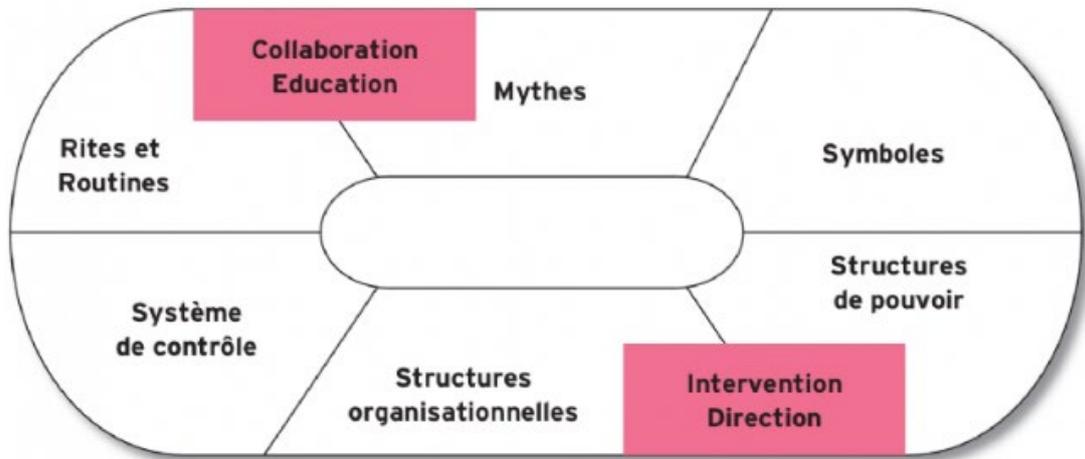


Figure 29 - Marguerite sociologique (Source : e-marketing.fr)

Annexe 25 - Changement et renforcement de la culture

 Symbole
  Organisation
  Processus d'appui
  Collaborateur

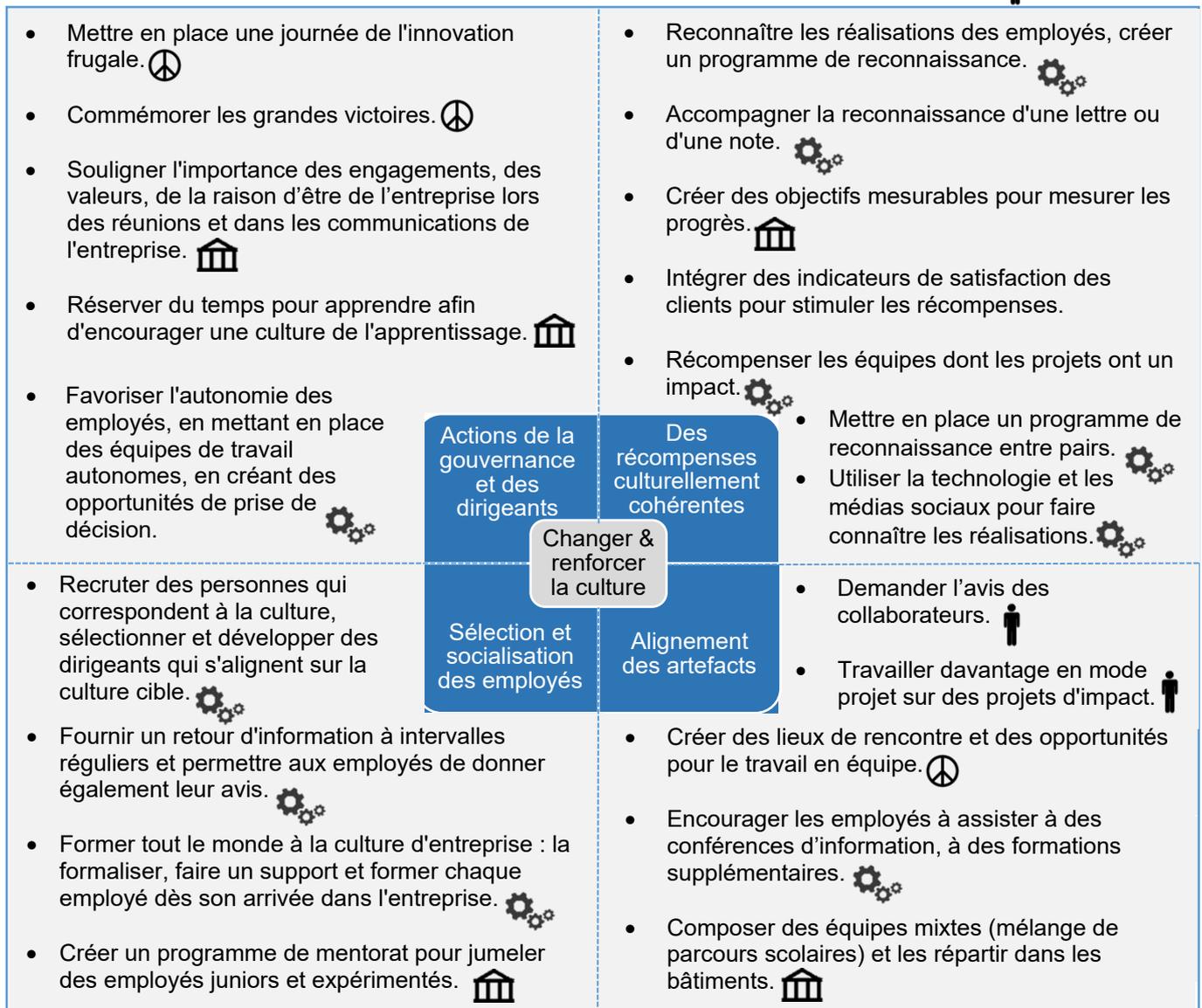


Figure 30 - Changement et renforcement de la culture

Annexe 26 - Grille d'auto-évaluation de la mission


LE SÉSAME : GRILLE D'AUTO-ÉVALUATION DE LA MISSION
 élaboré par   

Singularité de la mission	La mission pourrait être utilisée par n'importe quelle entreprise et le lien avec l'activité de l'entreprise n'est pas identifiable	La mission est clairement en lien avec le domaine d'activité de l'entreprise, mais pourrait s'appliquer à d'autres	La mission couvre toute l'activité de l'entreprise et un spécialiste du domaine identifie l'entreprise	La mission reflète sans ambiguïté l'identité, la culture et les convictions de l'entreprise et toute l'activité est couverte.
Mobilisation des parties prenantes	La mission ne pousse pas à l'action, et n'est pas impliquante pour ses parties prenantes	La mission implique quelques parties prenantes internes et clients, avec une appropriation et une incitation à l'action limitée	La mission est challengée par les parties prenantes internes et externes et génère une meilleure intégration de l'entreprise dans son écosystème	La mission est coconstruite avec des parties prenantes internes et externes clés, renforce les liens avec l'écosystème, et mène à l'action collective
Alignement avec le business model	Le chiffre d'affaires de l'entreprise est déconnecté de la mission, voire partiellement en contradiction	Une partie du chiffre d'affaires est aligné avec la mission, une autre partie ne l'est pas, voire crée des conflits d'intérêts potentiels	L'essentiel du chiffre d'affaires est aligné avec la mission, le complément n'est pas en contradiction	La totalité du chiffre d'affaires est alignée avec la mission
Contribution sociétale	La mission décrit l'écosystème actuel, sans remise en question sociale, sociétale, environnementale	La mission énonce des objectifs généraux et une conscience de la responsabilité de l'entreprise	La mission décrit un ou plusieurs objectifs de transformation environnementale, sociale et sociétale en lien avec des parties prenantes impactées	La mission décrit une contribution sociétale définie et intégrée, les parties prenantes sont impactées et la mission énonce des objectifs vérifiables de transformation
Ambition d'innovation	La mission décrit ce qu'est l'entreprise aujourd'hui, sans vision d'avenir	La mission décrit ce qu'est l'entreprise aujourd'hui et intègre des questions d'innovation et de transformation sociétale	La mission et les nouvelles relations avec l'écosystème aident à l'exploration de nouveaux champs d'innovation	La mission offre une vision d'avenir long-terme, réinventée et responsable, tant sur les enjeux que sur le rôle que jouera l'entreprise pour y répondre

Tableau 20 - Grille d'auto-évaluation de la mission (Source : www.entreprisesamission.org)

Table des figures

Figure 1 - Part des émissions mondiales de CO2 du numérique.....	23
Figure 2 - Typologies de Data Center et logiques d'implantation territoriale (27).....	28
Figure 3 - IA et Sciences des données (Source dilepix)	32
Figure 4 - Principe de fonctionnement de l'IA symbolique (Source : Conseil de l'Europe portail)	33
Figure 5 - Principe de fonctionnement de l'IA connexionniste (Source : Conseil de l'Europe portail) ..	33
Figure 6 - Représentation d'un neurone biologique et d'un neurone artificiel (Source : Brian Mwandau)	36
Figure 7 - Réseaux de neurones simple couche et multicouche (Source : Sufyan T. Faraj Al-Janabi)	36
Figure 8 - Architecture d'un CNN (Source : juripredis.com).....	37
Figure 9 - Bénéfices moyens réalisés grâce à l'utilisation de l'IA pour l'action climatique (Source: Capgemini)	43
Figure 10 - Évolution de la quantité de calcul en Péta FLOPS pour différents modèles CV, NLP et Speech (Source : Amir Gholami) (44)	45
Figure 11 – Exemples d'empreinte GES d'applications d'IA (40)	46
Figure 12 - Illustration de l'élagage d'un réseau de neurones (Souce : Google Research)	47
Figure 13 - Modèle Enseignant-Etudiant pour la distillation (Source : analyticsindiamag.com)	48
Figure 14 - Schéma de principe du transfert learning (Source : Young Jae Kim et al, 2021).....	48
Figure 15 - The Golden Circle	53
Figure 16 - Cartographie des alliés (Source : outilspourdiriger.fr)	60
Figure 17 – Augmentation de la température à la surface du globe depuis 1850-1900 en fonction des émissions cumulées de CO ₂ (GtCO ₂) (3).....	76
Figure 18 - Objectifs de Développement Durable (ODD).....	79
Figure 19 - Principaux pays producteurs de minerais rares (Source Sia Partners).....	81
Figure 20 - Cartographie des scopes 1, 2 et 3 (source : ADEME – Bilan GES)	84
Figure 21 - Cellule LSTM (Source : Pensée artificielle)	89
Figure 22 - Evolution du nombre de paramètres pour les modèles Transformer (Source: ysance)	90
Figure 23 - Évolution du volume de données dans le monde, 2010-2035 (Souce : ADVAES)	91
Figure 24 - Extrait du canevas Impact AI	92
Figure 25 – Flux de travail TensorFlow Lite Lite (Source : makeueof.com).....	101
Figure 26 - Tableau comparatif des modes d'apprentissage (Source : Business & Decision)	102
Figure 27 - Double boucle Manager / Leader (Source : emarketing.fr)	106
Figure 28- Cartographie des acteurs (Source : blog-gestion-de-projet.com).....	110
Figure 29 - Marguerite sociologique (Source : e-marketing.fr).....	111
Figure 30 - Changement et renforcement de la culture	112

Table des tableaux

Tableau 1 - Cibles carbone et engagements	23
Tableau 2 – Estimation par taille et par nombre de serveurs du nombre de sites dans le monde en 2017	28
Tableau 3 – Comparaison des MIPS par équipement (24) (30).....	30
Tableau 4 – Illustration de l'effet rebond par secteur	31
Tableau 5 - Performance comparée des modèles NLP GPT-3 et Switch Transformer.....	41
Tableau 6 - Comparatif d'architecture : Convolution standard vs Convolution séparable en profondeur (Source: Quantmetry)	49
Tableau 7 - Exemple d'objectifs - Résultats clés.....	55
Tableau 8 - IA Frugale - Trame d'audit.....	64
Tableau 9 – Évolution de la température mondiale à la surface de la Terre pour trois périodes de 20 au regard des cinq scénarios SSP étudiés par le Giéc (73)	75
Tableau 10 - Engagements des GAFAM pour l'environnement.....	82
Tableau 11 – Technologies d'émissions négatives (TEN)	83
Tableau 12 - Clean Energy Index par opérateur de centres de données (Source Greenpeace)	85
Tableau 13 - PUE moyen par type de data center (75)	86
Tableau 14 - R2E et % d'énergie utile par type de data center	88
Tableau 15 - Comparaison des grands modèles d'IA en fonction du nombre de paramètres.....	98
Tableau 16 – Arguments en faveur du TinyML en fonction des problématiques posées.....	101
Tableau 17 - Feuille de route conduite du changement (Source : emarketing.fr).....	107
Tableau 18 - Tableau de suivi des actions de conduite du changement (Source : emarketing.fr)	108
Tableau 19 - Exemple de questionnaire ICAP (Source e-marketing.fr).....	109
Tableau 20 - Grille d'auto-évaluation de la mission (Source : www.entreprisesamission.org)	113

Table des équations

Équation 1 - PUE (Power Usage Effectiveness)	86
Équation 2 - ERE (Energy Reuse Effectiveness)	86
Équation 3 – ERF (Energy Reuse Factor).....	87
Équation 4 - REF (Renewable Energy Factor).....	87
Équation 5 - CUE (Carbon Usage Effectiveness)	87
Équation 6 – WUE (Water Usage Effectiveness).....	87
Équation 7 - R2E (Ratio Efficacité Énergétique)	88

Table des matières

L'IA FRUGALE, UN ENJEU INDISPENSABLE POUR LE DEPLOIEMENT DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE	1
REMERCIEMENTS	2
LISTES DES SIGLES ET ABREVIATIONS	3
RESUME	4
LES PERSONNES INTERVIEWEES	5
SOMMAIRE	6
INTRODUCTION	7
PARTIE 1 - LES ENJEUX ENVIRONNEMENTAUX	9
1. CONTEXTE ENVIRONNEMENTAL	9
a. <i>Les rapports et enquêtes</i>	9
b. <i>Les accords et objectifs</i>	11
2. LES DOCUMENTS CADRES POUR L'IA ET LE NUMERIQUE	13
a. <i>Rapports et feuilles de route</i>	13
b. <i>Le cadre législatif</i>	16
3. TENDANCES SOCIETALES.....	18
a. <i>Génération climat</i>	18
b. <i>Paradoxe des digital natives</i>	20
c. <i>Des jeunes davantage formés</i>	20
PARTIE 2 - LES IMPACTS ENVIRONNEMENTAUX DU NUMERIQUE	21
1. L'EVALUATION DE L'IMPACT ENVIRONNEMENTAL	21
a. <i>Les indicateurs environnementaux</i>	21
b. <i>Les approches d'évaluation</i>	25
2. L'UNIVERS DU NUMERIQUE.....	27
a. <i>Les data centers</i>	27
b. <i>Les réseaux</i>	29
c. <i>Les terminaux</i>	29
3. L'EFFET REBOND	30
4. LES LABELS DU NUMERIQUE	31
PARTIE 3 - L'INTELLIGENCE ARTIFICIELLE ET SA SOUTENABILITE	32
1. QU'EST-CE QUE L'IA ?	32
a. <i>Définition</i>	32

b.	<i>IA connexionniste et IA symbolique</i>	33
c.	<i>Le machine learning</i>	34
d.	<i>Le Deep Learning</i>	36
e.	<i>L'IA distribuée</i>	38
2.	L'ECOSYSTEME BIG DATA ET IA.....	38
a.	<i>Le Big data</i>	39
b.	<i>Les processeurs</i>	39
c.	<i>L'ordinateur quantique</i>	42
d.	<i>Le edge computing</i>	43
3.	L'EMERGENCE DE L'IA POUR L'ENVIRONNEMENT	43
4.	LES IMPACTS ENVIRONNEMENTAUX DE L'IA.....	44
a.	<i>Consommation énergétique</i>	44
b.	<i>Les outils d'évaluation des émissions carbone</i>	45
5.	AMELIORATION DES PERFORMANCES ENERGETIQUES DE L'IA	46
a.	<i>Amélioration des algorithmes</i>	47
b.	<i>Amélioration de l'infrastructure matérielle et logicielle</i>	49
c.	<i>Amélioration des données</i>	50
d.	<i>Amélioration des pratiques</i>	50
	PARTIE 4 - VERS UNE CULTURE DE LA FRUGALITE	52
1.	L'INNOVATION FRUGALE.....	52
a.	<i>La frugalité</i>	52
b.	<i>Les 6 principes clés</i>	52
2.	LE CADRE DU CHANGEMENT.....	53
a.	<i>Le « Quoi »</i>	53
b.	<i>Le « Comment »</i>	53
c.	<i>Le « Pourquoi »</i>	53
3.	LE PILOTAGE ET L'ACCOMPAGNEMENT AU CHANGEMENT.....	54
a.	<i>Accompagnement des projets</i>	54
b.	<i>Développement humain</i>	56
c.	<i>Pilotage de la transformation</i>	58
d.	<i>Ouvertures et perspectives</i>	60
4.	DU CHANGEMENT A LA TRANSFORMATION.....	61
a.	<i>La raison d'être</i>	61
b.	<i>L'entreprise à mission</i>	62
5.	IA FRUGALE, UN AUDIT DES PRATIQUES NECESSAIRE.....	63
	CONCLUSION	65

BIBLIOGRAPHIE	66
ANNEXES	74
ANNEXE 1 - RAPPORT DU GIEC.....	75
ANNEXE 2 - LES OBJECTIFS DE DEVELOPPEMENT DURABLE (ODD)	79
ANNEXE 3 - PRINCIPAUX PAYS PRODUCTEURS DE MINERAIS RARES	81
ANNEXE 4 - ENGAGEMENTS DES GAFAM POUR L'ENVIRONNEMENT	82
ANNEXE 5 - TECHNOLOGIES D'EMISSIONS NEGATIVES (TEN).....	83
ANNEXE 6 – CARTOGRAPHIE DES SCOPES 1, 2 ET 3.....	84
ANNEXE 7 - CLEAN ENERGY INDEX PAR OPERATEUR DE CENTRES DE DONNEES	85
ANNEXE 8 - INDICATEURS DE PERFORMANCES DES DATA CENTERS	86
ANNEXE 9 - MODELES LSTM, GRU ET TRANSFORMER.....	89
ANNEXE 10 – ÉVOLUTION DU VOLUME DE DONNEES DANS LE MONDE	91
ANNEXE 11 - EXTRAIT DU CANEVAS IMPACT AI	92
ANNEXE 12 - ARTICLE LINKEDIN	93
ANNEXE 13 - COMPARAISON DES GRANDS MODELES D'IA	98
ANNEXE 14 - LES OUTILS D'ÉVALUATION DES EMISSIONS CARBONE.....	99
ANNEXE 15 - TINY ML	101
ANNEXE 16 - TABLEAUX COMPARATIF DES MODES D'APPRENTISSAGE	102
ANNEXE 17 - LES SIX PRINCIPES CLES DE L'INNOVATION FRUGALE.....	103
ANNEXE 18 - INR – PARCOURS GLOBAL DE MONTEE EN COMPETENCES « NUMERIQUE RESPONSABLE ».	105
ANNEXE 19 - DOUBLE BOUCLE MANAGER / LEADER.....	106
ANNEXE 20 – FEUILLE DE ROUTE CONDUITE DU CHANGEMENT.....	107
ANNEXE 21 – TABLEAU DE SUIVI DES ACTIONS DE CONDUITE DU CHANGEMENT	108
ANNEXE 22 – EXEMPLE DE QUESTIONNAIRE ICAP	109
ANNEXE 23 - CARTOGRAPHIE DES ACTEURS.....	110
ANNEXE 24 - MARGUERITE SOCIOLOGIQUE	111
ANNEXE 25 - CHANGEMENT ET RENFORCEMENT DE LA CULTURE	112
ANNEXE 26 - GRILLE D'AUTO-EVALUATION DE LA MISSION	113
TABLE DES FIGURES	114
TABLE DES TABLEAUX.....	115
TABLE DES EQUATIONS.....	116
TABLE DES MATIERES.....	117