Promouvoir des modèles d'intelligence artificielle frugale pour et par les politiques publiques

Maxime Amisse, Mélissa Faur, Lucie Gonard, André Orcesi

Promouvoir des modèles d'intelligence artificielle frugale pour et par les politiques publiques

Rapport de Groupe d'analyse de l'action publique

Mastère « Politiques et action publiques pour le développement durable » (PAPDD)

Année universitaire 2023/2024

Maxime AMISSE, Mélissa FAUR, Lucie GONARD, André ORCESI

Encadrés par Vincent Spenlehauer (Directeur du pôle de formation à l'action publique - ENPC)

pour le compte de Mathieu Aubry, Directeur de recherche au laboratoire IMAGINE / LIGM - ENPC et Romain Loiseau, laboratoire IMAGINE / LIGM - ENPC

Confidentialité : Toute information mentionnée issue des entretiens ne peut être diffusée ou utilisée à d'autres fins que l'évaluation de ce rapport.

L'Ecole des Ponts ParisTech, AgroParisTech et le laboratoire IMAGINE / LIGM n'entendent donner aucune approbation ni improbation aux thèses et opinions émises dans ce rapport ; celles-ci doivent être considérées comme propres à leurs auteurs.

Nous attestons que ce rapport est le résultat de notre travail collectif, qu'il cite entre guillemets et référence toutes les sources utilisées et qu'il ne contient pas de passages ayant déjà été utilisés intégralement dans un travail similaire.

Ce travail a bénéficié d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre de France 2030 portant la référence "ANR-23-PEIA-0008".

Remerciements

Nous souhaitons remercier nos commanditaires, MM. Mathieu Aubry et Romain Loiseau, pour leur confiance et leurs conseils tout au long du déroulé du groupe d'analyse de l'action publique. Nous espérons que ce rapport apporte des éléments de réponse aux questions posées sur l'IA frugale en vue de son utilisation pour et par les politiques publiques. Nous souhaitons également remercier notre encadrant académique, M. Vincent Spenlehauer, qui a su nous guider au fil du projet pour orienter au mieux nos travaux d'entretien et d'analyse. Nous remercions enfin tous les acteurs de l'IA qui ont accepté d'échanger avec nous sur l'IA frugale, et qui ont apporté une matière fertile pour la rédaction de ce rapport.

Table des sigles

AOM : Autorité Organisatrice des Mobilités

ASP : Agence de Service et de Paiement (établissement public du Ministère de l'Agriculture et de la Souveraineté Alimentaire)

CAF : Caisse d'Allocations Familiales

CE : Commission Européenne

Cerema : Centre d'Études et d'expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement

CGDD : Commissariat général au développement durable

CNIL : Commission Nationale de l'Informatique et des Libertés

CNRM : Centre National de la Recherche Météorologique

CNRS : Centre National de la Recherche Scientifique

CPU : Central Processing Unit, Unité Centrale de Traitement

DIAT : Appel à projets Démonstrateurs d'IA frugale pour les transitions écologique et énergétique dans les territoires

DINUM : Direction Interministérielle du Numérique

DGE : Direction Générale des Entreprises

Ecolab : Laboratoire d'innovation qui dépend du CGDD et du Ministère de la Transition Ecologique

ENSG : École Nationale des Sciences Géographiques

EPA : Établissement Public Administratif

EPCI : Établissement Public de Coopération Intercommunale

EPIC : Établissement Public à caractère Industriel et Commercial

ETALAB : Administrateur général des données, des algorithmes et des codes sources, département de la DINUM

FLAIR : French Land cover from Aerospace ImageRy

GENCI : Grand Équipement National de Calcul Intensif

GPU : Graphic Processing Unit, processeur graphique ou unité de traitement graphique

IA : Intelligence Artificielle

IDRIS : Institut du développement et des ressources en informatique scientifique

IGN : Institut National de l'information Géographique et forestière

Inria : Institut National de Recherche en sciences et technologies du numérique

INRAE : Institut National de la Recherche pour l'Agriculture, l'alimentation et l'Environnement

IPEF : Ingénieur des Ponts, des Eaux et des Forêts

LLM : large language model

LNE : Laboratoire national de métrologie et d'essais

ML : Model learning

MTECT : Ministère de la Transition Écologique et de la Cohésion des Territoires

NLP : Natural Language Processing (Traitement automatique du langage naturel)

OCDE : Organisation de Coopération et de Développement Économiques

ONF : Office National des Forêts

PAC : Politique Agricole Commune

PEPR : Programmes et Équipements Prioritaires de Recherche

SHARP : PEPR IA "Sharp Theoretical and Algorithmic Principles for frugal ML"

SNIA : Stratégie Nationale pour l'Intelligence Artificielle

TRL : Technology Readiness Level

UE : Union Européenne

ZAN : Zéro Artificialisation Nette

Résumé

Le développement de l'intelligence artificielle (IA) est une priorité pour la France qui a instauré une stratégie nationale pour l'IA depuis 2018. Ce développement correspond à divers enjeux sur le plan de la recherche, de l'économie, de la modernisation de l'action publique, de la régulation et de l'éthique.

Le contexte de la transition écologique et de la transition énergétique nous interroge sur nos usages, sur la consommation des ressources et de manière globale sur l'impact environnemental de nos activités dans différents domaines, dont l'IA. Le domaine de l'IA s'est notamment développé ces dernières années sur un modèle de données toujours plus volumineuses et de calculs de plus en plus complexes pour atteindre des niveaux de performance élevés. Ce modèle montre ses limites dans un contexte où l'urgence climatique et les préoccupations environnementales grandissantes nous conduisent à rechercher des compromis entre performance et poids environnemental. C'est ainsi que les développements actuels en France conduisent à intégrer une dimension de frugalité dans le domaine de l'IA, en termes d'optimisation des quantités de données utilisées, d'architecture des algorithmes, de choix de matériel, ou de source d'énergie utilisée, afin de réduire l'impact environnemental de l'IA tout en conservant un niveau de performance acceptable.

Sur la base d'une analyse bibliographie (publications, appels à projets, littérature grise) et d'entretiens avec les acteurs du domaine de l'IA, cette étude aborde divers aspects de l'IA frugale, dont sa définition, les différentes interprétations observées, la manière dont l'administration s'approprie ce concept et enfin le rôle que la recherche publique peut tenir pour participer à l'intégration de l'IA frugale dans nos usages.

Mots clés : IA, IA frugale, impact environnemental, frugalité, optimisation, données, éthique.

Abstract

The development of artificial intelligence (AI) is a priority for the French government who has enforced a national strategy for AI since 2018. Developing AI tackles various challenges in terms of research, economy, modernisation of public administration, regulation and ethics.

The necessary ecological and energy transitions make us question our uses, our resource consumption and more generally the environmental impact of our activities in different areas, including AI. Over the past few years, AI has soared and spawned bigger and bigger models, fed with larger and larger datasets, capable of running increasingly complex calculations to achieve high levels of performance, prediction and precision. However, we wish to quit this ever growing tendency as we are aware of the climate emergency and we share growing concerns regarding the environment. AI development should be balanced between increased performances and environmental impact. In France, recent AI developments pave the way to integrating “frugality” into AI systems. It consists in optimising the volume of data sets used for training and inference as well as the internal architecture of the AI algorithms, choosing hardware and its source of energy in order to reduce the total environmental impact of AI, while maintaining an acceptable level of performance.

Our study is based on a bibliographic research (scientific works and administrative documents mostly) and on interviews with AI stakeholders. We address various features of frugal AI, including its definition, different interpretations by the interviewees, how public administration is embracing this new concept and eventually how public research could play a part in disseminating new practices.

Keywords: AI, frugal AI, environmental impact, frugality, optimisation, data, ethics.

Sommaire

Introduction

La stratégie nationale pour l'intelligence artificielle (SNIA) a été lancée en 2018, à la suite du rapport Villani (Villani, 2018), avec l'objectif de positionner la France comme un leader européen et mondial en matière d'intelligence artificielle. La première phase 2018-2022, financée à hauteur de 1,85 milliards d'euros, a eu pour objectif de structurer l'écosystème de recherche et développement en IA, de mettre en place un certain nombre de communs numériques en IA dont le supercalculateur Jean Zay et de préparer l'écosystème pour saisir des opportunités de changements d'échelle de marché en IA.

Depuis 2022, la SNIA est entrée dans une deuxième phase, dotée de 1,5 milliards d'euros dans le cadre de France 2030. L'axe de développement est beaucoup plus centré sur comment influencer la prise en compte de l'IA dans l'écosystème économique français. Cette deuxième phase est élaborée en synergie avec le plan coordonné UE pour l'IA.

Trois piliers stratégiques constituent l'architecture de cette deuxième phase. Le premier pilier est le soutien à une offre deeptech qui fait référence à des innovations dans des domaines prioritaires comme l'IA embarquée (intégration de capacités d'IA directement dans des appareils, systèmes et objets connectés, par exemple le véhicule autonome), l'IA frugale (qui fait l'objet de ce rapport), l'IA de confiance (explicabilité des modèles d'IA), ou encore l'IA générative (IA capable de créer de nouveaux contenus et idées, notamment des conversations, des histoires, des images, des vidéos et de la musique). Le deuxième pilier fait référence à la formation et l'attraction de talents. Enfin, le troisième pilier porte sur le rapprochement de l'offre et de la demande de solutions en IA: passage à l'échelle de plusieurs accélérateurs, accompagnement des PME à l'intégration de solutions IA dans leur processus interne et dans leur processus de production…

Cette étude s'intéresse à l'IA frugale qui est un des domaines prioritaires de l'offre deeptech. Le terme d'IA frugale est apparu récemment dans le contexte d'initiatives diverses, notamment via la politique du MTECT, dans le cadre de recherches, de démarches citoyennes d'un certain nombre de collectivités, de démarches portées par des administrations, des associations ou encore des acteurs privés s'intéressant à la frugalité de l'IA, la frugalité de la donnée et le développement du numérique éco-responsable.

Pour l'IA frugale, la SNIA ambitionne de favoriser la recherche amont dans le cadre de programmes prioritaires de recherche (PEPR sur l'Intelligence artificielle piloté par l'Inria

portant sur les principes théoriques et algorithmiques de l'apprentissage frugal) (CNRS, 2023). De façon complémentaire vis-à-vis des besoins en aval, elle ambitionne de soutenir des démonstrateurs ou des développements technologiques d'IA frugale via un dispositif porté par le MTECT : appel à projets "Démonstrateurs d'IA frugale pour les transitions écologique et énergétique dans les territoires" (DIAT) (Ministère de la Transition Écologique et de la Cohésion du Territoire, 2023). D'une part, ce projet de la SNIA répond à l'enjeu de transition écologique en faisant un bilan coût avantage des solutions d'IA. D'autre part, cela le tourne fortement vers les territoires avec des projets locaux qui répondent à des thématiques locales de façon à diffuser la culture de l'IA et à proposer des solutions réplicables sur d'autres territoires.

Dans ce contexte, les objectifs de ce mémoire sont de dresser un panorama des enjeux de la frugalité de l'intelligence artificielle pour les politiques publiques et de comprendre le paysage d'acteurs concernés. Il vise à établir un lien entre les développements de recherche en IA frugale et leur intégration dans les politiques publiques, en ce qui concerne la prise de conscience des acteurs, les besoins d'acculturation, ainsi que les besoins de formation des acteurs publics à une IA plus frugale.

Une première partie consiste à proposer une définition du terme d'IA frugale, d'une part sur la base d'éléments bibliographiques, et d'autre part sur la base des échanges avec les professionnels du secteur.

Une deuxième partie s'intéresse à la manière dont le concept d'IA frugale est perçu et mis en place dans l'administration ainsi qu'à la diversité des usages observée et à la place de la frugalité dans ces usages.

Enfin, une troisième partie s'intéresse aux liens entre l'administration et la recherche et sur le développement de l'IA frugale.

I. IA frugale et positionnement du sujet

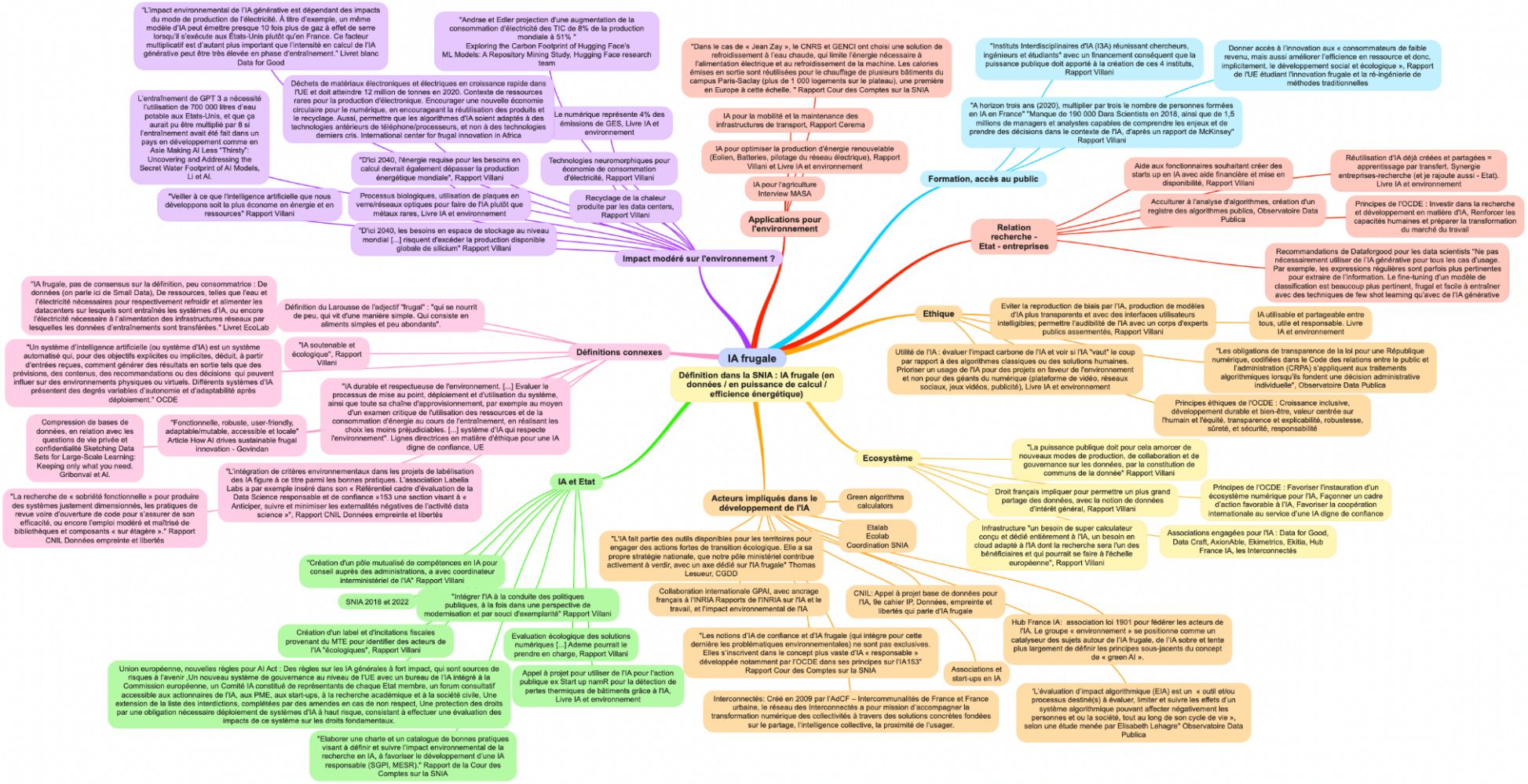
Cette première partie permet de circonscrire notre sujet d'étude, en commençant par une problématisation prolongée par une définition de l'IA frugale, de ses enjeux, et des acteurs du domaine, et de la vision que ces acteurs ont du concept de frugalité.

1. Problématique

A la découverte du sujet proposé par notre commanditaire, nous nous sommes d'abord posé la question de la nature de notre objet d'étude. L'Intelligence Artificielle est déjà en elle-même un concept scientifique et moral difficile à appréhender, et qui porte des sens très variés dans le langage courant. Ajouté à ce concept la qualification de « frugal » introduit une nouvelle dimension à cet espace déjà non métrique, qui complexifie davantage la typologie de ce que l'on appelle l'IA. Nous avons relevé le défi de répondre au sujet proposé par notre commanditaire, à savoir « promouvoir des modèles d'IA frugale pour et par les politiques publiques » en mettant en perspective notamment la relation entre la recherche en IA frugale (qui en France est financée dans le cadre d'un PEPR appelé SHARP et dans lequel le laboratoire de notre commanditaire est impliqué) et les acteurs des politiques publiques (ministères, établissements publics par exemple).
Pour cerner davantage la demande du commanditaire, nous avons tout d'abord effectué une recherche bibliographique pour définir de manière univoque ce qu'est l'intelligence artificielle, et analyser si le concept d'IA frugale apparaît déjà dans la littérature académique ou administrative, ou si cela se présente sous d'autres formes.

2. Définition de l'IA frugale par étude bibliographique

C'est par une étude bibliographique que nous avons commencé à aborder le sujet de l'IA frugale. Celle-ci nous a amené à parcourir des sources très diverses, que nous vous invitons à visualiser dans un diagramme pieuvre (Figure 1).


IA frugale
Définition dans la SNIA : IA frugale (en données / en puissance de calcul / efficience énergétique)
Impact modéré sur l'environnement ?
"L'impact environnemental de l'IA générative est dépendant des impacts du mode de production de l'électricité. À titre d'exemple, un même modèle d'IA peut émettre presque 10 fois plus de gaz à effet de serre lorsqu'il s'exécute aux États-Unis plutôt qu'en France. Ce facteur multiplicatif est d'autant plus important que l'intensité en calcul de l'IA générative peut être très élevée en phase d'entraînement." Livret blanc Data for Good
"Andrae et Edler projection d'une augmentation de la consommation d'électricité des TIC de 8% de la production mondiale à 51% " Exploring the Carbon Footprint of Hugging Face's ML Models: A Repository Mining Study, Hugging Face research team
Déchets de matériaux électroniques et électriques en croissance rapide dans l'UE et doit atteindre 12 million de tonnes en 2020. Contexte de ressources rares pour la production d'électronique. Encourager une nouvelle économie circulaire pour le numérique, en encourageant la réutilisation des produits et le recyclage. Aussi, permettre que les algorithmes d'IA soient adaptés à des technologies antérieurs de téléphone/processeurs, et non à des technologies derniers cris. International center for frugal innovation in Africa
Le numérique représente 4% des émissions de GES, Livre IA et environnement
L'entraînement de GPT 3 a nécessité l'utilisation de 700 000 litres d'eau potable aux Etats-Unis, et que ça aurait pu être multiplié par 8 si l'entraînement avait été fait dans un pays en développement comme en Asie Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models, Li et Al.
"D'ici 2040, l'énergie requise pour les besoins en calcul devrait également dépasser la production énergétique mondiale", Rapport Villani
Technologies neuromorphiques pour économie de consommation d'électricité, Rapport Villani
"Veiller à ce que l'intelligence artificielle que nous développons soit la plus économe en énergie et en ressources" Rapport Villani
Processus biologiques, utilisation de plaques en verre/réseaux optiques pour faire de l'IA plutôt que métaux rares, Livre IA et environnement
Recyclage de la chaleur produite par les data centers, Rapport Villani
"D'ici 2040, les besoins en espace de stockage au niveau mondial [...] risquent d'excéder la production disponible globale de silicium" Rapport Villani
Applications pour l'environnement
IA pour l'agriculture Interview MASA
IA pour optimiser la production d'énergie renouvelable (Eolien, Batteries, pilotage du réseau électrique), Rapport Villani et Livre IA et environnement
IA pour la mobilité et la maintenance des infrastructures de transport, Rapport Cerema
"Dans le cas de « Jean Zay », le CNRS et GENCI ont choisi une solution de refroidissement à l'eau chaude, qui limite l'énergie nécessaire à l'alimentation électrique et au refroidissement de la machine. Les calories émises en sortie sont réutilisées pour le chauffage de plusieurs bâtiments du campus Paris-Saclay (plus de 1 000 logements sur le plateau), une première en Europe à cette échelle. " Rapport Cour des Comptes sur la SNIA
Formation, accès au public
"Instituts Interdisciplinaires d'IA (I3A) réunissant chercheurs, ingénieurs et étudiants" avec un financement conséquent que la puissance publique doit apporté à la création de ces 4 instituts, Rapport Villani
Donner accès à l'innovation aux « consommateurs de faible revenu, mais aussi améliorer l'efficience en ressource et donc, implicitement, le développement social et écologique », Rapport de l'UE étudiant l'innovation frugale et la ré-ingénierie de méthodes traditionnelles
"A horizon trois ans (2020), multiplier par trois le nombre de personnes formées en IA en France" "Manque de 190 000 Dara Scientists en 2018, ainsi que de 1,5 millions de managers et analystes capables de comprendre les enjeux et de prendre des décisions dans le contexte de l'IA, d'après un rapport de McKinsey" Rapport Villani
Relation recherche - Etat - entreprises
Aide aux fonctionnaires souhaitant créer des starts up en IA avec aide financière et mise en disponibilité, Rapport Villani
Réutilisation d'IA déjà créées et partagées = apprentissage par transfert. Synergie entreprises-recherche (et je rajoute aussi - Etat). Livre IA et environnement
Acculturer à l'analyse d'algorithmes, création d'un registre des algorithmes publics, Observatoire Data Publica
Principes de l'OCDE : Investir dans la recherche et développement en matière d'IA, Renforcer les capacités humaines et préparer la transformation du marché du travail
Recommandations de Dataforgood pour les data scientists "Ne pas nécessairement utiliser de l'IA générative pour tous les cas d'usage. Par exemple, les expressions régulières sont parfois plus pertinentes pour extraire de l'information. Le fine-tuning d'un modèle de classification est beaucoup plus pertinent, frugal et facile à entraîner avec des techniques de few shot learning qu'avec de l'IA générative
Ethique
Eviter la reproduction de biais par l'IA, production de modèles d'IA plus transparents et avec des interfaces utilisateurs intelligibles; permettre l'audibilité de l'IA avec un corps d'experts publics assermentés, Rapport Villani
IA utilisable et partageable entre tous, utile et responsable. Livre IA et environnement
Utilité de l'IA : évaluer l'impact carbone de l'IA et voir si l'IA "vaut" le coup par rapport à des algorithmes classiques ou des solutions humaines. Prioriser un usage de l'IA pour des projets en faveur de l'environnement et non pour des géants du numérique (plateforme de vidéo, réseaux sociaux, jeux vidéos, publicité), Livre IA et environnement
"Les obligations de transparence de la loi pour une République numérique, codifiées dans le Code des relations entre le public et l'administration (CRPA) s'appliquent aux traitements algorithmiques lorsqu'ils fondent une décision administrative individuelle", Observatoire Data Publica
Principes éthiques de l'OCDE : Croissance inclusive, développement durable et bien-être, valeur centrée sur l'humain et l'équité, transparence et explicabilité, robustesse, sûreté, et sécurité, responsabilité
Ecosystème
"La puissance publique doit pour cela amorcer de nouveaux modes de production, de collaboration et de gouvernance sur les données, par la constitution de communs de la donnée" Rapport Villani
Droit français impliquer pour permettre un plus grand partage des données, avec la notion de données d'intérêt général, Rapport Villani
Principes de l'OCDE : Favoriser l'instauration d'un écosystème numérique pour l'IA, Façonner un cadre d'action favorable à l'IA, Favoriser la coopération internationale au service d'une IA digne de confiance
Infrastructure "un besoin de super calculateur conçu et dédié entièrement à l'IA, un besoin en cloud adapté à l'IA dont la recherche sera l'un des bénéficiaires et qui pourrait se faire à l'échelle européenne", Rapport Villani
Associations engagées pour l'IA : Data for Good, Data Craft, AxionAble, Ekimetrics, Ekitia, Hub France IA, les Interconnectés
Définitions connexes
Définition du Larousse de l'adjectif "frugal" : "qui se nourrit de peu, qui vit d'une manière simple. Qui consiste en aliments simples et peu abondants".
"IA frugale, pas de consensus sur la définition, peu consommatrice : De données (on parle ici de Small Data), De ressources, telles que l'eau et l'électricité nécessaires pour respectivement refroidir et alimenter les datacenters sur lesquels sont entraînés les systèmes d'IA, ou encore l'électricité nécessaire à l'alimentation des infrastructures réseaux par lesquelles les données d'entraînements sont transférées." Livret EcoLab
"IA soutenable et écologique", Rapport Villani
"Un système d'intelligence artificielle (ou système d'IA) est un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influer sur des environnements physiques ou virtuels. Différents systèmes d'IA présentent des degrés variables d'autonomie et d'adaptabilité après déploiement." OCDE
"IA durable et respectueuse de l'environnement. [...] Evaluer le processus de mise au point, déploiement et d'utilisation du système, ainsi que toute sa chaîne d'approvisionnement, par exemple au moyen d'un examen critique de l'utilisation des ressources et de la consommation d'énergie au cours de l'entraînement, en réalisant les choix les moins préjudiciables. [...] système d'IA qui respecte l'environnement". Lignes directrices en matière d'éthique pour une IA digne de confiance, UE
Compression de bases de données, en relation avec les questions de vie privée et confidentialité Sketching Data Sets for Large-Scale Learning: Keeping only what you need. Gribonval et Al.
"Fonctionnelle, robuste, user-friendly, adaptable/mutable, accessible et locale" Article How AI drives sustainable frugal innovation - Govindan
"La recherche de « sobriété fonctionnelle » pour produire des systèmes justement dimensionnés, les pratiques de revue voire d'ouverture de code pour s'assurer de son efficacité, ou encore l'emploi modéré et maîtrisé de bibliothèques et composants « sur étagère »." Rapport CNIL Données empreinte et libertés
"L'intégration de critères environnementaux dans les projets de labélisation des IA figure à ce titre parmi les bonnes pratiques. L'association Labelia Labs a par exemple inséré dans son « Référentiel cadre d'évaluation de la Data Science responsable et de confiance »153 une section visant à « Anticiper, suivre et minimiser les externalités négatives de l'activité data science », Rapport CNIL Données empreinte et libertés
IA et Etat
"Création d'un pôle mutualisé de compétences en IA pour conseil auprès des administrations, a avec coordinateur interministériel de l'IA" Rapport Villani
SNIA 2018 et 2022
"Intégrer l'IA à la conduite des politiques publiques, à la fois dans une perspective de modernisation et par souci d'exemplarité" Rapport Villani
Création d'un label et d'incitations fiscales provenant du MTE pour identifier des acteurs de l'IA "écologiques", Rapport Villani
Evaluation écologique des solutions numériques [...] Ademe pourrait le prendre en charge, Rapport Villani
Union européenne, nouvelles règles pour AI Act : Des règles sur les IA générales à fort impact, qui sont sources de risques à l'avenir ,Un nouveau système de gouvernance au niveau de l'UE avec un bureau de l'IA intégré à la Commission européenne, un Comité IA constitué de représentants de chaque Etat membre, un forum consultatif accessible aux actionnaires de l'IA, aux PME, aux start-ups, à la recherche académique et à la société civile, Une extension de la liste des interdictions, complétées par des amendes en cas de non respect, Une protection des droits par une obligation nécessaire déploiement de systèmes d'IA à haut risque, consistant à effectuer une évaluation des impacts de ce système sur les droits fondamentaux.
Appel à projet pour utiliser de l'IA pour l'action publique ex Start up namR pour la détection de pertes thermiques de bâtiments grâce à l'IA, Livre IA et environnement
"Elaborer une charte et un catalogue de bonnes pratiques visant à définir et suivre l'impact environnemental de la recherche en IA, à favoriser le développement d'une IA responsable (SGPI, MESR)." Rapport de la Cour des Comptes sur la SNIA
Acteurs impliqués dans le développement de l'IA
Green algorithms calculators
Etalab Ecolab Coordination SNIA
"L'IA fait partie des outils disponibles pour les territoires pour engager des actions fortes de transition écologique. Elle a sa propre stratégie nationale, que notre pôle ministériel contribue activement à verdir, avec un axe dédié sur l'IA frugale" Thomas Lesueur, CGDD
Collaboration internationale GPAI, avec ancrage français à l'INRIA Rapports de l'INRIA sur l'IA et le travail, et l'impact environnemental de l'IA
CNIL: Appel à projet base de données pour l'IA, 9e cahier IP, Données, empreinte et libertés qui parle d'IA frugale
Hub France IA: association loi 1901 pour fédérer les acteurs de l'IA. Le groupe « environnement » se positionne comme un catalyseur des sujets autour de l'IA frugale, de l'IA sobre et tente plus largement de définir les principes sous-jacents du concept de « green AI ».
"Les notions d'IA de confiance et d'IA frugale (qui intègre pour cette dernière les problématiques environnementales) ne sont pas exclusives. Elles s'inscrivent dans le concept plus vaste d'IA « responsable » développée notamment par l'OCDE dans ses principes sur l'IA153" Rapport Cour des Comptes sur la SNIA
Associations et start-ups en IA
"L'évaluation d'impact algorithmique (EIA) est un « outil et/ou processus destiné(s) à évaluer, limiter et suivre les effets d'un système algorithmique pouvant affecter négativement les personnes et ou la société, tout au long de son cycle de vie », selon une étude menée par Elisabeth Lehagre" Observatoire Data Publica
Interconnectés: Créé en 2009 par l'AdCF – Intercommunalités de France et France urbaine, le réseau des Interconnectés a pour mission d'accompagner la transformation numérique des collectivités à travers des solutions concrètes fondées sur le partage, l'intelligence collective, la proximité de l'usager.

i. Tout d’abord, définissons l’IA et ses enjeux récents

Le concept d’IA en lui-même englobe une multitude de concepts et d’enjeux que nous détaillons dans cette partie.

a. L’IA se décline en différents types d’algorithmes

La définition la plus consensuelle de l’IA est celle établie par l’OCDE, qui a travaillé à englober tous les aspects que peut revêtir ce concept informatique de manière à permettre aux Etats Membres de légiférer sur l’IA (OCDE, 2023). L’organisation définit l’IA comme un système par la formulation suivante : « Un système d’intelligence artificielle (ou système d’IA) est un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir d’entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influer sur des environnements physiques ou virtuels. Différents systèmes d’IA présentent des degrés variables d’autonomie et d’adaptabilité après déploiement. »

Cette définition date de novembre 2023, et est une version actualisée de définition établie par l’OCDE en 2019 : « système automatisé qui peut, pour des objectifs définis par un être humain, faire des prédictions, des recommandations et des décisions qui influencent les mondes réels et virtuels. Les systèmes d’IA sont créés pour opérer à différents niveaux d’autonomie. »

Nous voyons que la nouvelle définition de l’OCDE, sortie seulement quatre ans après la première, a fortement rectifié le concept d’IA en y supprimant l’intervention humaine dans la définition des objectifs du système, en y ajoutant la nuance d’objectifs « explicites ou implicites », en retirant le verbe « pouvoir » attribué au système d’IA pour le remplacer par « déduire », ce qui rappelle que ces systèmes sont des systèmes informatiques déterministes, qui ne fonctionnent pas nécessairement comme des boîtes noires.

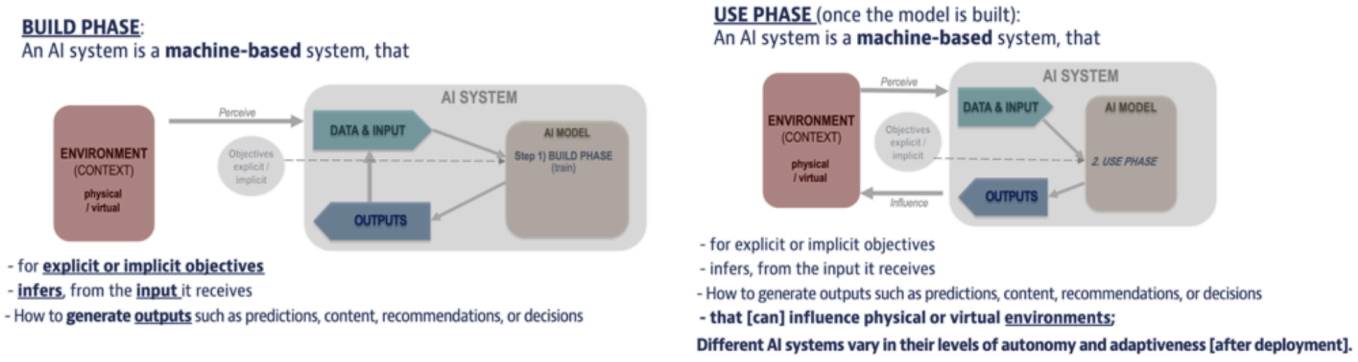


Figure 2 : Schémas blocs décrivant le fonctionnement d’un système IA (OCDE, 2023)

La nouvelle définition ajoute le mot « contenu » dans la liste des sorties d'un système d'IA, et nuance l'influence sur le monde extérieur en ajoutant la tournure « peuvent influer » plutôt que « influencent ». Enfin, l'expression « sont créés pour » est remplacée par « présentent » pour y intégrer le concept d'apprentissage non supervisé et d'apprentissage profond, dit Deep Learning, qui a été une révolution dans le monde de l'IA dans ses premiers développements à succès en 2012. Cette définition actualisée est présentée dans la Figure 2, extraite du site de l'OCDE.

Le mot IA fait donc référence à beaucoup de méthodes informatiques, dont les plus notoires sont : le machine learning (apprentissage automatique, supervisé continu, etc.), le deep learning (apprentissage profond), les algorithmes Random Forest - descente de gradient - SVM - Régressions statistiques - Plus proches voisins, les modèles de langage, les modèles génératifs, les réseaux de neurones artificiels (convolutifs, Transformers), le traitement du langage naturel et encore la vision par ordinateur. Nous vous invitons à consulter le glossaire de l'Intelligence artificielle de la CNIL si vous souhaitez explorer le champ lexical de l'IA plus avant (CNIL, 2023).

b. Les enjeux de l'IA, et les acteurs impliqués

L'OCDE propose une classification des algorithmes d'IA selon cinq critères : planète et peuples, contexte économique, données et entrées, modèle d'IA, tâches et sorties. Les membres de l'OCDE ont accepté en 2019 et renouvelé en 2023 leur allégeance aux principes de l'IA suivant (Figure 3) :

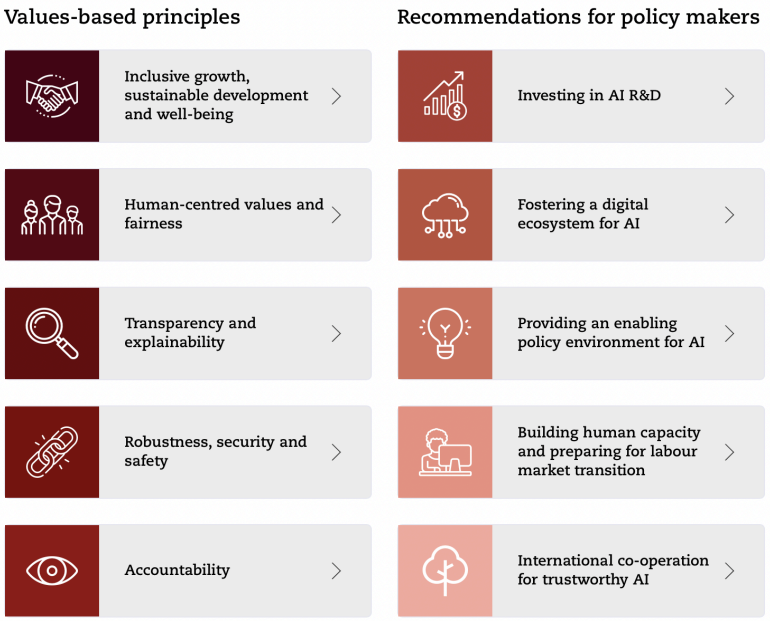


Figure 3 : Principes sur l'IA développés par l'OCDE (AI-Principles Overview - OECD.AI)

Ces principes cherchent à encadrer les premiers dangers découverts à l'utilisation de l'IA, à savoir :

- l'aspect « boîte noire » des modèles qui deviennent de plus en plus complexes et autonomes, et qui sont difficiles à expertiser par des non experts de l'IA;
- les enjeux d'éthique liés à l'IA et aux systèmes reposant sur des décisions suggérées par des IA, qui peuvent être biaisées (racisme, sexisme) selon les jeux de données utilisés en entrée du système;
- les enjeux d'impact environnemental de l'IA, qui ne pourraient être estimés que par une plus grande transparence dans la communication des acteurs de l'IA, et par une industrie de l'IA qui favoriserait l'exploitabilité des systèmes plutôt que la recherche incessante d'une précision accrue.

Cependant, ce n'est pas à l'ordre du jour des acteurs de l'IA, qui cherchent plutôt à développer des modèles d'IA gigantesques pour gagner en efficacité (GPT-4 100 fois plus gros en terme de volume du modèle que GPT-3), comme l'illustre l'article *Green AI* (Schwartz et all., 2020) et la Figure 4 qui présente le nombre d'articles scientifiques publiés au cours de trois grandes conférences académiques sur l'IA. Nous notons une nette majorité de publications en lien avec la recherche de précision (« accuracy ») plutôt que celle d'efficience, à savoir de modèles plus compacts mais avec des performances comparables.

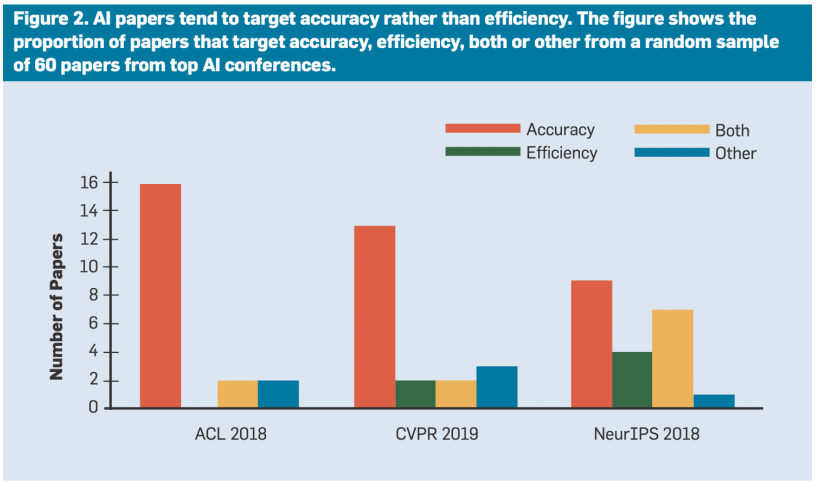


Figure 4: Occurrences du nombre de publications par thématique à l'occasion des trois dernières conférences internationales sur l'IA (Schwartz et al., 2020)

La démarche de l'OCDE a entraîné un rapprochement entre plusieurs pays qui ont fondé le Partenariat Mondial sur l'Intelligence Artificielle (GPAI, 2023), dirigé par la France et le

Canada, dans lequel des experts internationaux échangent sur les principes d'IA définis par l'OCDE dans le cadre de groupes de travail transdisciplinaires. Au niveau français, c'est Inria qui est partenaire du programme, et qui publie des travaux sur l'impact de l'IA sur la société (*Intelligence artificielle*, 2022) ou sur l'empreinte écologique des systèmes d'IA (Inria et al., 2023).

Face à ces nouveaux enjeux posés par l'IA, l'Union Européenne s'est saisie du sujet et les Etats membres se sont accordés sur une réglementation de l'IA, appelée « AI Act », qui est le premier règlement sur l'IA dans le monde (*Proposition de cadre réglementaire sur l'intelligence artificielle | Bâtir l'avenir numérique de l'Europe*, 2023). Ce règlement se fonde sur la définition de l'IA établie par l'OCDE. Le Conseil de l'Union Européenne et le Parlement européen ont récemment (Conseil de l'UE, 2023) adopté des nouveaux éléments à intégrer dans le règlement :

- Des règles sur les IA générales à fort impact, qui sont sources de risques à l'avenir;
- Un nouveau système de gouvernance au niveau de l'UE avec un bureau de l'IA intégré à la Commission Européenne qui veillera sur les récentes avancées en IA et l'application des règles de l'UE dans les États Membres, qui sera accompagné d'un panel d'experts scientifiques sur l'IA. A ce bureau s'ajoute un Comité IA constitué de représentants de chaque Etat membre qui sera un organe de conseil auprès de la Commission qui notamment rapportera sur l'implémentation des règles dans chaque État, et se fera conseiller par un forum consultatif accessible aux actionnaires de l'IA, aux PME, aux start-ups, à la recherche académique et à la société civile;
- Une extension de la liste des interdictions, complétées par des amendes en cas de non-respect;
- Une protection des droits par une obligation nécessaire au déploiement de systèmes d'IA à haut risque, consistant à effectuer une évaluation des impacts de ce système sur les droits fondamentaux.

La France, de son côté, a adopté une Stratégie Nationale pour l'Intelligence Artificielle en 2018, inspirée du rapport de Cédric Villani de 2017 (Villani, 2018), et l'a renouvelée en 2022 pour 4 ans. Cette stratégie est confiée à un coordinateur de l'intelligence artificielle, relevant du Ministère de l'Economie et des Finances. La stratégie vise à « faire de la France un leader mondial de l'intelligence artificielle » selon les mots du Président de la République. Pour cela, la stratégie a pour objectif de développer l'offre de formation en IA

en France, et de renforcer les démarches de recherche et développement portées par les acteurs publics (administrations et EPIC), acteurs privés et la recherche académique.

Depuis le début de la SNIA, la France est devenue riche de multiples start-ups et associations qui intègrent les enjeux sociétaux et environnementaux de l'IA mis en exergue par l'OCDE, et cherchent à impulser des réflexions dans la recherche académique et privée. C'est le cas notamment des associations DataCraft et Ekimetrics, qui organisent des séminaires sur l'IA frugale (Datacraft, 2023; Ekimetrics, 2023). C'est le cas plus généralement d'entités comme Hub France IA, Data for good, AxionAble et Ekitia qui portent auprès des acteurs de l'IA un accompagnement sur les sujets d'IA responsable, d'un point de vue éthique et environnemental (AxionAble, 2023; Data for Good, 2023; Ekitia, 2023; Hub France IA, 2023). La structuration des acteurs de l'IA en France débute à peine que des scandales éclatent déjà, notamment quand les algorithmes de la CAF et leurs biais ont été montrés du doigt par un article du Monde en décembre 2023 (Le Monde et al., 2023). L'enquête journalistique montre que les algorithmes utilisés par la CAF, qui se fondent sur du minage de données, utilisent des données personnelles des allocataires comme leur situation familiale, la composition de leur foyer, et leur vulnérabilité économique ou de santé, pour générer des « scores » qui seront la base d'une décision d'octroi d'allocations ou non. Ces algorithmes sont censés être audités pendant leur utilisation, ce qui n'a pas été fait depuis 13 ans selon la Commission Nationale Consultative des Droits de l'Homme (CNCDH).

Face aux dangers de l'IA, en termes de protection des données, d'éthique, et de lutte contre les discriminations, mais aussi face à l'énorme concurrence des modèles d'IA générative étrangers comme ChatGPT, différentes entités gouvernementales ont publié des rapports de préconisation en direction de l'Etat. C'est le cas de l'Observatoire DataPublica, qui s'interroge sur la transparence des algorithmes utilisés dans les administrations publiques (Data Publica & Banuls, 2023). Le rapport de l'observatoire, publié en septembre de cette année, souligne notamment que l'Etat est engagé constitutionnellement à être transparent, d'après déclaration des droits de l'homme et du citoyen et son article 15 : « *La société a le droit de demander compte à tout agent public de son administration.*» Cette nécessaire transparence, valeur de la République, est également inscrite dans la loi pour une République numérique qui impose une accessibilité en ligne aux «*traitements algorithmiques*» mis en place dans les administrations pour des prises de décision. La CNIL est dotée d'un pouvoir de contrôle et de sanction, et la CADA et les administrateurs ministériels des données sont compétents sur les algorithmes et doivent contribuer à leur

ouverture et leur publication destinée aux citoyens. Malgré tout, le rapport souligne qu'il n'existe « *pas d'autorité de contrôle externe et indépendante disposant d'une compétence générale de contrôle des algorithmes et de l'Intelligence Artificielle* ».

Dans son rapport d'avril 2023 sur la SNIA (Cour des Comptes, 2023), la Cour des Comptes souligne également que « les enjeux sociétaux prennent une importance particulière tant sur les perspectives éthiques que sur l'impact environnemental de l'IA » et cherche à vérifier si la SNIA prend en compte cela dans son deuxième volet lancé en 2022, qui définit comme nouvel objectif le développement de la "deeptech", et notamment de l'IA de confiance et de l'IA frugale. D'après la Cour, la SNIA n'a pas été assez ambitieuse sur le sujet de l'IA frugale, que nous allons définir juste après, notamment car le MTECT n'a pas défini de programme pour la recherche en IA frugale. Le Conseil d'État mentionne également la frugalité dans son étude sur l'intelligence artificielle et l'action publique (Conseil d'Etat, 2022).

c. L'écosystème de l'IA : infrastructure, énergie et données

Une dernière étape est nécessaire avant de définir l'IA frugale ; nous allons établir un panorama de l'écosystème d'une IA. Nous allons illustrer notre propos en utilisant une infographie du rapport ShapingAI rédigé par le MediaLab de SciencesPo (ShapingAI & MediaLab, 2023). Si nous revenons à la définition de l'OCDE, et aux images qui l'accompagnent, un système d'IA se crée en deux phases (Figure 2). La première est la phase d'apprentissage, où le modèle d'IA est nourri en entrée de données dites «d'entraînement » qui permettent à l'algorithme d'apprendre une tâche (par exemple détecter un chat sur une image, les données d'entrée étant alors des photos d'animaux agrémentées d'un label « chat » ou « pas chat »). L'apprentissage de cette tâche se conclut lorsque le modèle d'IA a « appris » la tâche, c'est-à-dire a défini des paramètres numériques qui lui permettent de reproduire cette tâche sur de nouvelles données. La deuxième phase, une fois que le modèle est complet (i.e que ses paramètres sont définis), est l'inférence : le modèle peut alors effectuer la tâche voulue sur tout type de données (pour reprendre l'exemple, une photo d'animaux sans label « chat » ou « pas chat », c'est le modèle d'IA qui dira s'il y a un chat ou pas).

Au fil des années, les systèmes d'IA se sont complexifiés de manière exponentielle (Figure 5) : les modèles les plus connus de deep learning utilisés aujourd'hui sont composés de millions de paramètres qui leur permettent d'effectuer leurs tâches. Le nombre de

paramètres internes au modèle représente directement la complexité de celui-ci, les plus gros modèles étant entraînés sur d'énormes bases de données (pour GPT-2, le modèle a été entraîné sur 8 millions de documents soit 40 Go de texte) et nécessitant également de lourdes infrastructures de traitement à chaque inférence. Nous illustrons cette course au gigantisme par le graphe suivant, issu du livre blanc de l'association Data for good sur l'IA générative (Data For Good et al., 2023).

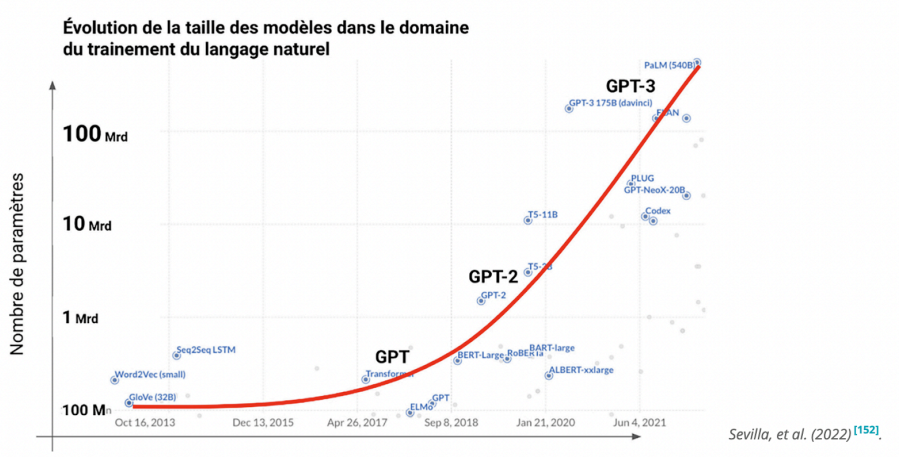


Figure 5 : Evolution de la taille (en nombre de paramètres) des modèles de langage naturel au fil des ans Semilog (Data For Good et al., 2023)

Mais alors, qu'est ce que physiquement une IA à part un processus algorithmique capable d'apprendre à effectuer une tâche à partir de données ? C'est en réalité toute une architecture d'infrastructures informatiques : les données sont stockées dans des serveurs, souvent regroupés en data centers. Les modèles d'IA les plus simples peuvent être entraînés sur des ordinateurs du commerce, où le modèle « apprend » en utilisant le processeur de la machine que l'on nomme le CPU. Pour les plus gros modèles d'IA, l'entraînement ne peut se faire qu'en utilisant des processeurs plus puissants nommés GPU, voire même plusieurs processeurs en même temps. Pour cela, les algorithmes « tournent » sur des clouds mis à disposition par des grandes entreprises d'informatique (Microsoft, Amazon, Google) qui sont constitués de réseaux de serveurs et processeurs accessibles en ligne, ou directement sur des superordinateurs qui peuvent être loués pour l'occasion, comme le supercalculateur français Jean Zay du CNRS. Enfin, l'inférence doit également se faire sur une infrastructure capable de « faire tourner » le modèle d'IA. Tout ceci constitue le système physique d'une IA.

Mais allons plus loin en regardant la Figure 6 : un système d'IA, au-delà de son fonctionnement et de son infrastructure attenante, est en réalité un écosystème, car le modèle créé va très certainement bouleverser les habitudes de ses créateurs.

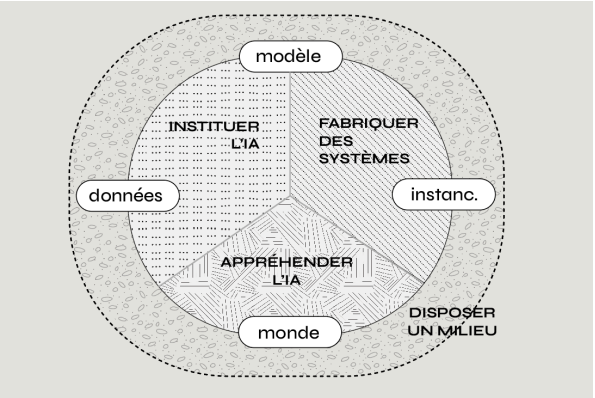
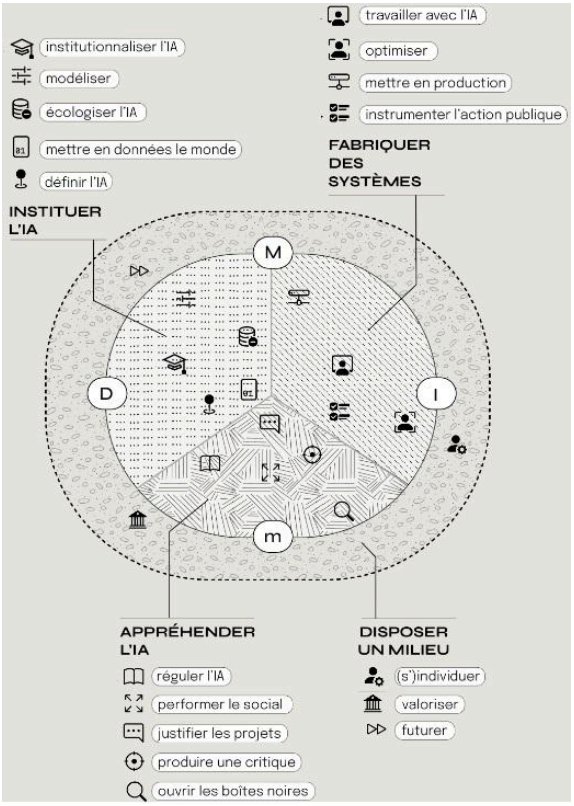


Figure 6 : Ecosystème de l'IA (ShapingAI & MediaLab, 2023)

Soit c'est un modèle de recherche qui vise à permettre une avancée dans des travaux académiques, soit c'est un modèle utilisé par une entreprise pour aider à la prise de décisions, ou un objectif commercial de vente, soit c'est un modèle utilisé par un site web pour optimiser ses performances (proposer des services personnalisés, améliorer sa publicité), soit c'est un modèle développé par l'administration publique comme une aide à la décision ou à l'instruction de procédures. Dans tous ces cas-là, le système d'IA occupe une nouvelle place irremplaçable pour ceux qui l'ont commandé. C'est en cela que les réflexions et réglementations naissantes sur l'IA cherchent à circonscrire différents domaines dans lesquels l'IA va avoir un impact, et réglementer les domaines les plus à risque d'être soumis à des externalités négatives comme l'éthique de l'IA. La Figure 7 illustre les enjeux relevés par la rapport Shaping AI, et la Figure 8 les enjeux qui sont déjà ou en cours d'être réglementés aux niveaux national et européen. Nous remarquons que les angles morts de ces premières règlementations sont principalement l'impact de l'IA sur l'environnement (item « écologiser l'IA »), la transparence algorithmique (item « ouvrir les boîtes noires ») et la mise en production. Nous verrons que ces trois angles morts sont traités par le concept de l'IA frugale par la suite.

L'écosystème de l'IA est également composé de toute la chaîne énergétique qui lui permet de fonctionner à savoir l'énergie qui alimente les serveurs, data centers, clouds ou superordinateurs, mais également le nombre d'utilisateurs (ou plutôt d'inférence) d'un système entraîné sur une certaine durée, ou encore la nature de la tâche qui est donnée à

l'IA. C'est cette composante de l'écosystème de l'IA qui va nous intéresser dans le cas de l'IA frugale. Nous cherchons d'abord à détailler les différents enjeux en lien avec la chaîne



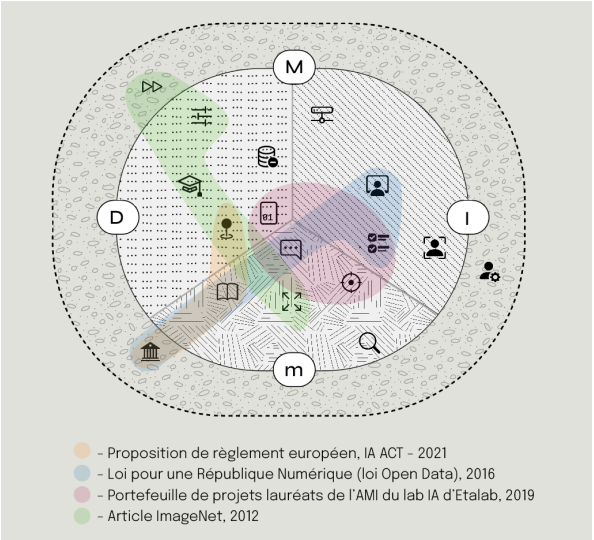


Figure 7 : Écosystème de l'IA et détail des enjeux de chaque domaine identifié (ShapingAI & MediaLab, 2023)
Figure 8 : Écosystème de l'IA et détail des réglementations récentes ou à venir suivant le domaine impacté

énergétique d'un système d'IA. Tout d'abord, cette chaîne est principalement fondée sur deux flux : un flux d'électricité et un flux d'eau. Nous l'illustrons par la figure ci-dessous qui présente ces deux flux, issu de l'article de Inria sur l'impact environnemental des systèmes d'IA (Inria et al., 2023; Li et al., 2023) :

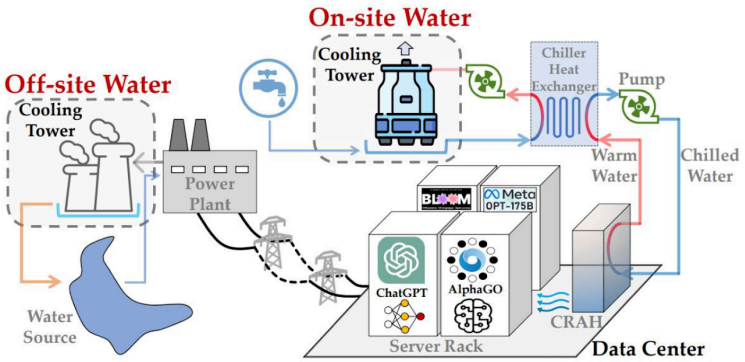


Figure 9 : Cycle de l'eau et de l'électricité pour alimenter les infrastructures hébergeant les systèmes d'IA (Inria et al., 2023) (Li et al., 2023)

Nous voyons en premier lieu que le circuit électrique est plutôt simple, et que l'impact environnemental d'un système d'IA dépend principalement du mix énergétique du pays dans lequel les infrastructures informatiques sont implantées. Le cycle de l'eau de refroidissement est quant à lui plus complexe : il faut compter à la fois l'eau qui peut servir à la production d'énergie (centrale nucléaire par exemple) mais également celle qui sert à refroidir les data centers où logent les données et le modèle d'IA, car les serveurs produisent de la chaleur en continu lorsqu'ils fonctionnent, et enfin ce que l'on fait de cette eau chauffée (soit refroidie et relâchée, soit réutilisée). En France, l'ADEME a mis en lumière toutes les ressources mobilisées par le numérique dans son ensemble dans son rapport (LEES PERASSO et al., 2022) qui détaille l'empreinte carbone du numérique et prévoit son augmentation pour 2030.

Toute une littérature se développe autour d'un usage plus raisonné des ressources de l'écosystème d'IA. Dans l'article de (Inria et al., 2023), il est noté qu'à l'avenir, la consommation d'énergie électrique liée aux technologies de l'information et de la communication (TIC) va fortement augmenter, ce qui va entraîner une plus grande empreinte carbone du secteur. Cela s'illustre dans la Figure 10, qui évalue le volume d'émissions de gaz à effets de serre dû aux TIC d'ici à 2040 selon différentes sources scientifiques. Nous voyons que cette croissance sera plus que linéaire dans tous les cas.

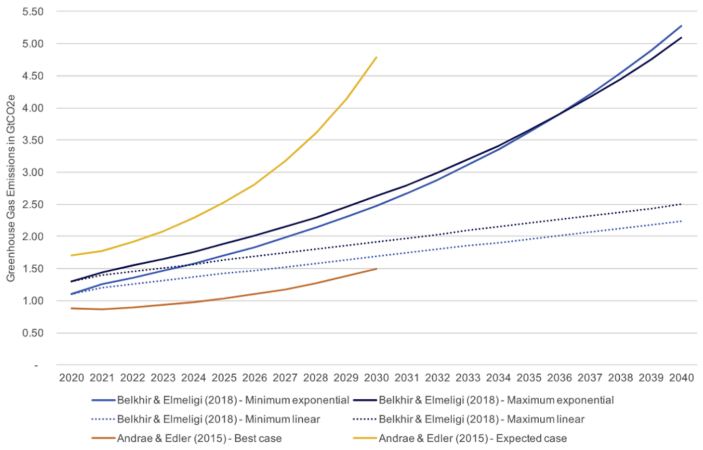


Figure 10 : Trajectoires d'évolution de la consommation électrique des TIC d'ici à 2050 (en GtCOAeq) (Inria et al., 2023)

De plus, la recherche en IA a abouti récemment à des modèles pratiques de calcul de l'impact carbone d'un système d'IA, comme cela a été fait pour le modèle de langage

BLOOM, développé par la start-up canadienne Hugging Face, et qui se veut être un Chat GPT moins gourmand en ressource, même si ce modèle est lui-aussi critiquable. Les Tableaux 1 et 2 sont issus du livret Data for good qui reprend les résultats des publications de Hugging Face (Strubell et al., 2019).

	GPT-3 (OpenAI)	BLOOM (BigScience)
Nombre de paramètres	175 milliards	176 milliards
Architecture	Transformer	Transformer
Energie consommée pour l'entraînement	1287 MWh	433 MWh
Intensité carbone du mix énergétique du pays hébergeur	423 gCO2eq / kWh	57 gCO2eq / kWh
Impact carbone total de la phase d'entraînement	552 tCO2eq	30 tCO2eq

Tableau 1 : Comparaison des consommations énergétiques de deux modèles d'IA génératives (Data For Good et al., 2023)

En particulier, nous voyons qu'un point déterminant de l'impact carbone en phase d'entraînement d'un système d'IA est le mix énergétique du pays où le modèle est hébergé et entraîné. Comme BLOOM a été entraîné en France et GPT-3 aux Etat-Unis, il y a un rapport d'environ 10 entre les deux intensités carbone. Il resterait à estimer le coût d'émission de GES des phases d'inférence, mais également de l'infrastructure informatique utilisée. Ce type d'analyse en cycle de vie des modèles d'IA commence à émerger auprès d'acteurs d'une IA plus responsable, comme des approches en cycle de vie qui sont explorées par la start-up Hugging face (Castaño et al., 2023) et également par l'Ademe dans sa méthode de calcul d'impact « Base Carbone » (Ademe, 2023). Cela se structure autour du concept de « green algorithms », ou algorithmes verts, des outils programmés pour calculer automatiquement l'empreinte carbone d'un code source informatique. Malheureusement ces premiers pas ne sont pas ceux suivis par les géants du numériques. Vous pouvez trouver ci contre le mix énergétique des trois pays développeurs de solutions de cloud pour les systèmes d'IA, et l'importance encore faible des énergies décarbonées dans le mix, issu d'un article de Hugging Face (Strubell et al., 2019) :

Pays / Service de Cloud Mix énergétique	Renouvelable	Gaz	Charbon	Nucléaire
Chine	22 %	3 %	65 %	4 %
Allemagne	40 %	7 %	38 %	13 %
Etats-Unis	17 %	35 %	27 %	19 %
Amazon Web Service	17 %	24 %	30 %	26 %
Google	56 %	14 %	15 %	10 %
Microsoft	32 %	23 %	31 %	10 %

Tableau 2 : Mix énergétiques des trois plus gros services de cloud en comparaison du mix de pays (Chine, Allemagne, Etats-Unis), (Strubell et al., 2019)

Il en va de même de la ressource en eau. En France, le supercalculateur Jean Zay a été construit de façon à ce que l'eau extraite pour le refroidir soit ensuite réutilisée dans le chauffage de bâtiments du campus universitaire Paris Saclay tout proche. Cela permet d'après la Cour des Comptes de fournir du chauffage pour 1000 logements par an. Cependant, si l'on prend l'exemple de Microsoft, l'article Li et al. explique que l'entraînement de GPT 3 a nécessité l'utilisation de 700 000 litres d'eau potable aux Etats-Unis, et que ça aurait pu être multiplié par 8 si l'entraînement avait été fait dans un pays en développement comme en Asie.

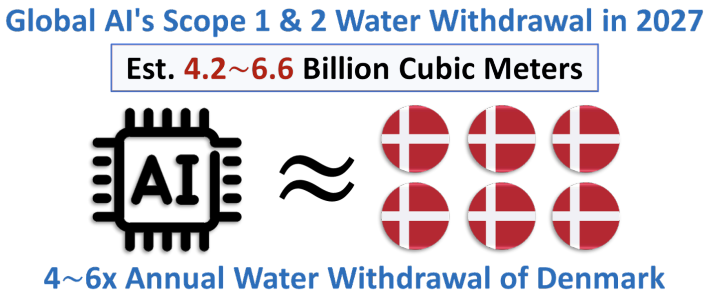


Figure 11 : Infographie de l'OCDE décrivant la consommation en eau de l'IA en 2027

L'OCDE rapporte également que cet article prédit que les systèmes d'IA consommeront d'ici 3 ans autant d'eau par an que quatre à six fois la consommation de toute la population du Danemark (Figure 11).

Enfin, l'écosystème de l'IA comporte également l'impact de l'IA sur la société. Dans une recherche de réalisation des objectifs du développement durable se pose la question des usages de l'IA. Comme le souligne la Cour des Comptes, une réduction de l'impact environnemental des systèmes d'IA ne peut être cohérente que si les IA sont utilisées pour lutter contre le changement climatique et la raréfaction des ressources. Ainsi, une IA responsable serait à la fois une IA sobre en termes d'écosystème informatique et énergétique, mais également une IA utilisée pour une tâche qui aide à la poursuite du développement durable. C'est l'objectif que fixait Cédric Villani dans son rapport de 2018, que la France devienne leader en IA soutenable et écologique. C'est ce que l'UE qualifie d'«*IA durable et respectueuse de l'environnement*» (Direction générale des réseaux de communication, 2019)**.** Nous nous rapprochons en cela du concept d'IA frugale.

ii. Ensuite, apposons-y « frugale »

Est frugal ce « qui se nourrit de peu, qui vit d'une manière simple. Qui consiste en aliments simples et peu abondants. » (Larousse, 2023). Le mot « IA frugale » est premièrement apparu dans les documents de synthèse de la première SNIA publiés par l'Etat Français, qui détaillent qu'un des piliers de la nouvelle SNIA sera les technologies deep tech, et en particulier l'IA frugale. Ce concept est donc 100% français, préfiguré par le rapport Villani qui souligné l'importance d'une IA soutenable, même si le terme « frugal » n'est pas présent dans le rapport. Le deuxième volet de la SNIA définit l'IA frugale comme frugale en données, en puissance de calcul et efficiente énergétiquement. A cette définition s'ajoute différents aspects selon les sources bibliographiques :

- Le concept présenté au-dessus d'une IA frugale et « écologique », appliquée à effectuer des tâches à impact positif dans la lutte contre le réchauffement climatique. C'est ce que soutient le rapport Villani, le rapport de l'UE sur une IA digne de confiance, et l'analyse de la SNIA effectuée par la Cour des comptes. C'est également ce que soutient Thomas Lesueur, commissaire général au développement durable, dans la toute récente feuille de route « Intelligence Artificielle et Transition Écologique » publiée par le ministère de la transition écologique, en ces mots : « *L'Intelligence Artificielle fait partie des outils disponibles […] pour engager des actions fortes de transition écologique. Elle a sa propre stratégie nationale […] avec un axe dédié sur l'Intelligence Artificielle frugale* »

(Ministère de la Transition Écologique et de la Cohésion du Territoire, 2023).

- Le concept d'une IA à comparer à d'autres méthodes informatiques à moindre coût en carbone, ce qui est soutenu par le livret Data for Good et le livre *IA et environnement* (Dunod et al., 2022): estimer par avance le coût de l'IA et les tâches précises qu'elle devra remplir avant de lancer son développement peut permettre de se tourner à rebours vers des solutions d'algorithmes plus classiques ou des solutions humaines.
- De plus, le concept d'IA frugale est généralement couplé au concept d'IA embarquée, qui est également un des axes de la nouvelle SNIA. L'IA frugale permettrait de construire des modèles d'IA moins volumineux qui pourraient être embarqués sur des appareils à faible capacité de traitement comme des téléphones par exemple. C'est ce que souligne Guillaume Avrin, Coordinateur National de la SNIA, dans la feuille de route du ministère : « *Les modèles d'IA frugaux sont également essentiels pour l'IA embarquée […] afin de pouvoir faire tourner des algorithmes localement sur tous types de véhicules* ». (Ministère de la Transition Écologique et de la Cohésion du Territoire, 2023)
- Un autre aspect souvent attribué à l'IA frugale serait un caractère d'accessibilité, de transparence et de partage des modèles. En effet, un pan des acteurs de l'IA frugale en France fondent leur recherche sur la création de modèles d'IA frugale qui seront réutilisables par d'autres acteurs sans nécessaire nouvelle phase d'entraînement, sans nécessaire investissement dans de grosses infrastructures, et même avec partage de données et de codes sources. C'est le cas du projet DIAT de la SNIA pour l'IA frugale qui est porté par le MTECT et son Ecolab, nommé « Démonstrateurs d'IA frugale dans les territoires ». Dans le livret blanc publié par l'Ecolab sont détaillés les critères d'évaluation des réponses à l'appel à projet : « la transparence des algorithmes, l'impact environnemental, la sécurité des systèmes et un usage raisonné et conforme des données. » (Ecolab et al., 2023). Sont également comptées la « réplicabilité des solutions proposées » dans un objectif de diffusion des solutions auprès d'autres acteurs publics. D'après le cahier *Données, empreinte et liberté* publié par la CNIL, l'organisme souligne qu'une bonne pratique pour IA frugale est « la recherche de *sobriété fonctionnelle* pour produire des systèmes justement dimensionnés, les pratiques de revue voire d'ouverture de code pour s'assurer de son efficacité, ou encore l'emploi modéré et maîtrisé de bibliothèques et composants *sur étagère*. » (CNIL et al., 2023).

- Un dernier aspect est celui d'une IA qui peut être développée par tous, par exemple par des acteurs publics comme les collectivités locales qui répondent à l'appel à projet DIAT et qui se sont rassemblées dans l'association Les Interconnectés. Mais cela pourrait également être toute personne qui pourrait accéder à d'algorithmes frugaux open source et explicables (transparents). Cela concerne également les pays en développement qui ne bénéficient pas des infrastructures et des volumes de données des pays développés, et qui souhaitent investir dans l'IA. Nous y reviendrons en fin de paragraphe avec l'exemple de l'Afrique.

Le concept d'IA frugale est de plus en plus représenté en France, que ce soit par le succès de l'appel à projet DIAT qui en est à sa deuxième vague d'appel, mais également par une communauté d'acteurs parapublics et privés structurante, que ce soit les établissements publics qui s'intéressent à l'IA frugale comme le Cerema, l'IGN, l'ADEME, le CSTB et d'autres encore cités dans la feuille de route du MTE, mais également les associations mentionnés déjà plus haut qui apportent conseil et orientation aux acteurs publics et privés qui veulent sauter le pas de l'IA frugale, ou plus généralement de l'IA responsable.
Cependant, nous avons remarqué que ce mot concept était très peu présent dans la littérature académique française, et encore moins à l'international. Nous avons retrouvé un concept similaire dans les publications du Dr Rémi Gribonval (Inria, ENS Lyon) qui travaille avec son équipe sur des modèles d'IA plus compacts (concept des matrices creuses et de la reproductibilité des articles de recherche en IA) (Gribonval et al., 2021). Nous avons également mentionné les rapports de Inria, beaucoup plus généraux, sur l'impact de l'IA sur le travail et le coût environnemental de l'IA. Cette littérature émerge depuis quelques années à l'international, comme nous l'avons déjà montré en citant les articles de la start-up canadienne Hugging Face. C'est également le cas d'un article de Govindan Kannan qui décrit comment l'IA peut accompagner une innovation soutenable et frugale en définissant cette dernière comme « fonctionnelle, robuste, user-friendly, émergente, avec un coût économique acceptable, et locale […] qui accompagne le développement durable » (Govindan, 2022). Ce concept d'innovation frugale est également partagé par l'Union Européenne et son rapport *Study on frugal innovation and reengineering of traditional techniques* qui veut donner accès à l'innovation aux « *consommateurs de faibles revenus, mais aussi améliorer l'efficience en ressource et donc, implicitement, le développement social et écologique* » (European Commission. Directorate General for Research and Innovation. et al., 2017).

Ce concept de frugalité de l'innovation, qui englobe l'IA frugale, est particulièrement présent dans des pays en développement comme le prouve les actions de l'*International center for frugal innovation in Africa*, qui cherche à ce que l'IA développée sur le continent soit « développée pour servir nos valeurs et engagements moraux, sociaux et légaux envers la soutenabilité, l'économie circulaire, les plus pauvres et le bien-être » (International Centre for Frugal Innovation, 2023). De même, dans un article du point sur ChatGPT est présenté une alternative à ChatGPT développée par une scientifique sud africaine, Pelonomi Moiloa, qui a voulu démontrer qu'il n'était « *pas nécessaire d'utiliser d'énormes machines gourmandes en énergie et d'investir des millions de dollars pour créer quelque chose de significatif.* » (Le Point & Grallet, 2023).
La France, qui a lancé le train de l'IA frugale il y a plus de deux ans maintenant, se positionne en leader sur ce nouveau type d'IA, qui diffère par tous les aspects présentés ci-dessus des modèles d'IA les plus connus développés par les GAFAM.

Le chapitre suivant s'attache à continuer à présenter le concept d'IA frugale par le prisme des entretiens menés avec les acteurs de l'IA rencontrés. Le paragraphe suivant met l'accent sur la multiplicité des sens que l'on peut accoler au concept d'IA frugale.

3. Définition de l'IA frugale par nos entretiens

Dans le cadre des entretiens que nous avons menés, nous avons cherché à cartographier le concept d'IA frugale pour différents acteurs : administrations publiques, établissements publics, associations spécialisées en IA, recherche académique, et villes innovantes. Notre première approche a été de parcourir la plateforme interministérielle de formation Mentor du Ministère de la Transformation et de la Fonction Publiques (Plateforme interministérielle Mentor de formation, 2023) dédiée aux agents du service public. A ce jour, il n'y a qu'une seule session dédiée à l'intelligence artificielle, « Objectif IA – Initiez-vous à l'intelligence artificielle » d'une durée de six heures, qui n'aborde que très simplement le concept d'IA par différentes définitions et une approche historique de son développement. Aucune référence n'y est faite de la SNIA, et des actions de l'administration publique pour l'IA, de même que le terme d'IA frugale n'y est pas mentionné (on note la même absence pour les autres démarches deeptech soutenues dans la SNIA 2). Une référence discrète peut y être trouvée dans une rapide description de ce que nous avons appelé l'écosystème d'une IA :

« *Bien que les algorithmes d'IA soient très énergivores, de nombreuses applications permettent d'améliorer les chaînes de production ainsi que notre consommation de ressources et d'énergie.* » Nous avons donc commencé dès le début de notre projet à mener des entretiens pour cerner l'utilisation de ce concept par les acteurs de l'IA.

Tout d'abord, la majorité des acteurs rencontrés sont familiers de ce mot, même si nombre d'entre eux considèrent que c'est plus un «mot valise» imposé par une volonté politique forte provenant du Président de la République, plutôt qu'un concept émergeant des acteurs de l'IA. En effet, plusieurs interlocuteurs nous ont signalé que l'IA frugale n'était pas une priorité, et que l'important est plutôt le niveau de performance. Dans le cadre de la GPAI, la collaboration internationale sur l'IA qui est portée par la France et le Canada, le terme d'IA frugale ne fait pas l'unanimité. Si nous prenons l'exemple de la création d'une IA pour réaliser une tâche administrative, l'important est ici plus la précision de l'analyse réalisée par l'IA plutôt que sa frugalité. Il y a aussi la nécessité que le modèle puisse « [passer] à l'échelle », c'est-à-dire fonctionner sur des données de tout un territoire alors qu'initialement le modèle n'a été testé que sur un petit ensemble de données plus localisées. Parfois, imposer la frugalité d'un système d'IA peut empêcher ce passage à l'échelle. De plus, d'autres soulignent que le mot « IA frugale » n'est pas utilisé en soi, car c'est encore un concept peu clair, un « mot parapluie », qui a émergé du fait de la médiatisation croissante des avancées en IA. Ce terme est utilisé parfois dans un jeu de dupes, car il ne fait pas référence à un type de technique informatique et mathématique mais plutôt à une question. Certains préfèrent parler de « soucis écologiques » en lien avec l'IA que d'IA frugale. Même parmi les acteurs d'administration publique que nous avons rencontrés, certains considèrent que le terme d'IA frugale est encore conceptuel, et que l'administration publique doit travailler à mettre en place et diffuser une définition de ce dernier auprès des acteurs techniques. D'autres décrient un terme à la mode, mais qui pose les bonnes questions, comme celle de « minimiser notre empreinte tous azimuts ».

Entrons plus dans le détail de la définition d'IA frugale qui nous a été donnée par les acteurs. Une majorité d'acteurs nous ont parlé tout d'abord d'une IA frugale comme économe en ressources et en énergie. Un interlocuteur nous donne la définition précise d'une IA frugale en données et en modes de calcul, qu'il faut analyser par un *green algorithm* (un algorithme vert, pas forcément celui développé par la société homonyme) pour vérifier son faible impact environnemental. Pour un acteur davantage positionné dans

la mise en oeuvre concrète de solutions d'IA, c'est une IA qui essaie d'utiliser moins de ressources « tout simplement », et qui n'est développée que si nécessaire (en comparaison avec d'autres méthodes informatiques ou ressources humaines) en calculant un « ratio » entre ce que ça coûte et ce que cela permettrait. D'autres complètent en expliquant qu'en plus d'une IA moins gourmande en capacité de calcul, il faut également contrôler la gourmandise en données et volume de calcul de la phase d'inférence et réguler le nombre d'inférence demandés à une IA. En particulier, un interlocuteur partage la définition consensuelle d'IA frugale, en soulignant la nécessité de penser à la frugalité du côté hardware comme du côté software. Nous avons découvert par nos entretiens que des acteurs intégrés dans l'écosystème de l'IA en France sous forme associative ou de start-up souhaitent intégrer dans leur réglementation interne la notion de frugalité dans un objectif de minimisation de l'impact environnemental des projets d'IA. Un intervenant préfère découper la définition en deux volets: une IA économe en énergie et en matière première qui ne porte pas de préoccupations environnementales en soi, et une IA pour l'environnement. Un autre s'intéresse surtout à l'architecture d'un modèle d'IA frugale qui, s'il y a peu de données disponibles, doit être capable, notamment en se fondant sur des modèles antérieurs déjà développés, de réaliser son apprentissage; et à l'inverse, en présence de beaucoup de données, d'extraire l'essentiel de l'information pour réaliser sa tâche. Il note que si un algorithme est entraîné sur peu de données, contrairement à ce que l'on peut penser, cela peut nécessiter davantage de calculs pour entraîner le modèle. Il insiste aussi sur la nécessité d'une architecture frugale à toutes les étapes du système d'IA (dans l'architecture même du réseau de neurone, mais également pour l'inférence).

Si nous reprenons les faisceaux de caractéristiques apposés au concept d'IA frugale qui ressortent de l'analyse bibliographique, nous en retrouvons certains dans les définitions issues des entretiens :

- Pour une IA frugale et écologique, nous avons déjà relevé une adhésion. C'est également le cas d'un interviewé, qui note que les entreprises qui développent l'IA frugale aujourd'hui sont poussées par une sensibilité vis-à-vis de l'environnement. Etonnamment c'est du côté de l'administration publique, porteuse de la SNIA, que certains soutiennent qu'une IA frugale n'était pas forcément écologique, en citant l'exemple de la start-up française Mistral qui développe un modèle d'IA générative similaire à ChatGPT, restant un très gros système d'IA par définition.

- Nous avons également remarqué dans l'analyse croisée des entretiens cette idée commune que l'IA frugale doit s'accompagner de processus de mesure d'impact et de coût d'un système d'IA, que ce soit son coût environnemental, économique ou « informatique » (volume de données, durée d'apprentissage et nombre d'inférences). Il est nécessaire d'évaluer a priori les coûts d'une solution de type IA pour y préférer d'autres solutions plus simples et ne pas vouloir faire de l'IA à tout prix. Cependant, un interlocuteur souligne la pluralité de métriques pour estimer l'impact d'un système d'IA, sans métrique commune définie pour l'instant. Des experts détaillent que le terme frugal doit être utilisé en valeur absolue : le deep learning ne sera jamais fondamentalement frugal, mais en comparatif « plus frugal que ». C'est ce que soutient également un autre intervenant : il faut comparer le coût énergétique et environnemental d'un projet avec les bienfaits que ce projet apporterait pour le développement durable.
- Quant à l'IA embarquée, elle est directement reliée à frugalité du système d'IA car la frugalité permet de créer un modèle compact qui pourra analyser les données issues d'un petit nombre de capteurs, et assure également sa réplicabilité. Nous retrouvons cette recherche de compacité du côté de l'administration publique, qui a travaillé sur des modèles plus légers pour pouvoir les faire tourner sur des ordinateurs portables. C'est nécessaire car il n'y a pas d'infrastructure étatique qui permette de faire tourner des IA, à part le supercalculateur Jean Zay qui est dédié à la recherche. Pour autant, un acteur attire notre attention sur le risque d'effet rebond en associant l'IA embarquée à la frugalité, car on va aussi vers une multiplication des applications embarquées (développement des objets connectés), et donc vers une potentielle hausse de la consommation énergétique associée.
- Passons maintenant au sujet de l'accessibilité, de la transparence et du partage des systèmes d'IA. Pour le public, il est nécessaire de réfléchir à la réutilisation de modèles existants, dans une approche de type open source, pour ne pas réentraîner des systèmes d'IA déjà existants et donc réduire l'empreinte carbone de l'IA. De même, il faut rendre les modèles d'IA accessibles à plus de personnes, et ne pas faire comme les modèles « géants » mais obscurs déployés par les GAFAM. Cela rejoint l'objectif du projet DIAT porté par l'Ecolab. Certaines associations et start-ups portent également cette volonté de créer un espace de partage de données sécurisé commun à différents acteurs de l'IA, où chaque acteur pourrait mettre à disposition ses données à un autre, contre rétribution. Cela permettrait d'abord d'éviter le

stockage des mêmes jeux de données sur différentes infrastructures, et ainsi économiser l'utilisation d'emplacements de stockage qui coûtent en énergie et en métaux rares. Un interlocuteur de la recherche insiste sur un autre aspect de la gestion des données à prendre en compte : celui de la confidentialité et du respect de la vie privée. C'est également ce que soutient un de ses confrères en indiquant qu'il faut développer des systèmes d'IA dans lesquels on ait « confiance », à savoir des systèmes robustes, interprétables, exploitables facilement et qui respectent la confidentialité des données. C'est également grâce à ces petits modèles d'IA « agile, maniable et déployable » que l'IA pourra se démocratiser et devenir un outil pour tous. Ici, nous retrouvons un lien direct entre IA frugale et l'éthique de l'IA. Pour un autre chercheur, la frugalité d'une IA permettrait de réduire le temps de pré-traitement de la donnée. Il souligne qu'aujourd'hui, ce pré-traitement se fait de manière générale en délégant à des prestataires dans des pays en développement qui engagent des travailleurs peu qualifiés et mal rémunérés pour réaliser la labellisation des grands jeux de données. De plus, un modèle d'IA frugale se devrait d'être robuste à des erreurs de labellisation, ce qui permettrait de réduire significativement le coût énergétique, humain et temporel du prétraitement.

Deux acteurs vont plus loin et suggèrent de réfléchir plus largement à toute innovation qui deviendrait frugale. Un acteur observe cela dans sa structure, pas seulement pour l'IA mais pour toutes les modélisations mathématiques et physiques utilisées. Les associations réfléchissent également avec cette approche plus globale, en intégrant un « principe d'innovation soutenable » dans des chartes partagées.

Nous avons découvert un nouvel enjeu de l'IA frugale qui n'apparaissait pas explicitement dans notre étude bibliographique mais qui est apparu en filigrane dans les entretiens : celui de la compétitivité et de la stratégie. L'IA est le domaine scientifique le plus stratégique aujourd'hui, et en cela, développer l'IA frugale c'est « réussir à faire exister d'autres pans que ceux que [les GAFAM] peuvent porter. » De plus, cet acteur explique que la frugalité permet au contenu scientifique du système d'IA d'être plus intéressant : comme la frugalité impose de ne pas faire de modèles avec de grandes quantités de paramètres qui pourraient, comme le font des modèles de deep learning, apprendre une tâche sans même avoir besoin de modèle d'algorithme intégré mais juste en absorbant un énorme volume de données, un modèle d'IA plus compact intègre nécessairement des considérations

scientifiques (mathématiques, physiques, informatiques), donc une architecture préétablie complexe (par exemple ce que l'on appelle le Physics Informed Machine Learning) pour que l'apprentissage se fasse sur un plus petit nombre de données et de manière plus efficace. Un intervenant de la recherche trouve également que l'IA frugale a de belles perspectives scientifiques à apporter, notamment en s'intéressant au biomimétisme en notant que : « pour une tâche cognitive simple comme reconnaître un élément dans une image, on estime qu'une machine consomme quatre à six ordres de grandeur d'énergie de plus qu'un cerveau humain ! ». Pour un confrère de ce dernier, s'investir dans le développement de l'IA frugale est une dynamique « plus raisonnée, voire nécessaire ».

Cependant, des critiques sont également exprimées face à l'usage multiplié du terme IA frugale. Même si certains soutiennent que les acteurs s'engageant dans l'IA frugale ne le font pas par « avantage économique ou concurrentiel », d'autres expliquent la prévalence de raisons économiques. Par exemple, une structure peut décider de mutualiser les infrastructures de calcul à destination des agents, et ainsi de réduire la quantité de serveurs et ordinateurs personnels, et cela pour des raisons économiques et non pour des raisons de réduction d'empreinte carbone. De même, un intervenant de l'administration publique explique que la frugalité n'est pas un but recherché dans la mise en place de systèmes d'IA pour l'administration, mais qu'on peut qualifier un système d'IA comme frugal par construction d'une infrastructure informatique au coût raisonnable (utilisation d'un cloud plutôt que de serveurs spécialisés plus onéreux). Un expert craint même que certains acteurs se saisissent de termes comme IA frugale ou soutenable juste pour faire de l'écoblanchiment. Il ajoute que la recherche académique française s'investit aujourd'hui dans l'IA frugale non par intérêt pour ce domaine, mais car c'est un sujet qui obtient des financements de recherche conséquents. Ainsi, la recherche en IA frugale est en fait issue de choix politiques. En dernier lieu, un acteur de la recherche soutient que parfois, comme frugalité et précision / performances sont inversement corrélées, le choix ne pouvait pas se porter systématiquement sur une solution d'IA frugale. C'est également ce qu'a souligné une intervention de l'administration publique.

Ainsi, cette première analyse des entretiens fait apparaître une définition de l'IA frugale équivoque, avec des faisceaux qui se recoupent avec l'analyse bibliographique, et explore d'autres considérations, notamment des critiques, que nous n'avions pas appréhendées

par l’analyse bibliographique.

La prochaine partie de ce rapport permet d’appréhender la mise en œuvre de l’IA dans quatre études de cas et d’en tirer des enseignements sur la frugalité vue à leur niveau, puis l’impact potentiel de la réglementation récente ainsi que les considérations éthiques qui peuvent accompagner les projets d’IA.

II. Etat des lieux: l'IA frugale dans l'administration

Comme indiqué dans la partie précédente, dans le cadre de notre étude, nous avons été amenés à nous entretenir avec différents acteurs sur la définition de l'IA frugale. Plus concrètement, sont présentés ci-dessous des projets d'IA mis en œuvre par l'administration pour illustrer dans quel cadre la frugalité s'inscrit. Les échanges menés lors des entretiens nous ont amenés à nous interroger sur la réglementation et l'éthique qui encadrait l'IA. L'état de ces réflexions est présenté dans une deuxième partie.

1. Etudes de cas menées dans l'administration

i. Méthodologie

Notre travail d'études de cas n'a pas visé l'exhaustivité des projets conduits dans l'administration mais d'appréhender la diversité des projets. Cela était également réaliste considérant le temps imparti, avec le fait que nous n'aurions pas pu conduire des entretiens avec l'ensemble des porteurs de projets.

Le champ des entretiens conduits a été volontairement restreint aux missions correspondantes à celles des IPEF, par exemple l'aménagement, l'agriculture, l'environnement.

Les entretiens ont été conduits selon une trame d'entretien qui se trouve en annexe de ce rapport (Annexe 1). La trame d'entretien correspondait plus à notre besoin que le questionnaire. Celui-ci peut se révéler trop directif et ne pas s'adapter au cadre de pensée spécifique des porteurs de projet dont nous recueillons le témoignage. Les trois phases de l'entretien portaient sur :

- Les projets menés par l'administration concernée, les compétences mobilisées ;
- Le rapport à l'IA frugale : quelle définition en donne l'acteur concerné ? Est-ce un concept qui fait sens dans son projet d'implémentation de l'IA ? Comment est-ce opérationnalisé le cas échéant ?
- Les liens et les attentes des porteurs de projet vis-à-vis de la recherche sur l'IA frugale.

ii. Les études de cas réalisées

Quatre études de cas sont présentées :

- Une première dans un établissement public où les contraintes interrogent sur les possibilité d'utilisation de l'IA frugale;
- Une deuxième au niveau d'une collectivité locale qui a répondu à l'appel à projets DIAT;
- Les deux dernières au niveau d'EPA qui abordent la frugalité de façon différente par rapport aux enjeux propres à leur établissement.

Détermination de l'usage des sols agricoles par l'ASP, sous contrainte réglementaire

La première étude de cas présentée concerne l'Agence de Services et de Paiement, le plus important organisme payeur européen. La Commission Européenne a exigé de l'ensemble des organismes payeurs la vérification exhaustive et systématique de l'utilisation de leurs parcelles par les exploitants dès la campagne d'aides 2023 pour les aides surfaciques (26 millions d'hectares sur 10 millions de parcelles).

L'Union Européenne met à disposition pour ce faire les données satellitaires de Sentinel, les images en infrarouge et radar régulières de l'ensemble du territoire.

Le paiement des aides de la PAC correspondant à des montants élevés, un enjeu de 9 milliards d'euros, les aides étant auditées régulièrement par la Commission Européenne, avec le risque d'un apurement (refus de la CE de rembourser l'État-Membre des sommes engagées) si les erreurs sont trop nombreuses, la fiabilité du système doit être maximale.

L'algorithme permet de "reconnaître" 73 classes de production, comprenant les 140 cultures qui peuvent être déclarées par les exploitations agricoles. Avec la généralisation du droit à l'erreur pour les exploitations agricoles, le système doit pouvoir déterminer la culture avant la mise en œuvre des contrôles, afin de permettre à l'exploitant de corriger sa déclaration si besoin sans subir de pénalité. Le fait que le travail de détermination des cultures doive être fait par des agents si l'IA ne le fait pas implique que la qualité de la détermination ainsi faite doit être maximale.

Le système doit également être robuste aux éventuelles difficultés : images non reçues d'un des satellites, présence de nombreux nuages lors de la prise des clichés, …

Le développement et la mise en œuvre de l'IA ont été confiés aux prestataires CERCO et Capgemini, après une première phase d'expérimentation avec l'IGN qui a des

compétences en propre. Le prestataire Capgemini est déjà en charge du système d'information de paiement des aides de l'ASP. L'organisme payeur nous a témoigné une réelle difficulté pour des compétences en interne. En effet, depuis 2018, la direction en charge du projet est passée de deux personnes compétentes en IA géographique à quatre, contre vingt-cinq personnes en charge du système informatique, donc pas que l'IA, chez le prestataire. Deux thèses ont également été financées sur le sujet, afin d'aider à ce développement. Les compétences en interne sont cruciales pour challenger le prestataire et avoir un regard critique sur ce qui est proposé.

Ce travail a été fait avec pour objectif d'optimiser l'utilisation de serveurs afin de permettre un passage à l'échelle de l'ensemble du territoire : utilisation de serveurs classiques sur le cloud, avec uniquement des CPU (Central Processing Unit, Unité Centrale de Traitement) et non des GPU (Graphic Processing Unit, processeur graphique ou unité de traitement graphique)[1]. L'infrastructure a été optimisée au regard du besoin : étant réalisés sur du CPU, les calculs ont été le plus possible parallélisés pour qu'ils soient optimisés. De plus, le recours à un cloud public permet que la puissance de calcul soit réallouée à un autre utilisateur quand l'ASP ne l'utilise pas. Cependant, l'importance des conséquences potentielles d'erreurs (conséquences financières et temps d'agent) ne permet pas d'avoir de marges de manœuvre pour favoriser un modèle plus frugal. Et il faut relativiser l'impact de cette IA par rapport au stockage de l'ensemble des images et des données exigées sur le paiement des aides de la PAC. De plus, ce modèle n'étant pas destiné au grand public, l'utilisation qui en est faite est bornée aux exigences réglementaires, ce qui limite son impact. Pour la personne interviewée, il faut bien cerner les domaines de l'administration où la frugalité fait sens.

Ce projet n'est pas le seul projet dans le champ de l'IA géographique : un appel à projets est en cours pour l'utilisation de l'IA autour de l'irrigation, le projet NIVA - Horizon 2020 permet de déterminer ainsi des indicateurs environnementaux comme le stockage de carbone. Un autre organisme payeur en charge des autorisations de plantation viticole, FranceAgriMer, attend le retour d'expérience de l'ASP pour éventuellement se lancer.

[1] Le CPU est un processeur tel Intel i9, réalisant des calculs généraux pour des tâches courantes. Le GPU est une carte graphique spécialisée pour des tâches graphiques intenses ou spécifiques tel que l'apprentissage automatique. Techniquement, le CPU peut faire les mêmes tâches que le GPU mais de façon beaucoup plus lente, car ce dernier a le potentiel de faire beaucoup plus de tâches en parallèle. Il est très efficace quand les tâches peuvent être fortement parallélisées.

Projet de la ville de Noisy-le-Grand dans le cadre de l'appel à projets "Démonstrateurs d'IA frugale pour les transitions écologique et énergétique dans les territoires" (DIAT)

En 2022, la ville de Noisy-le-Grand a répondu à cet appel à projet avec pour objectif de réduire la consommation énergétique de ses bâtiments *via* des recommandations d'usage fournies par l'IA grâce au suivi en temps réel de la consommation des bâtiments, afin de déterminer les aberrations de comportement.

L'appel à projet de la Banque des territoires voit son efficacité au travers de trois leviers :

- l'utilisation d'une IA frugale (frugalité en donnée et en énergie), évaluée par la solution open source Green Algorithms ;
- la raison d'être du projet proposé que l'IA appuiera (optimisation et réduction des transports, de la consommation d'énergie, meilleure gestion des déchets, …) ;
- une logique de territoires démonstrateurs, qui permettront ensuite l'évaluation du projet et sa dissémination dans d'autres territoires.

La ville de Noisy-le-Grand a appliqué cette logique de frugalité en prévoyant un minimum de capteurs, afin de ne pas les multiplier inutilement, une remontée d'information dans une fréquence raisonnable et des consignes au prestataire informatique de tester différents modèles afin de déterminer lequel est le plus efficient. L'objectif est également de favoriser la réplicabilité du modèle, sans exiger une organisation trop lourde pour les collectivités qui seraient tentées d'emboîter le pas. Leur vision de la frugalité est double : il faut utiliser la technologie et la donnée "quand c'est nécessaire" et non "pour le plaisir"; et il faut calculer le ratio entre le bénéfice apporté et le coût associé.

D'autres projets sont en cours dans la ville: l'optimisation des tournées des camions poubelles grâce à des capteurs présents sur certaines de celles-ci, avec encore une fois l'objectif de ne pas déployer des capteurs de façon exhaustive, et la prédiction du nombre d'élèves à la rentrée à partir des données de construction et de vente d'habitations.

Ils ont dû passer par un prestataire et un appel d'offres, n'ayant pas la compétence en interne, car même si la ville organise des formations pour ses agents, peu d'agents sont formés et cela ne fait pas d'eux des experts en IA. L'objectif de ces formations est plutôt une sensibilisation à la donnée, à la cybersécurité.

Il y a un véritable enjeu des compétences pour les collectivités qui ne peuvent se permettre chacune de recruter un expert en IA ou en données : c'est l'EPCI qui est en général le

niveau pertinent pour cela (intervention des compétences sur l'ensemble des projets de l'EPCI ou des communes) ou par exemple l'Autorité Organisatrice des Mobilités pour les problématiques de transport.

Une des questions qui se pose, cependant, est le modèle de diffusion et la souveraineté : doit-on aller sur des données propriétaires, ou open-source ?

Concernant le lien avec la recherche, enfin, la mairie a développé des partenariats avec un laboratoire de l'université Lyon II (ville régénératrice) et avec l'université Gustave Eiffel en lien avec deux masters sur des projets d'un an (data lab). Cependant, le temps de la recherche n'est pas celui des besoins des collectivités et même s'ils sont favorables à ces partenariats, cela peut présenter des difficultés (de même que l'utilisation de l'anglais dans les appels à projets européens, pour la relation avec les laboratoires participants).

Météo France : le cas d'un établissement avec des compétences très internalisées :

Météo France compte deux structures qui font de l'intelligence artificielle :

- Le Lab IA, qui correspond au choix de l'établissement de conserver des compétences en propre, est une structure de quatre agents permanents, à vocation opérationnelle, c'est-à-dire la résolution de problèmes concrets, utilisant de nombreux outils de l'IA : machine learning, régressions, réseaux de neurones, deep learning, avec également des IA plus "légères" si on considère l'entraînement et l'inférence.
- Le Centre National de la Recherche Météorologique, qui va expérimenter l'IA, en particulier les réseaux de neurones, à des problèmes de météo.

Au-delà de ces deux structures, d'autres agents de l'établissement travaillent sur l'IA mais sans réel lien avec les deux structures ci-dessus. Ils ont donc créé le Club deep learning, interne à l'établissement, afin de favoriser le partage sur les pratiques.

Ils ont de nombreux projets impliquant l'IA. Par exemples, le Lab IA cherche à estimer les précipitations avec des réseaux de neurones convolutionnels et des données satellitaires, la prévision immédiate à partir de l'évolution de grandeurs physiques, et à automatiser la prévision météo en tentant de copier l'expertise des prévisionnistes avec des neurones convolutionnels ou legion transformers . Pour le CNRM, un gros projet est de se substituer aux modèles physiques avec des réseaux de neurones entraînés : c'est un projet qui consomme de gros volumes de données, et nécessite des serveurs avec des cartes graphiques.

Une spécificité de Météo France est que l'établissement fait très peu appel à de la sous-traitance : les ingénieurs du Lab IA, les doctorants et chercheurs du CNRM ont des compétences en IA. Au Lab IA par exemple, ils combinent des profils de contractuels experts en IA avec des profils d'ingénieurs météo qui ont la compétence IA et la compétence des métiers de la météo. Cela permet de limiter grandement le coût d'entrée qu'aurait le fait de faire appel à un sous-traitant. Cependant, pour plus d'optimisation et d'efficacité, ils auraient intérêt à collaborer davantage avec des experts IA et optimisation, comme par exemple avec l'Université de Pau Pays de l'Adour qui travaille sur Green IA par exemple, avec qui une collaboration est déjà mise en place mais pourrait être approfondie.

Concernant les besoins en *hardware*, Météo France disposent d'une infrastructure dédiée, avec des noeuds de calcul et cartes graphiques sur le supercalculateur, et ont accès au European WeatherCloud (EWC), un accès à des GPU et du stockage sur lequel tous les pays partenaires ont droit à des quotas. Ils font seulement appel au CNRS pour du travail sur les réseaux génératifs.

Afin de limiter l'impact de leurs modèles, Météo France a beaucoup recours à du profiling de code, c'est-à-dire de l'optimisation de code. En passant 10 à 20% de leur temps de travail à faire cela, ils peuvent diviser par cinq ou dix le besoin de leurs modèles. Actuellement le Lab IA essaye de mettre en œuvre les conclusions d'un article de recherche qui permet de réduire de 75% un réseau de neurones pour le même résultat. Cependant, dans l'établissement, tout le monde n'est pas aussi sensible à l'optimisation ; certains des agents s'attachent à la performance uniquement. Au-delà de cette question d'optimisation, il y a la question de ne pas lancer certains projets qui seraient trop consommateurs en ressources au regard de l'attendu, mais c'est une question qui ne concerne pas que l'IA mais bien tous les modèles physiques de Météo France.

Cela passe par des formations à des agents moins experts sur le codage, des échanges de bonnes pratiques, par exemple avec le CNRM ils favorisent la réflexion collective autour des projets : "le thésard arrive avec son projet de code et on réfléchit", la mise en place d'ateliers thématiques avec l'Université de Pau.

Un levier très important pour la recherche et l'optimisation des codes est l'existence de librairies *open source* proposant des briques de codes. Elles peuvent être utilisées par de nombreux chercheurs et ingénieurs, et ont également de nombreux contributeurs, ce qui permet de corriger les contributions de chacun par le savoir-faire des utilisateurs si besoin.

Ces projets open source s'imposent d'eux-mêmes du fait de leur qualité, le travail qualitatif lié à l'existence d'une communauté autour de ces projets en fait un élément incontournable.

Et l'existence de ces bibliothèques est cruciale, ils n'auraient pas le temps de tout recoder.

Cependant, des marges de progrès peuvent encore exister, dans l'administration il y a parfois trop de fonctionnement "en silo", il faudrait faciliter le travail entre agents de différentes directions, le partage d'expérience, la création de code ensemble. Il faudrait "rassembler les faiseurs", entre différents établissements utilisant l'IA, et pas uniquement des "réunions de chefs". Il ne faut pas que de très bons ingénieurs soient "micro-managés".

Concernant l'optimisation et la frugalité, cela pourrait passer par des unités d'enseignement à l'école de la météo également. Sans oublier la formation continue, pour sensibiliser les agents de tout âge.

L'IGN, un établissement à la croisée de nombreux enjeux et de nombreuses données

Les nombreux enjeux de suivi de l'occupation et de la couverture des sols, en particulier suite à l'impulsion du "Zéro Artificialisation Nette", et l'immensité du territoire concerné fait que pour l'IGN, l'IA est un outil incontournable. Un fort renforcement de l'établissement a été réalisé du fait de ces enjeux, avec notamment la structuration d'une feuille de route de l'IA en 2022 et un fort effort de recrutement de compétences.

L'IGN a l'avantage également de disposer de laboratoires de recherche lui permettant de gagner en expertise sur ces sujets. Que ce soit pour la recherche ou le développement de solutions, ils ont choisi de faire appel à des compétences en propre, même si certains travaux ne peuvent être que sous-traités.

Afin de faciliter ce travail de sous-traitance et de partager avec d'autres administrations, ils ont créé le réseau DATALLIANCE, permettant aux administrations d'être aiguillées et conseillées pour évaluer la crédibilité technique des entreprises, en particulier au sujet de l'IA. Ils travaillent également en partenariat avec de nombreux acteurs : le Cerema, INRAE, l'ONF, d'autres laboratoires de recherche, et avec la DINUM et Ecolab du MTE.

Un autre enjeu réside dans le changement législatif induit par la loi Le Maire en 2016 pour l'établissement : d'un modèle de subventions publiques et de vente de données, l'objectif de l'établissement est maintenant de favoriser l'ouverture de données d'intérêt général. Cela s'accompagne également en favorisant les projets portés par d'autres acteurs utilisant des données cartographiques (par exemple la détection de panneaux solaires). Ils ont pour

cela monté les challenges techniques et scientifiques FLAIR, donnant accès à des jeux de données d'apprentissage de qualité, vastes et diversifiés, pour faire émerger des modèles et des publications. Ils sont ouverts à la recherche publique, en entreprise, mais aussi aux acteurs du numérique.

L'IGN ne travaille pas forcément beaucoup sur l'IA sobre même si cela fait partie de leurs préoccupations : une des pistes de recherche est par exemple de favoriser les modèles de convolution, avec des données riches, issues de capteurs, en 3D, des séries d'image, … afin de guider les apprentissages et ainsi gagner en efficacité. Le hardware, est aussi un sujet en réflexion car à l'origine chaque agent travaillait sur sa propre machine, puis il y a eu des moyens partagés. Ainsi le travail actuel est de créer une offre de service mutualisée à l'intérieur de l'établissement. Cet objectif entre autant dans une logique de capacité mais aussi du fait des tensions sur les marchés d'approvisionnement. L'empreinte écologique n'est pas le principal moteur, il y a des enjeux budgétaires, de souveraineté mais également de temporalité : petits projets tout au long de l'année qui peuvent être supportés en interne, déport extérieur sur les besoins ponctuels les plus intensifs en calcul.

Le deuxième challenge FLAIR intègre la notion de contrainte de sobriété sur les temps d'inférence et d'apprentissage, ainsi que le volume de calcul. L'objectif est par exemple de déterminer les meilleures données pour atteindre l'objectif fixé : forte fréquence mais résolution faible, forte résolution mais fréquence des images faibles, plus ou moins de bande spectrale,… Des réponses à ces questions peuvent permettre de faire avancer la science mais aussi la frugalité des modèles. C'est un progrès même si de nombreux éléments (matériaux, composants, renouvellement du parc, …) ne peuvent pas être évalués.

Il y a également un enjeu de partage de pratiques, ce qui marche, comme ce qui se fait dans le cadre de l'Écolab. Cependant, l'IA concernant les données géographiques et géomatiques n'intéresse pas forcément les autres administrations, et ne sont pas transposables à leurs besoins.

Concernant la formation des acteurs, l'ENSG a développé des formations IA en tronc commun et spécialisation d'ingénieurs, et c'est en réflexion pour l'intégrer dans les cursus de géomètres, à un niveau de licence professionnelle.

iii. Ce que ces études de cas nous montrent sur les projets d'IA dans l'administration :

Les projets d'IA dans l'administration sont très divers et nous ont permis d'appréhender à travers des cas pratiques la place de la frugalité dans ces projets :

- Tout d'abord, la frugalité n'est pas une fin en soi de ces projets, qui répondent tous à un besoin, parfois lié à l'environnement. Cependant, des critères comme dans le cadre de l'appel à projet des Démonstrateurs d'IA frugale (DIAT) ou les critères des projets FLAIR peuvent imposer la frugalité aux projets financés / soutenus ;
- ensuite, les porteurs de projets ne font pas de "frugalité pour la frugalité". Ils prennent en compte les contraintes techniques qui sont les leurs et vont optimiser le modèle développé afin qu'il soit compatible avec les moyens à disposition (*hardware*, espace de serveur, temps de supercalculateur, …) dans un contexte de difficulté à se procurer du matériel et de sobriété budgétaire ;
- Enfin, si le besoin impérieux d'une solution existe, et qu'elle n'est pas frugale, elle sera développée quand même.

Au-delà de la question de la frugalité des modèles, des problématiques des administrations sont ressorties lors des entretiens :

- La question des compétences, la difficulté de recruter des ingénieurs en IA dans la fonction publique, l'arbitrage entre faire et faire-faire. Ce sont des problèmes pratiques qui vont également avoir un impact sur ce qu'une administration peut faire et ce qu'elle ne peut pas faire ;
- La question du partage des compétences et des avancées, centrale dans tous les projets, prend des formes très diverses : partage d'expériences entre États-membres de l'Union Européenne, diffusion des résultats d'un appel à projets, participation à des bibliothèques de code open source… Dans une logique d'économie de moyens, de temps d'expert réduit.
- La question de la formation : formation initiale, formation continue, recrutement de jeunes ingénieurs, d'ingénieurs plus expérimentés, et la complémentarité avec les compétences plus techniques des administrations concernées.

2. Réglementation et éthique

i. L'IA frugale, un champs pour l'instant éloigné des réflexions actuelles en matière de réglementation

La réglementation représente actuellement un sujet majeur, mais sur l'IA dans sa globalité. Il n'y a pas vraiment de règlement sur l'IA, ni aux USA (même si un cadre serait en réflexion au sein de l'administration Biden), ni en Chine. L'Europe souhaite être précurseure avec l'AI act ou RIA (en français : réglementation pour l'IA). Une phase de trilogue a récemment eu lieu entre la commission européenne, le parlement européen et le conseil de l'UE (L'Agefi & Cousin, 2023). Après trois mois de négociations, les Etats membres et le Parlement européen ont trouvé un « accord politique » le 8 décembre 2023 sur un texte qui doit favoriser l'innovation en Europe en matière de développement de l'intelligence artificielle. Cependant, cela pose la question d'une réglementation européenne dans un cadre international et global au sein duquel seul ce territoire applique cette réglementation.

Quatre niveaux de risque doivent être identifiés pour les systèmes d'IA (Figure 12). En fonction de la classification du risque associé, il y aura soit interdiction de la mettre sur le marché, soit un certain nombre de contraintes à respecter pour que l'IA produite soit mise sur le marché (ou continue d'être mise sur le marché). Plus précisément, la proposition de la commission européenne sur les différents risques en matière d'IA est la suivante (Proposition de cadre réglementaire sur l'intelligence artificielle | Bâtir l'avenir numérique de l'Europe, 2023):

- **risque inacceptable :** tous les systèmes d'IA considérés comme une menace évidente pour la sécurité, les moyens de subsistance et les droits des personnes seront interdits,
- **risque élevé :** les systèmes d'IA identifiés comme à haut risque seront soumis à des obligations strictes avant de pouvoir être mis sur le marché. Ils comprennent la technologie de l'IA utilisée dans :
    - les infrastructures critiques (par exemple les transports), susceptibles de mettre en danger la vie et la santé des citoyens;
    - la formation éducative ou professionnelle, qui peut déterminer l'accès à l'éducation et au cours professionnel de la vie d'une personne (par exemple, la notation des examens);

- composants de sécurité des produits (par exemple, application d'IA en chirurgie assistée par robot);
- l'emploi, la gestion des travailleurs et l'accès au travail indépendant (par exemple, un logiciel de tri de CV pour les procédures de recrutement);
- services publics et privés essentiels (par exemple, notation de crédit refusant aux citoyens la possibilité d'obtenir un prêt);
- les services répressifs susceptibles d'interférer avec les droits fondamentaux des personnes (par exemple, évaluation de la fiabilité des preuves);
- gestion des migrations, de l'asile et du contrôle aux frontières (par exemple, vérification de l'authenticité des documents de voyage);
- administration de la justice et processus démocratiques (par exemple application de la loi à un ensemble concret de faits).

- **risque limité :** pour les systèmes d'IA assortis d'obligations de transparence spécifiques. Lors de l'utilisation de systèmes d'IA tels que les chatbots, les utilisateurs doivent être conscients qu'ils interagissent avec une machine afin qu'ils puissent prendre une décision éclairée de continuer ou de prendre du recul;
- **risque minimal :** inclut des applications telles que les jeux vidéo compatibles avec l'IA ou les filtres anti-spam. La grande majorité des systèmes d'IA actuellement utilisés dans l'UE relèvent de cette catégorie.

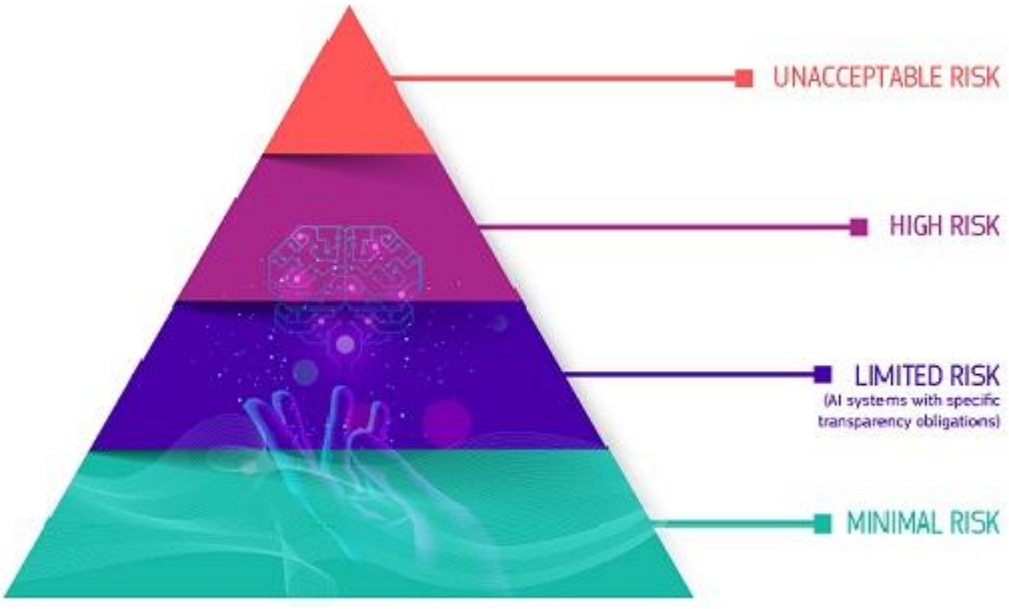


Figure 12 : Quatre niveaux de risque en matière d'IA: inacceptable, élevé, limité et minimal (*Proposition de cadre réglementaire sur l'intelligence artificielle | Bâtir l'avenir numérique de l'Europe*, 2023)

Pour les systèmes d'IA à haut risque, la commission européenne propose de les soumettre à des obligations strictes avant de pouvoir être mis sur le marché :

- des systèmes adéquats d'évaluation et d'atténuation des risques;
- haute qualité des ensembles de données alimentant le système afin de minimiser les risques et les résultats discriminatoires;
- l'enregistrement des activités afin d'assurer la traçabilité des résultats;
- une documentation détaillée fournissant toutes les informations nécessaires sur le système et son objet pour permettre aux autorités d'évaluer sa conformité;
- des informations claires et adéquates pour l'utilisateur;
- des mesures de surveillance humaines appropriées pour minimiser les risques;
- haut niveau de robustesse, de sécurité et de précision.

Les discussions et débats autour de l'AI act se sont par ailleurs complexifiées avec l'arrivée de l'IA générative en février-mars-avril 2023. Les discussions ayant débutées depuis plusieurs mois, il a fallu intégrer la question de l'IA générative et mettre à jour tous les travaux déjà engagés.

Concernant certains secteurs, comme le secteur financier, les enjeux sont tels que la prudence dans sa mise en œuvre par les acteurs doit aller au-delà des exigences de cette nouvelle réglementation.

Quid de l'IA frugale?

Il semble être encore tôt pour traduire le concept d'IA frugale dans la réglementation, parce que la réflexion est en cours pour l'IA elle-même. A l'échelle européenne, l'élaboration de l'AI act apparaît un préalable en matière de réglementation pour pouvoir envisager ensuite une norme ou une réglementation spécifique à l'IA frugale. Par ailleurs, l'IA frugale est beaucoup moins mature que sur l'IA dans sa globalité et ne représente pas une préoccupation majeure étant donné qu'il n'y a pas de pression actuellement de la part des acteurs privés ou publics pour inscrire cette notion dans un règlement.

ii. Des incitations diverses à la frugalité

Si l'étape de la réglementation semble encore éloignée des réflexions autour de la mise en place de l'IA frugale, la question se pose de ce qui peut motiver un porteur de projet à mettre en avant la frugalité.

“*Les entreprises qui poussent actuellement le sujet de l’IA frugale le font plus par sensibilité vis-à-vis de l’environnement, moins par avantage économique ou concurrentiel. Pourtant cette question mérite d’être regardée de façon globale.*”

En premier lieu, un modèle frugal signifie utiliser des modèles plus légers avec un temps d’entraînement significativement réduit. Sur un secteur concurrentiel, une IA frugale peut représenter un avantage de compétitivité si une solution d’IA peut être déployée plus rapidement (par exemple en trois ou quatre mois, contre une année pour une IA classique).

D’autre part, utiliser un modèle léger signifie aussi une inférence moins coûteuse. Le coût à la requête se compte certes en centimes d’euros mais est à considérer de manière globale, par rapport au nombre de requêtes. En effet, ce coût à la requête doit être supporté par une des parties prenantes. Il est soit facturé au client (en prompt, en token, à la requête), soit supporté par le fournisseur et les entreprises. Le recours à une IA frugale peut alors devenir intéressant d’un point de vue économique.

Enfin, un aspect pouvant motiver la mise en avant de l’IA frugale est la réglementation RSE, notamment sociétale et environnementale. En effet, les entreprises considèrent de plus en plus les aspects suivants (i) leurs activités directes, (ii) leurs activités indirectes, support et autres, qui viennent alimenter les activités directes, (iii) toutes les tierces parties, les fournisseurs, les prestataires pour lesquels il est plus difficile d’avoir une vision sur leurs bonnes pratiques.

Pour une entité, pouvoir dire que parmi ses modèles d’IA utilisés au sein de son entreprise, la plupart développent des modèles d’IA frugale peut être un critère qui vient alimenter le critère RSE. Néanmoins, pour l’instant, il n’y a pas de labellisation IA frugale et ces bonnes pratiques se font essentiellement sur le mode déclaratif.

En matière de sobriété, il peut être aussi intéressant de se poser la question de l’utilité ou pas d’utiliser une IA, même frugale, en fonction du domaine d’application. Dans certains cas, le meilleur modèle frugal est celui que l’on n’a pas fait pour une raison qui aurait pu être futile.

“*On peut par exemple faire un très beau modèle, qui est très frugal, pour aider à reconnaître des chats et des chiens, juste pour faire une campagne publicitaire. Mais est-ce vraiment utile de lancer un projet d’IA pour faire cela. Quelle en est la valeur ajoutée ?*”

iii. Un label sur la frugalité?

Un label IA Frugale apparaît à terme comme une solution intéressante, en imaginant un certain nombre de points de contrôle (par exemple une liste de 40 à 50 critères à satisfaire).

Idéalement, si on connaissait les limites fondamentales des compromis entre les ressources à utiliser et la performance qu'on peut en attendre, cela permettrait d'avoir une échelle. On pourrait ainsi dire que pour une performance donnée, telle méthode est dans la consommation minimale de ressource et qu'il ne faut pas utiliser plus de ressource. Cela pose la question de qu'est-ce que le degré de performance légitime dans chaque cas d'usage. Il paraît difficile d'avoir une notion absolue d'IA frugale. Il peut y avoir des notions relatives, mais il y a toujours derrière un curseur.

"J'imagine que la frugalité pour gérer un système de décision sur l'armement nucléaire n'est peut-être pas la même que recommander la boîte de haricots verts que l'on prendra dans notre prochain caddie."

Se dirige-t-on alors vers un nutriscore de la frugalité?

L'IA frugale apparaît à ce jour encore en phase de développement via des recherches et des expérimentations, et encore un peu éloignée de l'introduction d'un label qui est généralement une étape envisageable à partir du moment où le sujet est suffisamment mature. Plusieurs questions se posent en lien avec l'introduction d'un label. Tout d'abord sur ce qu'il contient et sur ce qu'il représente. Cette étape nécessite de faire appel à plusieurs parties prenantes. Se pose ensuite la question de désigner un intervenant neutre de confiance qui va venir auditer le système pour accorder le label. Serait-ce un bureau de contrôle privé qui devra être formé ? Ou bien un laboratoire public comme le LNE ? Il s'agit surtout de désigner un intermédiaire professionnel du sujet et dont la légitimité et la crédibilité ne sont pas remises en cause.

Une fois le référentiel élaboré, il se pose la question de (i) qui porte le label, et de faire appel pour cela à des organismes certifiés indépendants qui puissent auditer et certifier; (ii) la temporalité autour de la vie du label (pour une entreprise auditée et certifiée, est-il nécessaire de renouveler une certification si son modèle dérive ou s'il est réentraîné au bout d'un certain temps ?).

Il sera important également de tenir compte du grand nombre de labels développés ces

dernières années par divers acteurs sur le sujet de l'IA et plus généralement des usages du numérique (numérique responsable, label éthique, IA responsable, Garantie humaine de l'IA…). La DGE pourrait avoir un rôle à jouer pour clarifier l'offre qui existe autour de ce sujet. Pour les destinataires de cette offre, ce n'est en effet pas évident de savoir quel label a quel périmètre, lequel est le plus adapté pour son projet. Certains pensent plus à labelliser une organisation. Pour d'autres, cela peut davantage concerner un projet, i.e. la solution numérique.

iv. Une question d'éthique

Cette partie se fonde principalement sur les éléments recueillis lors de nos entretiens. Ceux-ci étant globalement convergents, ces éléments nous apparaissent comme assez robustes.

L'Etat se doit d'être exemplaire, en matière d'éthique et de transparence, et à ce titre a un rôle à jouer dans le développement d'une IA frugale satisfaisant à des principes de base: la transparence, l'explicabilité, la robustesse technique des solutions d'IA utilisées. Cela a une forte importance notamment vis-à-vis du deep learning, qui peut apparaître à certains égards comme une boîte noire (notamment dans le cas des réseaux de neurones). Certaines start-ups commencent à développer des technologies se connectant au modèle durant la phase d'entraînement, pour le rendre explicable tout au long de son entraînement. Le but est de pouvoir expliquer le cheminement quand la solution d'IA arrive à une décision.

La question de l'éthique est un sujet majeur, d'autant plus si le modèle d'IA est déployé sur des données concernant les concitoyens, ou s'il peut émettre une décision. Il est alors nécessaire d'avoir un humain derrière pour valider ou infirmer la décision. Plus précisément, lorsque le modèle d'IA émet un avis, favorable ou non, cela ne reste que l'émission d'un avis, et l'agent concerné doit alors faire lui-même son analyse sur la base de cette recommandation, pour in fine prendre la décision. Cela signifie d'avoir un modèle transparent, comme expliqué précédemment, pouvant expliquer toutes les étapes qu'il a empruntées, pour amener à ce niveau de recommandation ou de décision.

Le Label éthique Ekitia est par exemple un des labels introduits en France sur les enjeux autour des usages des données. Ce label vise à valoriser les projets respectueux de la Charte éthique des usages des données créée par Ekitia en 2020.

Le LNE a également récemment créé une certification qui permet aux utilisateurs de

disposer de critères de choix objectifs pour faire leur sélection, et aux développeurs de démontrer qu'ils maîtrisent toutes les étapes du cycle de vie d'une IA et répondent aux exigences de performance, réglementation, confidentialité et éthique de leurs clients.

v. Un positionnement stratégique

La souveraineté est une préoccupation forte en matière d'IA globale. Mais en soutenant la souveraineté, plusieurs problématiques apparaissent. Il y a certes de nombreux acteurs IA français et européens qui permettent de produire et d'entraîner les modèles. La question concerne plus les infrastructures utilisées. Y-a-t-il aujourd'hui toutes les infrastructures nécessaires pour entraîner les modèles ? Les entreprises françaises fournisseuses en IA ont-elles connaissance de ces infrastructures et peuvent-elles y accéder ? On peut citer plusieurs supercalculateurs, dont Jean Zay qui a par exemple permis d'entraîner le langage Bloom de Hugging Face. La question se pose par contre de savoir si les entreprises connaissent ces infrastructures ? Y-ont-elles accès ? Et ont-elles intérêt à les utiliser ? Est-ce plus intéressant économiquement que de passer par un programme d'abonnement classique (par exemple Microsoft azure). Est-ce plus pratique ? Y-a-t-il autant de services qu'avec un écosystème bien développé ?

La cybersécurité représente une autre question essentielle. Le fait d'entraîner un modèle d'IA peut ajouter des failles supplémentaires (en augmentant la surface d'attaque). Par exemple, une entreprise souhaitant entraîner un modèle d'IA en ligne sur un cloud le fait généralement sur des plateformes (Amazon web service, Google cloud platform, Microsoft azure). Cela signifie que les données d'entraînement sont mises en ligne pour que le modèle soit entraîné dessus, pour avoir une puissance de calcul suffisante sur le data center. Cela pose le problème, pour une administration ayant des données en interne sécurisées, de les envoyer sur un site externe via un prestataire. Le prestataire a-t-il des pratiques conformes vis-à-vis de la cybersécurité ? Est-il certifié ISO 27001 ? Ce prestataire externe met-il lui-même les données sur une tierce partie via des plateformes en ligne ? Comment sécuriser alors les données par rebond ? Cet aspect de cybersécurité est très important, d'autant plus si l'administration publique veut être vertueuse et montrer l'exemple, en privilégiant des modèles d'IA frugaux.

Pour compléter cette analyse, la prochaine partie de ce rapport interroge le rôle de la recherche en ce qui concerne l'IA frugale et les pistes pour un accompagnement plus soutenu de son déploiement dans l'administration.

III. La recherche peut-elle favoriser l'IA frugale dans l'administration?

Cette partie présente les liens perçus entre la recherche et l'administration sur la base des publications sur l'IA frugale et des entretiens menés, puis propose des voies à explorer pour renforcer les liens existants.

1. Liens entre la recherche et l'administration

Dans la synthèse de son évaluation des politiques publiques sur la stratégie nationale de recherche en intelligence artificielle publiée en avril 2023, la cour des comptes indique à propos de la confiance et de la frugalité: "*La consultation des chercheurs en IA opérée par la Cour montre que ces thématiques sont actuellement peu prises en compte dans les travaux de recherche".* (Cour des Comptes, 2023)

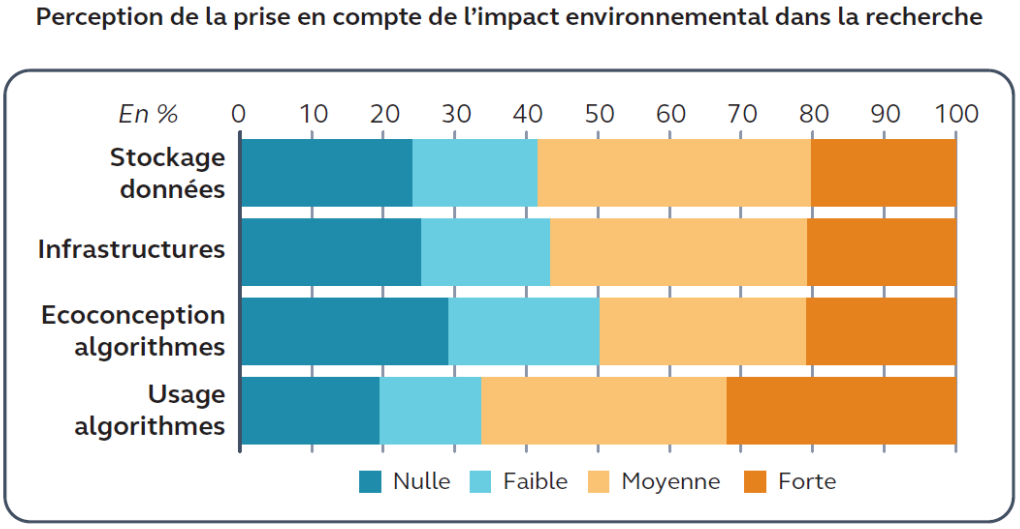


Figure 13 : Perception de la prise en compte de l'impact environnemental dans la recherche

Etalab partage cette analyse car ils ont le sentiment que la frugalité n'est pas la préoccupation première de la recherche. Il y a pourtant un programme stratégique appelé "numérique éco-responsable" qui est très lié à une stratégie d'accélération dans le cadre de France 2030. Ce programme est certes plus large que la seule IA, mais sa mise en place fait partie des priorités du gouvernement.

La deuxième phase du SNIA, débutée en 2022, comporte toutefois dans un de ses trois piliers le soutien à l'offre Deeptech: IA embarquée, IA frugale et IA générative. Cette troisième partie analyse comment la recherche et l'administration interagissent sur le thème de l'IA frugale. Les réponses ne sont pas fermes et figées car l'IA est un sujet d'actualité et la recherche travaille sur une échelle temporelle plus grande. La cour des comptes souligne d'ailleurs le risque d'un financement de la recherche "*via des instruments financiers à courte durée*" qui ne répondent pas forcément aux besoins et tempère également les conclusions de son évaluation: « *En raison du temps long de la recherche, les effets concrets de la stratégie sur la production scientifique ne sont néanmoins pas encore appréhendables de manière fiable.*»(Cour des Comptes, 2023)

i. Les actions liées au SNIA

Dans une interview donnée en avril 2023, Guillaume Avrin, coordinateur national pour l'intelligence artificielle, indique que la première phase du SNIA a servi à se mettre en ordre de bataille et à avoir des écosystèmes prêts à l'action. Présenté comme une réussite, les effets de cette stratégie sont visibles avec le nombre de publications sur l'IA qui continue de croître et le nombre de laboratoires d'IA et de startups spécialisées dans ce domaine en France: 81 laboratoires d'IA et 502 startups spécialisées en 2021(Ministère de l'économie des finances, 2023).

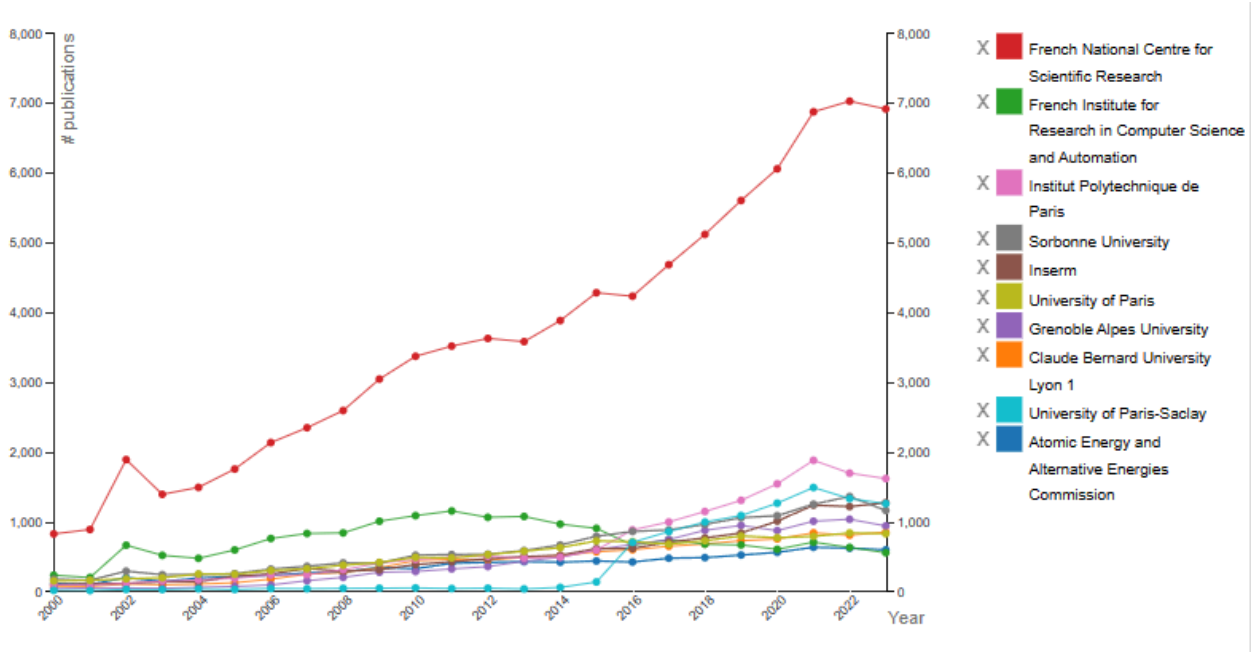


Figure 14 : Nombre de publications en France concernant l'IA *Source :* [AI Strategies and Policies in France - OECD.AI]

Guillaume Avrin positionne également l'IA frugale comme un axe prioritaire dans son interview et le confirme dans sa contribution à la feuille de route Intelligence artificielle et transition écologique: « *Il n'y aura pas de diffusion large et pérenne de l'IA dans notre société et nos entreprises si l'impact environnemental de ces technologies intelligentes n'est pas maitrisé.* » et incite à « *la montée en puissance du sujet de l'IA frugale pour tous les domaines d'application de l'IA* » (B SMART, 2023).

La deuxième phase du SNIA a été construite autour de "*comment influencer la prise en compte de l'IA dans l'économie et donc dans l'écosystème économique français. Durant cette deuxième phase, on va avoir trois différents piliers: Soutien à l'offre Deeptech (IA embarquée, IA frugale, IA de confiance et IA générative), rapprochement offre et demande de l'IA et formation (comment attirer et former le maximum de talents en IA)*." (Coordination de la SNIA, Ministère de l'Economie et des Finances, communication personnelle, 14 décembre 2023)

Financement du deuxième volet de la SNIA – stratégie d'accélération (en M€)

En M€	Prog-ramme de Recherche	IA décent-ralisée et embarquée	IA de confiance	Diffusion de l'IA & démons-trateurs d'IA responsable	Compé-tences et talents	Total
Financement public	134	265	111	259	776	1 545
PIA 4	73	263,5	97,5	123		557
France 2030					700	700
Autres Crédits État et collectivités	61	1,5	13,50	136	76	288
Financement privé		310	105	86	5	506
Union européenne		60	10	16		86
Total	134	635	226	361	781	2 137

Source : Retraitement Cour des comptes d'après le dossier de presse du 8 novembre 2021 et les données issues du coordonnateur national

Tableau 3 : Financement du deuxième volet de la SNIA (en millions d'euros)

Fer de lance de cette stratégie, Guillaume Avrin découpe son action en trois parties: soutien aux champions et à la formation par le biais d'investissement, coordination de l'action au niveau des différents ministères et animation de l'écosystème.

Le lancement du PEPR IA ayant pris du retard (kick-off de lancement fin mars 2024 au lieu de début octobre 2023), il est difficile d'estimer l'effet de l'élan donné par la SNIA concernant l'appropriation de la frugalité par la recherche. Le projet PEPR SHARP se

positionne clairement sur ce sujet. Sur les neufs projets du PEPR, il devrait y en avoir un, maximum 2 autres, abordant également ce sujet. Les entretiens menés montrent qu'au sein même du projet PEPR SHARP, le sens du mot frugalité est interprété de façons diverses. Les chercheurs interrogés sur l'IA frugale ne citent d'ailleurs pas d'eux-même la SNIA dans les échanges. Cela interroge sur la perception de la SNIA comme un ensemble cohérent qui embarquerait les différents acteurs de la chaîne, de la recherche jusqu'aux utilisateurs.

ii. Les appels à projet ou projets

Les appels à projet et les projets autour de l'IA se multiplient.

Aucun projet IA du pôle ministériel Economie Energie Territoires listé dans la feuille de route 2023 ne mentionne de centrage sur la frugalité. La DINUM concentre ses efforts sur l'IA générative et travaille pour trouver le juste équilibre entre performance et frugalité. L'objectif de l'optimisation étant d'abord de faire un outil utile et ensuite d'en optimiser la frugalité.

Placer la frugalité comme priorité a l'air essentiellement destinée aux territoires. Ainsi, les appels à projets DIAT font explicitement mention de la frugalité. Thomas Cottinet, d'Ecolab, le souligne dans le livre blanc de la communauté des acteurs de l'IA en territoires: « *L'intelligence artificielle frugale s'avère un réel atout différentiel parmi les outils disponibles pour les territoires afin d'engager des actions fortes de transition écologique. Les expériences positives de démonstrateurs IA diffusent entre territoires, au-delà de l'appel à projet « Démonstrateurs d'intelligence artificielle frugale au service de la transition écologique dans les territoires » financé par France 2030.* »(Ecolab et al., 2023)

Cet appel à projet porté par Ecolab demande explicitement , dans sa deuxième vague, aux collectivités et aux établissements publics territoriaux de mesurer l'impact environnemental de leur proposition (Ecolab, Ministère de la Transition Écologique et al., communication personnelle, 3 octobre 2023). L'Ecolab motive les participants à s'associer avec des instituts de recherche ou des universités mais ce n'est pas obligatoire. L'appel à projet est toutefois diffusé auprès des laboratoires de recherche pour favoriser leur intégration dans un consortium.

La seconde vague d'appel à projet se terminait le 1er décembre 2023. Pour la première vague, neuf dossiers ont été déposés et quatre ont été retenus. Ces chiffres ont été perçus comme un succès. Sur les quatre projets retenus, deux ont un laboratoire de recherche

dans leur consortium:

- La ville de Noisy-Le-Grand s'est associée avec l'institut Efficacity;
- La ville de Metz s'est associée avec le laboratoire d'Eau et Environnement de l'université Gustave Eiffel de Nantes.

L'objectif est ensuite que ces démonstrateurs soient réplicables sur d'autres territoires, le critère d'interopérabilité est d'ailleurs examiné dans l'analyse des dossiers déposés.

Si la volonté d'associer les territoires avec les laboratoires de recherche dans le cadre des appels à projet lancés par l'Ecolab est louable, elle ne paraît pas assez motrice pour inciter les laboratoires à s'impliquer. Il sera intéressant de voir les résultats de la deuxième vague pour constater si les consortiums entre recherche et administrations sont plus nombreux. Le nombre de dossiers déposés étant trois fois plus élevé que l'année dernière, on peut espérer que ce soit le cas.

Un chercheur avec qui nous avons échangé voyait d'ailleurs la frugalité comme un moyen de rendre l'IA diffusable et démocratique dans le sens où si elle est déployable dans des plus petites structures, elle devient accessible à tous.

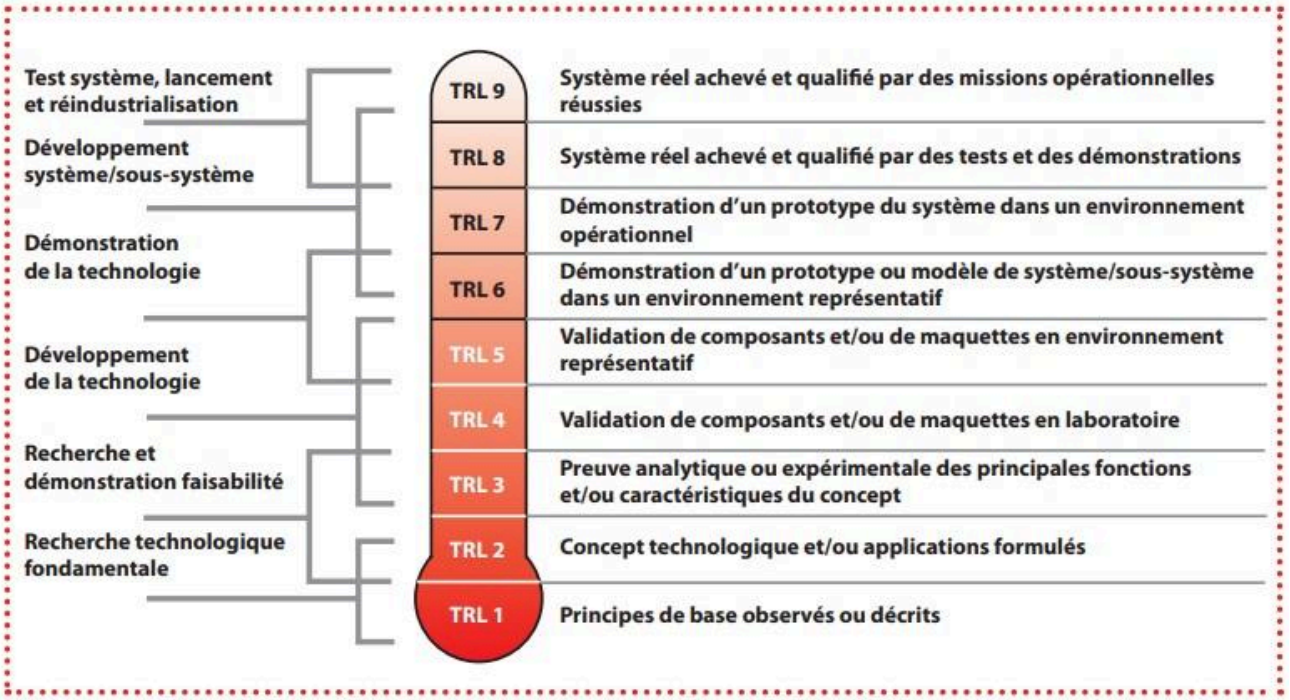


Figure 15 : Échelle TRL (source: technologies-cles-2015-annexes.pdf (entreprises.gouv.fr))

Concernant le lien entre le PEPR IA et le démonstrateur DIAT, un acteur de la SNIA nous indique qu'il n'y a pas à ce jour de réel pont entre ces deux programmes, parce qu'ils sont à des échelles de TRL très différentes (Figure 15). Il pourrait y avoir une animation in fine, mais à ce jour, il s'agit d'une recherche très fondamentale pour le PEPR (TRL de 1 à 3), et de quelque chose de très aval sur le DIAT (TRL 9). Les échelles de maturité sont donc trop différentes à ce stade pour envisager des liens directs. Pour autant, il a déjà organisé dans le cadre du PEPR une réunion qui rassemblait un certain nombre d'industriels et d'entreprises pour connaître les directions stratégiques de la recherche fondamentale en IA, dans le cadre de la stratégie, et pour avoir un échange. L'acteur indique que cette réunion avait été jugée très intéressante. Ce type d'initiative a été demandé aux pilotes du PEPR IA (Inria, CEA, CNRS) pour qu'il puisse y avoir des mécanismes de consultation des industriels. C'est-à-dire être en capacité de faire connaître les projets de recherche amont. Les chercheurs ont ainsi une visibilité à long terme sur les cas d'usage potentiels, et sur les besoins des entreprises. Et en même temps, les entreprises sont au courant des différents thèmes de recherche dans les instituts de recherche publique. Les échelles de temps sont très différentes, mais des applications peuvent être identifiées in fine, même si elles ne se réalisent pas dans l'immédiat. Il y a également un certain nombre de recherches amont qui ne donneront peut-être pas lieu à des solutions, ou qui resteront peut-être juste de l'ordre de la recherche sans forcément mener à une innovation. Ce n'est pas forcément évident, mais des pistes de liens sont quand même étudiées dans le cadre de la SNIA.

Enfin, la frugalité peut également s'imposer dans certains projets sans que cela soit l'objectif premier pour des contraintes techniques ou par manque de données. En effet, un chercheur soulignait que parfois l'IA avait un côté frugal du fait du manque de données pour entraîner le modèle. Par ailleurs, le deuxième volet du SNIA met l'accent sur l'IA embarquée. Guillaume Avrin souligne que « *Les modèles d'IA frugaux sont également essentiels pour l'IA embarquée, un autre axe de notre stratégie, afin de pouvoir faire tourner des algorithmes localement sur tous types de véhicules* » (Ministère de la Transition Écologique et de la Cohésion du Territoire, 2023). Un chercheur nous témoignait cependant du risque à associer l'IA embarquée à une forme de frugalité, du fait de possibles effets rebond (si on regarde le développement des objets connectés, on est plutôt sur une augmentation de l'impact environnemental qui fragilise d'une certaine manière l'association de l'IA embarquée à une forme d'IA frugale).

iii. Un lien plus important chez les opérateurs de l'Etat ?

Les liens entre la recherche et l'administration paraissent plus faciles à mettre en place chez les opérateurs de l'Etat qui disposent dans leur organisation d'une direction liée à la recherche. Les travaux réalisés trouvent directement une application en lien avec les missions des opérateurs concernés et la question de la frugalité s'est posée sans qu'il y ait d'impulsion de l'Etat sur le sujet. Ainsi, comme présenté dans la partie II 1), l'IGN, sans avoir de politique spécifique liée à la frugalité, a lancé un challenge, FLAIR, qui intègre la notion de contrainte de sobriété sur les temps d'inférence et d'apprentissage, ainsi que le volume de calcul. A Météo-France, le Lab IA a confirmé se questionner sur les questions de frugalité en raison de contraintes liées à la puissance des machines.

Toutefois, Météo-France s'appuie beaucoup sur des briques de codes en open source. Une coopération importante entre les différents laboratoires de recherche paraît donc essentielle pour avancer concrètement sur le sujet.

Stefan Duffner, enseignant chercheur au laboratoire LIRIS, indique dans une interview: « *Aujourd'hui, beaucoup de modèles sont surdimensionnés et consomment beaucoup plus d'énergie que le besoin ne requiert* » (Duffner, Stéphane, 2023) et identifie comme un frein au développement de l'IA frugale le fait que cela implique des compétences dans plusieurs domaines, notamment pour trouver le bon compromis entre sécurité, robustesse et réponse aux besoins. Le lien entre la recherche et les utilisateurs, définissant le besoin, paraît donc essentiel pour avancer sur la frugalité dans l'IA.

2. Des voies à poursuivre ou à explorer

Les axes à développer entre la recherche et l'administration concernant l'IA frugale peuvent être résumés avec trois verbes: **"Définir, diffuser et former**".

i. Partager une définition de la frugalité entre la recherche et l'administration

La partie I de ce rapport a mis en évidence la diversité des définitions de l'IA frugale. Définir ce terme paraît être une base pour ensuite échanger sur le sujet. D'autant plus que c'est un sujet de discussion dans plusieurs pays. France Science, service pour la science et la technologie dépendant de l'ambassade de France aux Etats-Unis a d'ailleurs publié un

article en ligne le 19 décembre 2023 traitant spécifiquement du sujet de l'IA frugale: *« L'IA frugale bouleverse les codes technologiques : décryptage des solutions techniques innovantes depuis la Silicon Valley »* (France Science, 2023).

Définir l'IA frugale par le biais de l'impact environnemental est une piste qui se développe et qui présente l'avantage de faire abstraction de la raison de la frugalité. En effet, lors des entretiens, il est apparu que certaines utilisations de l'IA étaient frugales par construction car soit les données étaient peu nombreuses, soit les capacités de calcul étaient limitées, soit une combinaison des deux.

Diverses études sont parues ces derniers mois:

- Une étude américaine publiée en novembre 2023 sur l'empreinte carbone de l'IA (Luccioni et al., 2023). Ce type d'étude fournit des pistes pour le développement de l'IA frugale. A titre d'exemple, l'article montre que l'utilisation d'un modèle polyvalent pour des tâches spécifiques est plus énergivore que l'usage de modèles spécifiques pour ces mêmes tâches.
- Une étude française publiée en septembre 2023 sur l'impact environnemental de l'IA (Inria et al., 2023). Cette étude conclut notamment sur le fait qu'il ne faut pas se limiter au calcul de l'impact carbone lorsqu'on regarde l'impact environnemental de l'IA et sur la nécessaire vigilance des effets rebond des différentes applications.

Cet appel à la vigilance se retrouve d'ailleurs dans le livre blanc de la communauté des acteurs de l'IA en territoires: *« Pour aller jusqu'à une analyse en cycle de vie du projet IA, il est possible de s'intéresser à l'impact de la production des serveurs et de tous les équipements qui permettent à la solution de fonctionner. Dans un contexte d'évolution technologique, les porteurs de projet peuvent recenser en amont les effets rebond et d'obsolescence potentiels, afin de préparer des contre-mesures »* (Ecolab et al., 2023). On note toutefois l'absence d'obligations et les études complètes sont rarement réalisées notamment par manque d'outils. Si on se limite à la consommation énergétique et à l'empreinte carbone, ce même rapport cite les outils à disposition: *« Il est possible d'estimer la consommation énergétique et l'impact carbone de l'entraînement et des inférences à partir d'outils simples : Green Algorithms, ML COA Impact, Code Carbon et Cloud Carbon Footprint »*. Il serait intéressant de savoir si les développeurs s'attachent à utiliser ces outils simples lorsqu'ils mettent en place une nouvelle utilisation de l'IA. Concernant l'administration, dans sa feuille de route sur l'IA et la transition écologique, le ministère de

l'écologie en a fait une action: *Fiche action 12 : "utiliser les outils disponibles pour la mesure de l'impact environnemental des modèles d'IA et progresser sur les indicateurs"* (Ministère de la Transition Écologique et de la Cohésion du Territoire, 2023)

Partager une définition de l'IA frugale qui soit commune aux différents acteurs et mettre en place ou consolider des outils permettant d'estimer la frugalité paraissent être deux étapes importantes pour renforcer les liens entre la recherche et l'administration. Ce constat a d'ailleurs été fait par l'Ecolab qui lance le 15 janvier 2024 un groupe de travail sur l'impact environnemental de l'intelligence artificielle. L'objectif auquel ce groupe de travail se donne six mois pour répondre est notamment d'aboutir à un choix et à la définition d'indicateurs les plus pertinents et de méthodes de calcul associées pour évaluer l'impact environnemental de l'IA (AFNOR, 2023). Le nombre de personnes souhaitant participer à la réunion (plus de 200) et la diversité de leur origine (public et privé) montre l'intérêt suscité par ce type de questions.

ii. Un accompagnement nécessaire des administrations

Il paraît nécessaire d'accompagner les administrations dans leur utilisation de l'IA pour qu'elles puissent en exploiter le potentiel de façon utile. Par ailleurs, les besoins des territoires pouvant avoir des points communs, faire connaître et partager les différents projets mis en place devraient permettre d'éviter de mettre en œuvre différents systèmes pour le même type d'utilisations. L'appel à projet DIAT va clairement dans ce sens, ainsi que la feuille de route intelligence artificielle et transition écologique du pôle ministériel Ecologie, Energie Territoires dont la fiche action 3 consiste à capitaliser pour aider la généralisation de l'IA avec la *« mise en place d'une bibliothèque commune »* qui *« intégrerait des jeux de données et de modèles d'intelligence artificielle »* (Ministère de la Transition Écologique et de la Cohésion du Territoire, 2023). La fiche action 8 va également dans le même sens: il s'agit de "*monter en compétence à l'exploitation de l'IA: L'Ecolab agit comme acteur central de de réseau de partage connu sous le nom de « Club de l'IA »".* Toutefois sur cette action, il est dommage que les chercheurs ne soient pas cités comme partie intégrante du réseau.

Plusieurs initiatives de la part d'associations peuvent être mentionnées et seraient à poursuivre.

Ainsi, Céline Colucci, déléguée générale des Interconnectés, indique à propos de l'écosystème français autour de l'IA côté collectivités: *« Nous aidons cette communauté à se solidariser dans le cadre de notre Forum annuel national, les rencontres experts en région ou notre programme de formation Territoir'Prod »* (Ecolab et al., 2023).

Par ailleurs, l'association France HubIA, association ayant pour objectif l'accélération du développement et l'adoption d'une IA responsable, éthique et souveraine, accompagne l'Ecolab dans le DIAT et différents groupes de travail et participent activement à la réflexion sur l'impact de l'IA sur l'environnement.

Plus orientée vers les entreprises privées, l'association DataCraft indique avoir 700 chercheurs parmi ses membres et a organisé au moins un atelier sur l'IA frugale: ETAT DE L'ART - Embarquez dans l'IA frugale ! - Datacraft

Concernant les actions menées sur l'IA frugale au sein du PEPR IA, la question se pose du lien entre la recherche, les développements et les applications. Comment accompagner ce qui est financé dans le cadre du PEPR, au bénéfice des administrations et entreprises, c'est un enjeu qui est crucial et sur lequel, par exemple, Inria oeuvre beaucoup, avec une politique très axée sur les start-ups studios où les chercheurs sont sollicités pour monter des entreprises à partir d'un travail de recherche. Un acteur Inria nous indique que "*c'est une politique très assumée au niveau d'Inria par rapport à d'autres centres de recherche publique et qui peut être une clé de réponse dans le sens ou si on a cette intention-là, on va vraiment aller voir les travaux qui ont une finalité potentielle, une innovation pouvant se déployer dans le cadre d'une solution numérique IA à l'échelle par la création d'un business. On va chercher la personne sur le PEPR qui monte un travail qui est prometteur et on la contacte. Au sein d'Inria et en dehors, il y a des passerelles.*" Il s'agit donc de mettre en place une politique de transfert au sein du PEPR. Ce même acteur nous indique que cette question peut également créer des tensions si on considère que le travail des chercheurs est de mener une recherche fondamentale et non de créer un modèle économique associé à ce travail (par exemple en créant une start-up).

Inclure la recherche dans ces différentes initiatives paraît essentiel pour faire un pont entre la recherche et les administrations afin d'orienter les différents projets vers une valeur ajoutée maximale de l'IA. L'idée de l'institut Paris Région paraît intéressante dans cet objectif. En effet, l'institut organise des petits déjeuners décideurs-chercheurs intitulés «Inventons nos futurs » pour lancer le dialogue entre les enseignants-chercheurs et les

décideurs (L'Institut Paris Région, 2023)

Enfin, la démarche mise en place par la Grande-Bretagne donne un bon exemple de rapprochement entre la recherche et l'administration : le pairing scheme (*Pairing scheme | Royal Society*, 2023) favorise l'acculturation en mettant en place des échanges entre les scientifiques et les décideurs publics lors desquels chacun découvre le travail de l'autre.

iii. L'apport de la formation initiale

Lors d'un entretien, un membre de France Hub IA soulevait qu'il était rare dans l'éducation et l'enseignement supérieur d'avoir une discipline qui évolue aussi rapidement que l'IA actuellement. Maintenir des programmes qui soient toujours d'actualités présente un véritable défi (Hub France IA, communication personnelle, 19 octobre 2023).

Apprendre à se préoccuper de l'impact environnemental et introduire le concept d'IA frugal dès la formation initiale permettrait de poser de bonnes bases pour le futur que ce soit au niveau des décideurs dans les administrations, des utilisateurs ou des chercheurs.

Sur Parcoursup (Parcoursup, 2023), cette année, 13 formations post-bac citent explicitement le terme Intelligence artificielle (licence, écoles d'ingénieurs, école de commerce, privé ou public). Les termes frugal ou impact environnemental n'apparaissent pas directement dans les descriptifs mais cela ne veut pas dire qu'ils ne sont pas traités. Par exemple, pour la formation au titre d'ingénieurs, la commission des titres CTI indique que « *En termes de responsabilité environnementale, l'école vise à maîtriser les impacts environnementaux de son activité* » et « *L'école permet à ses élèves d'acquérir les compétences nécessaires pour accompagner la transition écologique et énergétique en privilégiant une approche systémique* » (*Critères et procédures – Ingénieur – CTI – Commission des Titres d'Ingénieur*, 2023).

Les enseignants-chercheurs avec qui nous nous sommes entretenus sont conscients de la nécessité d'aborder la question de l'impact environnemental de l'IA dès la formation initiale. Ils y accordent une place de plus en plus importante même s'il n'y a pas de cours spécifiques identifiés sur le sujet.

Pour l'administration, attirer les jeunes talents issus de la formation initiale et ayant des connaissances en IA tout en ayant saisi l'importance de l'impact environnemental constitue un défi important à relever.

iv. La formation des administrations

Sur la plateforme de la formation en ligne des agents de la fonction publique de l'Etat, Mentor, existe une initiation à l'intelligence artificielle (Plateforme interministérielle Mentor de formation, 2023). L'impact environnemental de l'IA y est à peine cité. Compléter l'offre de formation en ligne pour sensibiliser les agents et introduire la notion de frugalité permettrait de diffuser la notion dans la fonction publique. Il serait intéressant d'agir relativement rapidement de façon à s'inscrire dans les programmes de formation déjà mis en place sur la transition écologique ou celles à venir sur la transformation numérique. En effet, dans une de ses interventions, Stanislas Guerini, alors ministre de la transformation et de la fonction publique, indiquait « *Aussi, dès cette année, l'ensemble des directeurs d'administration bénéficieront d'une* ***formation aux enjeux de la transformation numérique****.* » (Guerini,Stanislas, 2023)

Combiner une action de sensibilisation des agents publics avec des actions permettant de faire le lien entre les administrations et la recherche devrait permettre à l'IA frugale de prendre un essor nécessaire et utile à la transition écologique et énergétique.

Conclusion

Sur la base d'entretiens avec les acteurs du domaine de l'IA, cette étude a analysé divers aspects de l'IA frugale, sa définition et son interprétation, la manière dont l'administration s'approprie ce concept et enfin le rôle que la recherche publique peut tenir pour participer à l'intégration de l'IA frugale dans nos usages.

Nous proposons, pour conclure, une analyse stratégique ci-dessous sur l'IA frugale permettant de mettre de manière synthétique les points forts et les points faibles de l'IA frugale à ce jour, ainsi que les menaces et les opportunités associées à cette thématique.

STRENGTHS S	WEAKNESSES W	OPPORTUNITIES O	THREATS T
Points forts	**Points faibles**	**Opportunités**	**Menaces**
• Une approche vertueuse de la frugalité à la fois pour les administrations (exemplarité) et les entreprises (RSE) ; • Un réseau d'acteurs diversifié s'étant approprié la thématique de l'IA (PME, start-up, associations…) et un écosystème de recherche publique à la pointe (Inria, CNRS, CEA) ; • Des moyens financiers mis à disposition par l'Etat ; • Un réseau d'acteurs mobilisés au sein de l'administration publique (Ecolab, Etalab) ; • Des passerelles possibles pour diffuser l'IA au sein des administrations (par exemple les EIG: postes créés et déployés dans l'administration pour recruter des talents du numérique au sein des administrations) ; • Des supercalculateurs de GENCI dont Jean Zay ; • Un mix énergétique français favorable à l'entraînement des modèles et à la phase d'inférence	• Une participation modérée de la communauté scientifique à l'élaboration de la stratégie globale sur l'IA frugale ; • L'IA frugale, un terme "valise" non clairement défini, à multiple facette pouvant mener à différents niveaux d'interprétation suivant les acteurs (allant jusqu'à un sentiment de Green washing) ; • Peu de littérature scientifique sur ce sujet ; • Manque de référentiel normatif sur le cadre de l'IA frugale ; • En France, une maturité des utilisateurs potentiels encore relativement faible sur l'IA et a fortiori sur l'IA frugale - notamment peu d'agents formés au numérique dans les ministères ; • Essor de la frugalité à mettre en balance avec l'acceptabilité d'une baisse de la performance des modèles ; • Pas de lien à ce jour entre la recherche sur l'IA frugale, menée au sein du PEPR IA et le programme DIAT.	• Bénéficier d'un essor économique pouvant être apporté par l'IA frugale (avantage concurrentiel) ; • Pouvoir déployer des solutions d'IA sur du matériel existant plutôt que de devoir renouveler du matériel (économie de l'utilisation de terres rares). Démocratisation possible des techniques d'IA pour les entreprises, les associations, les citoyens ; • Souveraineté numérique ; • Attirer des talents dans le monde académique, qui vont pouvoir continuer à former les générations suivantes ; • Créer des passerelles pour de futurs liens (i) entre la recherche sur l'IA frugale menée au sein du PEPR IA et les applications à l'échelle des territoires dans l'appel à projet DIAT, ou (ii) entre les organismes de recherche impliqués dans le PEPR IA sur des projets liés à la frugalité et des opérateurs de l'Etat (Cerema, IGN, Météo France…) ; • Créer des passerelles entre la recherche et la formation - utiliser des solutions de type MOOC, ou des contenus courts équivalents de sensibilisation à la thématique d'IA frugale, pour un ensemble large d'étudiants ou d'élèves ; • Acculturer et sensibiliser les différents ministères à l'IA frugale ; • Faire connaître aux entreprises françaises fournisseuses en IA les différentes solutions d'infrastructures pouvant être utilisées pour les phases d'entraînement et d'inférence des modèles ; • Développer un cadre normatif pour l'IA frugale; y associer une métrique faisant consensus et un label à la fois qualitatif et quantitatifs	• Gestion de l'utilisation de données qui peuvent être critiques, confidentielles, personnelles, sensibles ; • Problèmes d'éthique ; • Cybersécurité ; • Augmentation de l'impact environnemental par effet rebond - Non maîtrise des coûts de capture de ces données, de transmission, de stockage d'analyse de ces données ; • Essoufflement de la dynamique sur l'IA frugale si l'utilité n'est pas reconnue in fine et non transposition des résultats de la recherche ; • Des financements temporaires ne permettant pas de façonner de manière durable et qualitative des générations d'ingénieurs et d'universitaires pour alimenter l'industrie de l'IA.

Recommandations

À l'issue de notre travail, il apparaît que deux familles de recommandations peuvent être formulées. Les premières, à destination des pouvoirs publics et de l'ensemble des acteurs concernés par l'intelligence artificielle, dont le but principal est de fournir un écosystème permettant de favoriser l'émergence de modèles d'intelligence artificielle plus frugaux et aux acteurs de pouvoir objectiver et communiquer sur l'impact de leurs modèles. Les secondes, à destination plus spécifiquement de la recherche, visent à préciser les premières pour ces acteurs et à répondre plus spécifiquement à la commande qui nous a été faite.

Recommandations générales :

La recherche de frugalité des modèles va globalement dans le sens de l'intérêt économique des acteurs et cela peut être stratégique pour la France de viser un leadership, éventuellement international, dans l'IA frugale. Elle peut permettre de développer des modèles plus rapidement, pour une mise sur le marché accélérée. Elle limite l'investissement dans le hardware, sachant que les cartes graphiques sont à ce jour difficiles à trouver et très chères, et le besoin de location de temps de serveur, etc. Il nous apparaît donc qu'il n'y a pas lieu de mettre en place de normes visant à contraindre l'ensemble des acteurs à tendre vers plus de frugalité. D'autant plus que concernant certaines solutions, il n'est pas envisageable de perdre même 1% d'efficacité du modèle.

Cependant, afin de faciliter cette recherche de frugalité, il nous apparaît qu'un environnement favorable est nécessaire pour embarquer l'ensemble des acteurs. Des acteurs pressentis sont identifiés à la suite de chaque recommandation pour initier l'action associée. Dans le cadre de ce travail, il n'a pas été possible de rencontrer le coordinateur national de la SNIA et de recueillir sa vision sur la gouvernance pressentie sur la thématique de l'IA frugale. Ces propositions ne présument donc pas des prérogatives futures de chacun des acteurs dans un contexte évoluant très vite.

<u>Recommandation 1 :</u> favoriser l'émergence d'un consensus sur la définition et le champ de l'IA frugale.

Veiller à faire aboutir la dynamique initiée mi-janvier 2024 par l'Ecolab et l'AFNOR sur l'aspect environnemental de l'IA et à obtenir l'adhésion de l'ensemble des acteurs, recherche comprise.

Acteur pressenti : Ecolab

<u>Recommandation 2 :</u> favoriser les échanges de pratiques et de codes entre les acteurs.

Que ce soit à l'intérieur d'une structure ou entre acteurs de différentes structures, les échanges de pratiques, les retours d'expérience peuvent permettre d'embarquer des acteurs favorables à plus de frugalité et d'accélérer la mise en œuvre de ces modèles dans leurs structures.

Acteurs pressentis : acteurs de l'IA au sein des différentes structures sous l'impulsion de l'Idris (opérateur du calculateur Jean Zay)

<u>Recommandation 3 :</u> généraliser la connaissance et les outils de mesure d'impact des modèles

Connaître, favoriser les outils de mesure d'impact des modèles et pouvoir comparer des modèles mis en œuvre par différents acteurs devrait permettre d'objectiver les résultats obtenus par chacun et, à l'intérieur de l'environnement, de permettre la diffusion des solutions qui ont fait leurs preuves. Intégrer les impacts environnementaux dans les formations portant sur l'IA. Cette recommandation vise avant tout à créer de l'émulation et une attention sera portée à ce que cela ne tende pas vers du "name and shame".

Acteurs pressentis : DINUM (Etalab) et commission d'attribution des titres d'ingénieurs

<u>Recommandation 4 :</u> tendre vers la mise en place d'un label

Dans un premier temps, valoriser les acteurs qui participent de cette dynamique d'échange de bonnes pratiques, puis dans un second temps, si cela est possible (de l'objectivation sera nécessaire pour comparer les modèles), valoriser via un label l'utilisation de modèles frugaux, sur le modèle de "greentech innovation" de l'Ecolab ou du réseau "datalliance" de l'IGN.

Concernant la frugalité qui est permise par des modèles d'IA permettant de diminuer la consommation de ressources et l'impact sur l'environnement, le cadre actuel des appels à projets dans les territoires visant à aider les projets et diffuser les résultats obtenus nous apparaît comme une solution intéressante. Cependant, une évaluation de la capacité de ces projets à diffuser sera à mettre en œuvre afin de déterminer si cette diffusion est réelle et quels seraient les leviers d'accélération de celle-ci.

Acteurs pressentis : Coordinateur national de la SNIA en lien avec l'Ecolab et l'AFNOR

Recommandation 5 : faciliter la diffusion des projets dans l'ensemble de l'administration

Une fois un retour d'expérience des projets faits, déterminer la meilleure solution pour que ces projets diffusent également à l'ensemble de l'administration : par exemple en favorisant la diffusion de ces projets ou en créant des appels à projets spécifiques pour des besoins particuliers des administrations qui n'existent pas dans les territoires.

Acteurs pressentis : l'Ecolab et l'Etalab

Recommandation 6 : évaluer la capacité de diffusion des projets dans les territoires de l'appel à projets DIAT

Si l'appel à projet à bien permis de favoriser la mise en œuvre de solutions locales, il est important de déterminer l'impact sur d'autres territoires et la capacité de diffusion de celles-ci.

Acteur pressenti : l'Ecolab

Recommandation 7 : nouer des partenariats

Créer des passerelles pour de futurs liens (i) entre la recherche sur l'IA frugale menée au sein du PEPR IA et les applications à l'échelle des territoires dans l'appel à projet DIAT, ou (ii) entre les organismes de recherche impliqués dans le PEPR IA sur des projets liés à la frugalité et des opérateurs de l'Etat (Cerema, IGN, Météo France…).

Acteurs pressentis : responsables du PEPR IA

Recommandations à destination de la recherche :

La recherche est un maillon essentiel dans la production et l'évaluation de modèles d'intelligence artificielle. Il est nécessaire que celle-ci soit partie prenante des groupes d'échanges, pour faciliter la diffusion de leurs travaux, avoir des retours des utilisateurs et être à l'interface pour connaître les besoins spécifiques des acteurs concernant des formations.

Recommandation 8 : favoriser la présence et la visibilité des chercheurs dans l'écosystème de l'IA frugale et dans les groupes d'échanges.

Favoriser et valoriser la présence de chercheurs à la fois dans les cercles du conseil aux décideurs et dans la mise en œuvre des projets d'IA, faire des liens entre l'administration et la recherche à ce sujet. Quand des initiatives existent, il serait intéressant de les favoriser et de les diffuser.

Acteurs pressentis : responsables du PEPR IA

Recommandation 9 : favoriser l'open science et les échanges d'algorithmes

La mise en œuvre pratique des modèles d'IA passe souvent par l'utilisation de parties de code éprouvées diffusées par d'autres acteurs, ce qui permet également de proposer des corrections sur ces codes. L'optimisation du travail des administrations et de l'ensemble des acteurs passe par cette fluidité des échanges, qui peut également favoriser le rapprochement d'acteurs qui n'avaient pas de lien auparavant.

Acteurs pressentis : responsables du PEPR IA et les porteurs de projets

Recommandation 10 : favoriser la vulgarisation des travaux de la recherche

Un engagement de la recherche pour vulgariser au plus grand nombre, pas seulement via la formation ou par plus de communication avec le public ou le privé, mais vraiment une démarche d'ouverture de la connaissance à tout un chacun qui passerait par un engagement des chercheurs.

Acteurs pressentis : responsables du PEPR IA

L’ensemble des recommandations proposées converge dans une visée générale de créer, ou favoriser quand il existe, un écosystème efficace de partage autour de l’IA, et en particulier l’IA frugale. Cependant, cet écosystème n’est pas seulement nécessaire en ce qui concerne la frugalité mais permettra également une meilleure connaissance et diffusion des solutions existantes, favorisant par là la mise en œuvre de projets innovants et de confiance.

Liste des entretiens

Entretiens avec des chercheurs:

(Réunion de lancement du PEPR SHARP IA, communication personnelle, 16 octobre 2023)

(Inria, communication personnelle, 16 octobre 2023)

(CEA, communication personnelle, 16 octobre 2023)

(Inria & University College London, communication personnelle, 20 décembre 2023)

(Inria, communication personnelle, 4 janvier 2024)

Entretiens avec des agents d'administration:

(Ecolab, Ministère de la Transition Écologique et al., communication personnelle, 3 octobre 2023)

(Etalab, Ministère de l'Economie et des Finances et al., communication personnelle, 11 octobre 2023)

(Coordination de la SNIA, Ministère de l'Economie et des Finances, communication personnelle, 14 décembre 2023)

(Direction générale du Trésor, Ministère de l'Economie et des Finances, communication personnelle, 31 octobre 2023)

(Agence de Service et de Paiement, Ministère de l'Agriculture, communication personnelle, 14 novembre 2023)

(Mairie de Noisy-le-Grand, communication personnelle, 13 novembre 2023)

(Ecolab, Ministère de la Transition Écologique, communication personnelle, 10 janvier 2024)

Entretiens avec des opérateurs de l'Etat:

(IGN, communication personnelle, 3 octobre 2023)

(L. Météo France, communication personnelle, 24 octobre 2023)

Entretiens avec des associations

(Ekitia, communication personnelle, 6 décembre 2023)

(Hub France IA, communication personnelle, 19 octobre 2023)

Interviews en ligne

- La grande interview de Guillaume Avrin - SMART TECH (B SMART, 2023)
- Webinaire du 13/06/2023 de l’appel à projets Démonstrateurs d’IA frugale dans les territoires (Greentech Innovation, 2023)

Bibliographie

Articles de presse

- L'Agefi, & Cousin, C. (2023, octobre 13). Comment l'intelligence artificielle irrigue secteurs et entreprises. *L'Agefi*. https://www.agefi.fr/news/entreprises/comment-lintelligence-artificielle-irrigue-secteurs-et-entreprises
- Le Monde, Geiger, P., Pénicaud, S., Romain, M., & Sénécat, A. (2023, décembre 4). *Enquête sur les dérives de l'algorithme des caisses d'allocations familiales*. https://www.lemonde.fr/les-decodeurs/article/2023/12/04/profilage-et-discriminations-enquete-sur-les-derives-de-l-algorithme-des-caisses-d-allocations-familiales_6203796_4355770.html
- Le Point, & Grallet, G. (2023, novembre 21). *Intelligence artificielle : Les dessous de l'incroyable feuilleton Sam Altman*. https://www.lepoint.fr/economie/les-vrais-enjeux-du-limogeage-de-sam-altman-21-11-2023-2544044_28.php

Articles de recherche

- Castaño, J., Martínez-Fernández, S., Franch, X., & Bogner, J. (2023). Exploring the Carbon Footprint of Hugging Face's ML Models : A Repository Mining Study. *2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 1‑12. https://doi.org/10.1109/ESEM56168.2023.10304801
- CNRS. (2023). *PEPR d'accélération Intelligence artificielle*. https://www.cnrs.fr/fr/pepr/pepr-dacceleration-intelligence-artificielle
- Govindan, K. (2022). How Artificial Intelligence Drives Sustainable Frugal Innovation : A Multitheoretical Perspective. *IEEE Transactions on Engineering Management*, *PP*, 1‑18. https://doi.org/10.1109/TEM.2021.3116187

- Gribonval, R., Chatalic, A., Keriven, N., Vincent Schellekens, Jacques, L., & Schniter, P. (2021). *Sketching Data Sets for Large-Scale Learning : Keeping only what you need*. https://ieeexplore.ieee.org/document/9524547
- Inria, Delort, E., Riou, L., & Srivastava, A. (2023). *Environmental Impact of Artificial Intelligence* (p. 1) [Report, INRIA ; CEA Leti]. https://inria.hal.science/hal-04283245
- *Intelligence artificielle : Quels impacts sur le monde du travail ? | Inria*. (2022, septembre 27). https://www.inria.fr/fr/intelligence-artificielle-quels-impacts-sur-le-monde-du-travail
- International Centre for Frugal Innovation$. (2023). *Frugal Innovation for Science and Society*. https://www.icfi.nl/home
- Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). *Making AI Less « Thirsty » : Uncovering and Addressing the Secret Water Footprint of AI Models* (arXiv:2304.03271). arXiv. http://arxiv.org/abs/2304.03271
- Luccioni, A. S., Jernite, Y., & Strubell, E. (2023). *Power Hungry Processing : Watts Driving the Cost of AI Deployment?* (arXiv:2311.16863). arXiv. http://arxiv.org/abs/2311.16863
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54-63. https://doi.org/10.1145/3381831
- Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP* (arXiv:1906.02243). arXiv. http://arxiv.org/abs/1906.02243

Livres

- CNIL, Pailhès, B., & et Al. (2023). *Données, empreinte et libertés*.
- Data Publica, & Banuls, J. (2023). *La Transparence des Algorithmes Publics*.
- Dunod, Pachot, A., & Patissier, C. (2022). *Intelligence artificielle et environnement : Alliance ou nuisance ? : L'IA face aux défis écologiques d'aujourd'hui et de*

demain—ScholarVox Université. https://univ-scholarvox-com.extranet.enpc.fr/reader/docid/88937709/page/1?searchterm=Intelligence%20artificielle

- Ecolab, Les Interconnectés, & Hub France IA. (2023). *Livret Blanc de la Communauté des Acteurs de l'IA en Territoires*.
- Larousse, É. (2023). *Définitions : Frugal, Pays frugaux, (groupe des) frugaux*. https://www.larousse.fr/dictionnaires/francais/frugal/35451

Rapports

- Conseil d'Etat. (2022, août 31). *Intelligence artificielle et action publique : Construire la confiance, servir la performance*. Conseil d'État. https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance
- Cour des Comptes. (2023, avril 3). *La stratégie nationale de recherche en intelligence artificielle*. https://www.ccomptes.fr/fr/publications/la-strategie-nationale-de-recherche-en-intelligence-artificielle
- Data For Good, Lespagnol, R., & et Al. (2023, juillet 20). *Les grands défis de l'IA générative*. https://issuu.com/dataforgood/docs/dataforgood_livreblanc_iagenerative_v1.0
- Direction générale des réseaux de communication, du contenu et des technologies (Commission européenne). (2019). *Lignes directrices en matière d'éthique pour une IA digne de confiance*. Office des publications de l'Union européenne. https://data.europa.eu/doi/10.2759/74304
- European Commission. Directorate General for Research and Innovation., Fraunhofer ISI., & Nesta. (2017). *Study on frugal innovation and reengineering of*

traditional techniques. Publications Office. https://data.europa.eu/doi/10.2777/94587

- LEES PERASSO, E., VATEAU, C., DOMON, F., ADEME, ARCEP, & Bureau Veritas. (2022, janvier). *Evaluation de l'impact environnemental du numérique en France et analyse prospective*. La librairie ADEME. https://librairie.ademe.fr/consommer-autrement/5226-evaluation-de-l-impact-environnemental-du-numerique-en-france-et-analyse-prospective.html
- *Proposition de cadre réglementaire sur l'intelligence artificielle | Bâtir l'avenir numérique de l'Europe*. (2023, novembre 16). https://digital-strategy.ec.europa.eu/fr/policies/regulatory-framework-ai
- Villani, C. (2018, mars 28). *Donner un sens a l'intelligence artificielle : Pour une strategie nation | vie-publique.fr*. http://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation

Sites Internet

- Ademe. (2023). *Bilans GES*. https://bilans-ges.ademe.fr/
- AFNOR. (2023, décembre 15). Intelligence artificielle frugale. *AFNOR Normalisation*. https://normalisation.afnor.org/nos-solutions/afnor-spec/intelligence-artificielle-frugale
- Axionable. (2023). *1er Spécialiste du conseil en IA durable*. https://www.axionable.com/
- B SMART (Réalisateur). (2023, avril 12). *SMART TECH - La grande interview de Guillaume Avrin | Coordonnateur national pour l'IA*. https://www.youtube.com/watch?v=7fRMNOK6lWA
- CNIL. (2023). *Glossaire de l'intelligence artificielle (IA)*. https://www.cnil.fr/fr/intelligence-artificielle/glossaire-ia
- Conseil de l'UE. (2023, décembre 9). *Artificial intelligence act : Council and*

Parliament strike a deal on the first rules for AI in the world. https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

- *Critères et procédures – Ingénieur – CTI – Commission des Titres d'Ingénieur*. (2023, décembre 1). https://www.cti-commission.fr/fonds-documentaire
- Data for Good. (2023). *Le numérique pour l'intérêt général*. https://dataforgood.fr/
- Datacraft. (2023). The Club for data scientists and their companies. *Datacraft*. https://datacraft.paris/
- Duffner, Stéphane. (2023, mars 13). *« Il est possible d'aller vers une IA plus frugale »*. INSA Lyon. https://www.insa-lyon.fr/fr/actualites/il-est-possible-d-aller-vers-ia-plus-frugale
- Ekimetrics. (2023). *Préparez votre business à l'IA de demain.* Ekimetrics. https://ekimetrics.com/fr/
- Ekitia. (2023). Espace de confiance et éthique de l'économie de la donnée. *Ekitia*. https://www.ekitia.fr/accueil/
- France Science. (2023, décembre 19). *L'IA Frugale bouleverse les codes technologiques*. https://france-science.com/lia-frugale-bouleverse-les-codes-technologiques-decryptage-des-solutions-techniques-innovantes-depuis-la-silicon-valley/
- GPAI. (2023). *A responsible AI strategy for the environment*. https://gpai.ai/projects/responsible-ai/environment/
- Guerini,Stanislas. (2023, octobre 9). *Expérimentation de l'intelligence artificielle au sein des services publics*. https://www.transformation.gouv.fr/ministre/actualite/stanislas-guerini-lance-une-experimentation-de-lintelligence-artificielle-au

- Hub France IA. (2023). *Groupes de travail*. Hub France IA. https://www.hub-franceia.fr/groupes-de-travail/
- L'Institut Paris Région. (2023, septembre 12). *Que faisons-nous ? - Institut Paris Région*. L'Institut Paris Region. https://www.institutparisregion.fr/institutparisregion.html
- Ministère de la Transition Écologique et de la Cohésion du Territoire. (2023, novembre 28). *Feuille de route intelligence artificielle et transition écologique*. Ministères Écologie Énergie Territoires. https://www.ecologie.gouv.fr/feuille-route-intelligence-artificielle-et-transition-ecologique
- Ministère de l'économie des finances. (2023, octobre 3). *La stratégie nationale pour l'intelligence artificielle*. https://www.economie.gouv.fr/strategie-nationale-intelligence-artificielle
- OCDE. (2023). *The OECD Artificial Intelligence Policy Observatory : Policies, data and analysis for trustworthy artificial intelligence*. https://oecd.ai/en/
- *Pairing scheme | Royal Society*. (2023, décembre 1). https://royalsociety.org/grants-schemes-awards/pairing-scheme/
- Parcoursup. (2023, novembre 1). *Carte Parcoursup—Parcoursup*. https://dossier.parcoursup.fr/Candidat/carte
- Plateforme interministérielle Mentor de formation. (2023). *Formation Objectif IA*. https://mentor.gouv.fr/badges/badge.php?hash=08e531fb6526a65d0fbaf2ede8c8814e4b5d4952&trk=public_profile_certification-title
- ShapingAI, & MediaLab. (2023). *Les formes de participation au développement de l'IA en France*. https://medialab.github.io/ShapingAI/

Table des figures

Table des tableaux

Annexes

Annexe 1 : trame d'entretien utilisée lors des entretiens hors recherche et experts

Présentation du projet :

Nous sommes une équipe de 4 étudiants travaillant sur un projet d'ingénieur commandité par le Laboratoire d'Informatique Gaspard Monge (groupe de recherche IMAGINE), de l'École des Ponts ParisTech.

Ce projet pose la question de l'usage de l'intelligence artificielle au sein des administrations (au sens large, administrations centrales, opérateurs de l'État, entreprises publiques) et des liens existants avec la recherche sur ce sujet.

L'État et l'IA :

Quels sont les usages actuels de l'IA par votre administration ?

Des projets en cours de développement de l'IA ?

Comment avez-vous assuré ce développement ? Compétences en interne, prestation

informatique, compétences inter-ministérielles / inter-opérateurs ? Si prestation informatique, compétences pour gérer l'appel d'offre et la sélection ?

Quels investissements avez-vous dû faire au-delà de la compétence en IA ? (ex: serveurs, données, infrastructure de partage de données …)

Quels sont les leviers et les limites actuelles de ce développement ?

Des besoins particuliers qui pourraient nécessiter son utilisation mais pour lesquels vous n'avez pas trouvé de solution ?

Est ce que vous/votre ministère est impliqué dans des appels à projet/ relation avec des acteurs de l'IA sur le territoire ? (ex: IA pour l'agriculture, appel à projet du MTE sur les villes avec IA frugale)

L'IA frugale :

Avez-vous entendu parler d'IA frugale ?

Comment définissez-vous l'IA frugale? Cette notion est-elle clairement définie selon vous ? Quels sont les facteurs qui la définissent ?

Est-ce que la distinction IA / IA frugale fait sens selon vous quand vous développez un projet d'IA pour répondre à un besoin ? Ou pour adapter un outil actuel pour le rendre plus frugal ?

Quelles sont / pourraient être vos motivations pour développer l'IA frugale : efficience ? réduction de la taille des données récoltées ? temps de traitement ? … Est-ce seulement un choix du prestataire ? Dans le cas d'un appel à prestataire, est-ce que la frugalité est attendue dans le cahier des charges ? Un facteur permettant d'arbitrer entre deux offres ? Ou la main est-elle laissée au prestataire pour développer l'outil qui lui semble le plus adapté ?

<u>Liens avec la recherche / formation :</u>

Est-ce que les avancées de la recherche sur l'IA, et en particulier l'IA frugale, diffusent dans votre administration ? Comment (agents qui se forment, recrutement de docteurs, partenariats, …) ? Est-ce via les développements opérés par les entreprises privées ?

Quels sont les leviers et les freins à cette diffusion de la recherche vers votre administration ?

Quels pourraient être les outils pour favoriser la diffusion des avancées de la recherche en IA frugale vers votre administration (formations) ?

Contacts ressource dans les administrations : mise en œuvre de l'IA voire IA frugale.

Annexe 2 : trame d'entretien utilisée lors des entretiens recherche et experts

Présentation du projet :

Nous sommes une équipe de 4 étudiants travaillant sur un projet d'ingénieur commandité par le Laboratoire d'Informatique Gaspard Monge (groupe de recherche IMAGINE), de l'École des Ponts ParisTech.

Ce projet pose la question de comment la recherche pourrait permettre de favoriser le développement de l'intelligence artificielle frugale au sein des administrations (au sens large, administrations centrales, opérateurs de l'État, entreprises publiques).

L'entretien proposé doit permettre d'explorer les thématiques suivantes :

- L'État et l'IA : quels sont les usages actuels par l'État de l'IA ? Les projets en cours ? Quelle maturité de l'État (appropriation, expertise, projets internalisés, externalisés) ?
- L'IA frugale : objectivation de la notion d'IA frugale, quels sont les projets d'IA frugales développés par et pour l'administration ? Quels sont les leviers et freins à ce développement ?
- Les liens de l'État et de la recherche/formation pour le développement de l'IA, en particulier l'IA frugale

L'État et l'IA :

Quels sont les usages actuels de l'IA par l'État ? Comment comprendre ce développement : par thématique (eau, transport, énergie…) par administration,... ? Des exemples ? Des contacts (services…) ?

Si besoin, si le champ "État" est trop vaste, restreindre avec l'accord de l'interviewé aux champs des Ministères de la Transition écologique et énergétique, voire les Ministères économiques et financiers.

Quels sont les leviers et les limites actuelles de ce développement ?

Comment l'État s'approprie et développe l'IA ? Compétences en propre, externalisation ? Interne à chaque administration / opérateur ou équipes interministérielles ?

Est-il pertinent de considérer le développement de l'IA dans les entreprises publiques /

dans le champ de l'État ou celui-ci s'assimile-t-il plus à celui des entreprises privées / concurrentielles ? Êtes vous en lien avec l'Etat sur vos travaux d'IA ? (formation, conseil, financements)

L'IA frugale :

Comment définissez-vous l'IA frugale? Est-ce que cette notion vous semble clairement définie ? Quels sont les facteurs qui la définissent (plutôt sur les données ou sur les data centers, soft vs hard) ?

Est-ce que la distinction IA / IA frugale fait sens selon vous quand un opérateur développe un projet d'IA pour répondre à un besoin ? Ou quand un opérateur adapte un outil actuel pour le rendre plus frugal ? Est-ce que le concept d'IA frugale fait sens pour les porteurs de projets ?

Quelles sont / pourraient être les motivations d'un porteur de projet pour développer l'IA frugale : efficience ? réduction de la taille des données récoltées ? temps de traitement ? … Est-ce seulement un choix du prestataire ? Dans le cas d'un appel à prestataire, est-ce que la frugalité est attendue dans le cahier des charges ? Un facteur permettant d'arbitrer entre deux offres ? Ou la main est-elle laissée au prestataire pour développer l'outil qui lui semble le plus adapté ?

Connaissez-vous des développements d'IA frugale dans le champ de l'État ? Qui, comment, en propre ou via un prestataire ?

Liens avec la recherche /formation :

Est-ce que les avancées de la recherche sur l'IA, et en particulier l'IA frugale, diffusent dans l'administration ? Comment (agents qui se forment, recrutement de docteurs, partenariats, …) ? Est-ce via les développements opérés par les entreprises privées ?

Quels sont les leviers et les freins à cette diffusion recherche - État ? Pistes avec la formation ?

Quels pourraient être les outils pour favoriser la diffusion des avancées de la recherche en IA frugale vers les administrations ? (ex: données de l'Etat à transmettre aux centres de recherche, ou inversement algorithmes frugaux développés par les chercheurs en IA à donner aux services de l'Etat "prêt à l'emploi")

Contacts ressource dans les administrations : mise en œuvre de l'IA voire IA frugale.