

OBSERVATIONS DÉFINITIVES

**L'INTELLIGENCE ARTIFICIELLE
DANS LES POLITIQUES
PUBLIQUES : L'EXEMPLE DU
MINISTRE DE L'ECONOMIE ET
DES FINANCES**

Exercices 2018 – 2023

Le présent document, qui a fait l'objet d'une contradiction avec les destinataires concernés,
a été délibéré par la Cour des comptes, le 18 juillet 2024.

TABLE DES MATIÈRES

TABLE DES MATIÈRES	2
SYNTHÈSE	5
RECOMMANDATIONS	10
INTRODUCTION	11
1 UN RECOURS A L'INTELLIGENCE ARTIFICIELLE INÉGAL ET ENCORE INSUFFISAMMENT PILOTE	13
1.1 Un recours à l'IA diversifié mais dominé par la gestion fiscale et la détection des fraudes	13
1.1.1 Des projets qui mobilisent l'ensemble des technologies associées aux dernières générations d'IA	13
1.1.2 Des missions variées mais dominées par la gestion fiscale et la lutte contre les fraudes.....	16
1.1.3 Des directions d'état-major qui recourent encore insuffisamment à l'IA	19
1.2 Une capacité à expérimenter l'IA hors de portée pour certains services	20
1.2.1 Des coûts de développement principalement portés par les projets de la DGFIP	20
1.2.2 Des compétences internes sur le champ de l'IA inégalement réparties	23
1.2.3 Un accès différencié aux données de masse et à la puissance de calcul	23
1.3 Un pilotage qui doit se structurer pour garantir un développement cohérent du recours au <i>data-mining</i> à l'IA.....	25
1.3.1 Une répartition des ressources qui n'est pas corrigée par les dispositifs ministériels et interministériels	25
1.3.2 Un pilotage ministériel à établir pour garantir la cohérence du recours à l'IA au sein du MEFSIN.....	27
1.3.3 Une offre de services à définir en faveur des directions métier en mobilisant la DINUM et les instruments financés par la stratégie nationale d'IA.....	29
2 IA ET GAINS DE PRODUCTIVITÉ : UN ENJEU D'EFFICACITÉ, DES EFFETS AMBIVALENTS	33
2.1 Les systèmes d'IA améliorent la productivité mais ne permettent pas nécessairement de libérer des emplois	34
2.1.1 Un nombre limité de systèmes d'IA réduisent le coût d'une action à périmètre constant	34

2.1.2	Une majorité de systèmes d'IA stabilisent le coût d'une action tout en augmentant ses résultats	35
2.1.3	Certains systèmes d'IA peuvent prendre en charge des missions nouvelles.....	36
2.1.4	Le recours à l'IA peut entraîner un effet de rebond sur l'activité des services	36
2.2	Des économies avérées, mais moins importantes qu'estimées initialement.....	37
2.2.1	Une cible totale de 61 M€ d'économies annuelles pour cinq systèmes d'IA	37
2.2.2	Des économies constatées pour un montant de 20 M€ en 2022.....	38
2.2.3	Un écart qui témoigne d'un suivi incomplet et d'hypothèses budgétaires parfois optimistes.....	39
2.2.4	Le FTAP, un dispositif structurant dont les indicateurs peuvent être enrichis	40
2.3	Une affectation des gains de productivité qui doit gagner en visibilité	41
2.3.1	L'affectation des économies réalisées grâce aux IA soulève des enjeux particuliers pour l'administration	41
2.3.2	Les conséquences des IA en termes de ressources humaines ne sont pas suffisamment anticipées.....	42
2.3.3	Une prospective à construire à l'échelle des projets et à l'échelle ministérielle.....	43
3	UN DEVELOPPEMENT DES SYSTEMES D'IA AU MEFSIN QUI NE TRAITE QU'IMPARFAITEMENT LES ENJEUX DE CONFIANCE ET LES DEFIS ENVIRONNEMENTAUX	46
3.1	Une prise en compte peu structurée des considérations de confiance.....	47
3.1.1	Des principes partagés mais non règlementés, un besoin d'outil évaluatif.....	47
3.1.1.1	Un appel partagé pour lutter contre les craintes vis-à-vis de l'IA, la non-explicabilité et le biais d'automatisation.....	47
3.1.1.2	Maturité des systèmes d'IA en termes de confiance au moyen d'une grille d'analyse élaborée par la Cour	50
3.1.2	Un niveau de maturité non standardisé, non piloté, non optimisé	55
3.1.2.1	Un niveau de maturité ministériel estimé à 2 sur 5, soutenu par des pratiques fortes en matière de gouvernance des données.....	55
3.1.2.2	La maturité des systèmes d'IA n'est pas corrélée au stade de développement mais semble plus liée à l'effet « organisation ».....	57
3.1.2.3	« Signaux Faibles », un système d'IA au niveau de confiance 4 sur 5.....	58
3.1.3	Une stratégie d'IA de confiance à consolider pour assurer la sécurité juridique des systèmes d'IA.....	59
3.2	Une intégration affichée des enjeux environnementaux mais une démarche non coordonnée.....	60
3.2.1	L'impact environnemental croissant de l'IA.....	60
3.2.2	Une prise en compte affichée de l'impact environnemental des systèmes d'IA du MEFSIN	61

3.2.3 Une possibilité d'intervention sur les données, les modèles d'IA, les infrastructures et l'énergie utilisée tout au long du cycle de vie d'un SIA	64
3.2.3.1 Agir sur les données	64
3.2.3.2 Agir sur les modèles	64
3.2.3.3 Agir sur les infrastructures	65
3.2.3.4 Agir sur l'énergie utilisée	65
3.2.4 Un risque de dilution des responsabilités	65
3.3 L'IA frugale, le point de convergence entre l'efficacité, la confiance et les intérêts environnementaux et de souveraineté	66
3.3.1 L'IA frugale, une efficacité de la dépense	67
3.3.2 La frugalité peut contribuer à l'IA de confiance	68
3.3.3 Classifier les IA selon leur frugalité.....	68
ANNEXES.....	72
Annexe n° 1. Glossaire	73
Annexe n° 2. Description des systèmes d'IA mis en œuvre ou projetés par le MEFSIN entre 2018 et 2023	77
Annexe n° 3. Présentation de la méthode d'évaluation de la maturité des systèmes d'IA en termes de confiance.....	82
Annexe n° 4. Sept principes d'IA de confiance	87

SYNTHÈSE

Nées dans les années 1950, les technologies d'intelligence artificielle (IA) ont connu un essor particulier depuis les années 2010 grâce au développement de la puissance de calcul des systèmes d'information et des nouvelles techniques d'apprentissage par la machine (*machine learning*). Ces techniques permettent la fabrication de systèmes d'intelligence artificielle (SIA), qui peuvent automatiser certaines tâches, générer du texte et des images et analyser un grand volume de données à des fins de prévision ou d'aide à la décision. Le recours à l'intelligence artificielle conjugué au facteur humain permet ainsi de produire des prévisions souvent plus fiables tout en diminuant très fortement leur coût. Les systèmes d'IA présentent à ce titre un potentiel économique qui a motivé l'adoption en France d'une stratégie nationale pour l'intelligence artificielle (SNIA) définie en mars 2018. Dans les administrations publiques, le recours à l'IA génère des gains de productivité importants qui peuvent être redéployés pour augmenter le volume et la qualité de leurs travaux, ou restitués pour contribuer à la maîtrise des dépenses publiques.

L'intelligence artificielle, une technologie expérimentée depuis 2015 par le ministère de l'économie, des finances, de la souveraineté industrielle et numérique

Depuis 2015, le développement de l'IA permet au ministère de l'économie, des finances et de la souveraineté industrielle et numérique (MEFSIN) de fonder certaines de ses interventions sur l'analyse de données plus complètes, en mobilisant par exemple le programme *Signaux Faibles*, qui améliore la détection des difficultés des entreprises dès l'apparition des premiers signes. Mobilisant la plupart des technologies d'IA les plus récentes, d'autres applications permettent au MEFSIN de mieux détecter les situations frauduleuses, tels que le programme *Foncier innovant* qui aide à l'identification de certains biens non déclarés à l'administration fiscale, ou de proposer aux usagers des robots conversationnels capables de répondre à certaines de leurs demandes sans interruption dans le temps.

En outre, la mobilisation de ces outils à des fins de diminution des dépenses, qui a concerné cinq projets sur les trente-cinq systèmes d'IA développés au sein du MEFSIN, a permis de réaliser plus de 20 M€ d'économies annuelles à compter de 2022, contre 46,6 M€ attendus, la trajectoire d'économies devant atteindre 60 M€ en 2024 sur ce périmètre. Le recours à la *data-science* pour contribuer à la programmation du contrôle fiscal a par exemple permis de réduire le nombre d'emplois mobilisés pour cette tâche et de vérifier le potentiel de ces outils en termes de contribution à la maîtrise des finances publiques.

Une expérimentation qui laisse certaines directions à la marge et mobilise trop peu les ressources interministérielles

Huit directions du MEFSIN recourent aux systèmes d'IA pour un coût total évalué à 66 M€ depuis 2015. La DGFIP et l'agence pour l'informatique financière de l'État (AIFE) regroupent la moitié de ces systèmes et 95 % des moyens mis en œuvre pour leur développement. De fait, les directions disposant historiquement d'un parc informatique

important, de services informatiques étoffés et d'un grand nombre de données de masse sont plus à même d'identifier des cas d'usages prometteurs pour l'emploi de l'IA et de solliciter les dispositifs d'appui et les financements interministériels existants.

Certaines directions, telles que la direction générale du Trésor et la direction du budget ne portent actuellement aucun projet d'utilisation de l'IA, en dépit du potentiel de ces technologies pour améliorer les fonctions de modélisation, de prévision ou d'organisation des données dans les domaines économiques et financiers.

Si les ressources proposées par les directions interministérielles du numérique (Dinum) et de la transformation publique (DITP), ou financées par la stratégie nationale d'IA (SNIA), pourraient compléter les moyens de ces services, elles restent peu connues et peu mobilisées. Le service du numérique (Snum) du MEFSIN, instance transversale susceptible d'accompagner l'ensemble des directions en interne, ne dispose pas des attributions qui seraient nécessaires pour accompagner les directions restées en marge du recours à l'IA et ne s'est pas doté d'une stratégie d'action concernant spécifiquement ce domaine. Dans ce contexte, le Snum n'est pas en mesure de remplir la fonction d'interface entre l'offre interministérielle et les directions métier du MEFSIN.

Une maîtrise inégale et encore insuffisante des risques spécifiques liés à l'IA

Certains risques liés au recours à l'IA semblent, à ce stade, maîtrisés par le MEFSIN.

Ainsi, des clauses d'internalisation des compétences sont prévues lorsque ses directions recourent à des prestataires extérieurs pour développer certains systèmes d'IA, ce qui leur permet de conserver la maîtrise technique des outils mis en production et de limiter leurs coûts d'entretien. De même, l'exposition des systèmes d'IA aux menaces cybernétiques paraît maîtrisée, dès lors que ces systèmes reposent sur des infrastructures généralement communes aux autres systèmes d'information sécurisés du MEFSIN.

Par ailleurs, les risques concernant la protection et le cadre juridique des données sont pris en compte de façon satisfaisante. Les systèmes d'IA du MEFSIN bénéficient de l'action des responsables de la sécurité des systèmes d'information et de la politique mise en place pour appliquer le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

En revanche, les enjeux de l'IA en termes d'éthique, de ressources humaines et de coût écologique n'apparaissent pas suffisamment maîtrisés.

Sur le plan éthique, le recours à l'IA suscite une appréhension de la part des usagers et des agents publics. Certains d'entre eux peuvent craindre que l'IA remplace progressivement le facteur humain au lieu de le compléter et limite la capacité de l'administration à rendre compte de ses décisions et à en identifier les responsables. Différents biais sont, par ailleurs, susceptibles affecter l'usage et les résultats de l'IA, tels que le biais d'automatisation, qui consiste à accorder un crédit excessif aux machines au détriment du jugement individuel des utilisateurs, ou les biais de sélection et la tendance à reproduire les déséquilibres historiquement présents dans les données d'entraînement. Il est ainsi nécessaire de mieux informer et de mieux accompagner les utilisateurs en cas de recours à l'IA, afin de garantir l'intégrité des résultats, la qualité des interactions avec le public et la prévention des risques contentieux.

Le recours à l'IA présente en outre des enjeux en termes de ressources humaines, en modifiant la nature et le volume de l'activité sur l'ensemble de la chaîne de travail. Ces nouveaux outils peuvent ainsi augmenter la charge de travail des services du fait de leurs performances, en détectant par exemple d'avantage d'irrégularités à traiter, ou du fait de leurs limites, en générant par exemple des réponses erronées qu'il revient aux humains de corriger. De plus, leur mise en œuvre modifie la répartition des métiers et des compétences nécessaires à la conduite des politiques publiques du MEFSIN. Ces enjeux ne sont pas suffisamment documentés dans les projets et ne sont quasiment pas traités par les instances de dialogue social du ministère. L'évaluation des conséquences de l'IA sur les métiers du MEFSIN a commencé en 2023, mais demeure très limitée.

Enfin, l'évaluation du coût environnemental des systèmes d'IA du MEFSIN est encore embryonnaire alors que l'État s'est engagé dans une démarche de réduction de l'impact environnemental du numérique. Il n'y a ni calcul systématique d'impact, ni harmonisation de la méthode calcul au sein du MEFSIN. Les actions entreprises pour réduire la taille des données, la longueur des algorithmes, la durée des entraînements, l'énergie utilisée, la provenance des matériaux ne sont ni documentées, ni coordonnées, et ne concernent, au mieux, qu'une partie du cycle de vie des systèmes d'IA utilisés au MEFSIN.

Un pilotage ministériel nécessaire pour maîtriser les risques éthiques et évoluer vers une IA de confiance

En l'absence d'un pilotage ministériel robuste et structuré du recours à l'IA, le MEFSIN n'est pas en mesure d'expérimenter ces technologies sur l'ensemble des processus pour lesquels elles présentent un intérêt potentiel. Les enjeux en termes d'éthique, de ressources humaines et de coût environnementaux demeurent perçus et traités inégalement au sein de chaque service, tandis que les outils disponibles au niveau interministériel restent peu connus et peu mobilisés.

Tout en préservant les logiques ascendantes, qui permettent aux directions de tester la validité du recours à l'IA au plus près de leurs besoins, une instance de pilotage ministériel est nécessaire pour garantir une réponse cohérente, efficiente et durable à l'ensemble de ces enjeux à l'échelle du MEFSIN.

Cette instance viserait d'une part à garantir que le potentiel de l'IA est expérimenté sur l'ensemble des fonctions dans lesquelles elle peut apporter des gains de productivité, de qualité et d'efficience. Ce pilotage devrait sécuriser dans la durée les prérequis à la mise en œuvre de l'IA que sont la disponibilité des données, des infrastructures et des compétences de *data science* au sein du MEFSIN.

Elle devrait d'autre part impulser la mise en œuvre d'une stratégie de « l'IA de confiance », qui a pour objet de sécuriser le recours à l'IA en répondant aux craintes inspirées par cette technologie. Cette instance de pilotage devrait ainsi définir des critères garantissant la transparence dans l'emploi de l'IA, son développement dans le respect de l'autonomie humaine, l'explicabilité de ses résultats, l'équité et la non-discrimination, le respect de la vie privée, ainsi que la sécurité et la robustesse des infrastructures et des modèles, tout au long du cycle de vie des solutions. L'établissement d'une charte ministérielle de l'IA de confiance constituerait un premier pas afin d'imposer des standards minimaux et des impératifs de primauté humaine et d'explicabilité des décisions ayant mobilisé l'IA au sein du MEFSIN.

Cette instance pourrait s'appuyer sur la méthodologie conçue par la Cour à l'occasion de ce contrôle pour examiner les systèmes d'IA à l'aune des critères de l'IA de confiance. Ce

référentiel permet d'évaluer les garanties apportées concernant l'explicabilité des résultats, le niveau de transparence envers les usagers, la prévention des discriminations, la robustesse technique, la sécurisation des données et la soutenabilité énergétique des systèmes d'IA.

Ce pilotage permettrait enfin de mobiliser plus systématiquement au profit des administrations du MEFSIN les outils interministériels proposés par la direction interministérielle du numérique, qui est responsable de la conception et de la mise en œuvre de la stratégie numérique de l'État des évolutions structurelles.

Pour garantir le développement d'une IA de confiance par un pilotage ministériel efficace, le MEFSIN peut s'inspirer des pratiques développées par France Travail et par la Banque de France, ou des réformes de la gouvernance du numérique récemment conduites par les ministères de l'Intérieur ou de l'Europe et des affaires étrangères¹. L'instance de pilotage ministériel devrait en effet disposer de compétences plus étendues que l'actuel service du numérique, en mobilisant les effectifs actuels du Snum et des agents issus des différentes directions métiers du MEFSIN, pour constituer une filière ministérielle du numérique et de la *data science* au service de l'ensemble des directions et services.

Des gains d'efficience à amplifier en anticipant l'évolution des métiers et en garantissant la frugalité dans le recours à l'IA

Si les systèmes d'IA tendent à générer des gains de productivité à l'échelle d'une étape de travail ou au sein d'un service déterminé, leur contribution durable à l'efficience globale du MEFSIN requiert une meilleure anticipation des évolutions en termes de ressources humaines et suppose une doctrine d'emploi attentive à leur frugalité.

D'une part, le MEFSIN doit anticiper avec plus de précision les conséquences du recours à l'IA en termes de ressources humaines, afin de maîtriser ses effets sur la nature des travaux confiés aux agents, de fidéliser les compétences nécessaires et de répartir de façon pertinente les gains de productivité permis par ces technologies.

À l'échelle de chaque projet, les implications en matière de productivité, d'emplois, de formation et de conditions de travail devraient être mieux documentées dès la phase d'études préalables, par une preuve de concept consacrée aux ressources humaines. Pour les plus grands projets, ces données permettraient d'enrichir les études MAREVA². À l'échelle ministérielle, un recensement des processus pour lesquels l'IA présente des gains potentiels significatifs pourrait être conduit en complément du rapport prospectif commandé par le MEFSIN à l'Institut national de recherche en sciences et technologies du numérique (INRIA), dont le champ est limité aux métiers du contrôle fiscal et douanier. Ainsi, le MEFSIN disposerait dès 2025 d'une vision globale des enjeux associés à l'IA en termes de métiers, de conditions de travail et de gains de productivité potentiels.

D'autre part, les services du MEFSIN doivent garantir la frugalité du recours à l'IA. Le concept d'IA frugale soutient, qu'au-delà des avantages écologiques, la sobriété en matière de

¹ Entre autres : IGF et conseil général de l'économie de l'industrie, de l'énergie et des technologies, *Les ressources humaines de l'État dans le numérique*, janvier 2023 ; Insee et DINUM, *Évaluation des besoins de l'État en compétences et expertises en matière de donnée*, septembre 2021.

² Les études dites « MAREVA » (Méthode d'Analyse et de Remontée de la Valeur) estiment la valeur prévisionnelle des projets de système d'information, en rapportant leur coût estimé et leurs bénéfices attendus, afin de produire un indice de retour sur investissement.

données, de modèles et d'infrastructures favorise l'efficacité de la dépense, la robustesse des systèmes d'IA et la capacité des organisations à en conserver la pleine maîtrise. Cette approche suppose que le choix de recourir à un système d'IA prenne en compte les coûts associés à l'ensemble du cycle de vie des systèmes. Elle implique également que le dimensionnement des données et des infrastructures soit rigoureusement adapté aux besoins et que les économies d'échelle possibles dans leur gestion soient identifiées et réalisées.

Cette démarche existe au sein de certaines équipes portant les systèmes d'IA au sein du MEFSIN, mais elle n'est pas animée à l'échelle ministérielle, au détriment du partage de bonnes pratiques et des gains d'efficacité. Or, la décision de recourir à l'IA doit être pilotée par l'efficacité de la dépense, à l'échelle des projets comme à l'échelle du MEFSIN. De même que la maîtrise des risques éthiques et de l'empreinte écologique associés à l'IA, la frugalité dans l'usage de l'IA requiert un renforcement du pilotage ministériel, afin de garantir dans la durée la cohérence et l'efficacité des initiatives développées au sein des différentes directions métier du MEFSIN tout en veillant à favoriser l'appropriation de l'IA par les administrations.

RECOMMANDATIONS

Recommandation n° 1. (SG des MEF) : Identifier l'instance ministérielle en charge du pilotage stratégique de l'intelligence artificielle et des données numériques.

Recommandation n° 2. (SG des MEF, Dinum) : Créer un incubateur d'IA au sein du MEFSIN avec l'aide de la Dinum.

Recommandation n° 3. (SG et directions du MEFSIN) Identifier en 2025 les missions et les processus pour lesquels l'IA est susceptible d'apporter des gains d'efficacité et de productivité significatifs.

Recommandation n° 4. (SG des MEF, Dinum) : Promouvoir auprès des directions des ministères économiques et financiers la mesure de l'impact environnemental des systèmes d'IA tout au long de leur cycle de vie en vue de le réduire.

Recommandation n° 5. (SG des MEF) : Promouvoir auprès des directions une démarche de frugalité ministérielle lors de la mise en œuvre de systèmes d'IA au MEFSIN.

INTRODUCTION

Depuis le premier réseau neuronal apprenant élaboré en 1957, les progrès accomplis par la recherche en intelligence artificielle ont permis d'établir dans les années 1980 des réseaux neuronaux multicouches capables de traiter des données complexes. À partir de années 2000, l'essor de l'IA a notamment reposé sur les techniques d'apprentissage profond (*deep learning*) par les machines. Ces techniques permettent ainsi la fabrication de systèmes d'intelligence artificielle (SIA), qui se définissent comme des « systèmes fondés sur une machine qui, pour répondre à des objectifs explicites ou implicites, déduit, à partir des données qu'ils reçoivent, comment générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer des environnements physiques ou virtuels. Les différents systèmes d'IA varient dans leurs niveaux d'autonomie et d'adaptabilité après le déploiement »³.

Le recours à l'intelligence artificielle conjugué au facteur humain permet de produire des prévisions souvent plus fiables tout en diminuant très fortement leur coût⁴. L'IA présente à ce titre un potentiel économique qui a motivé l'adoption en France d'une stratégie nationale pour l'intelligence artificielle (SNIA) définie en mars 2018 pour accélérer sa diffusion dans l'économie et son utilisation par les acteurs publics⁵.

En effet, lorsque sa mise en place est réussie dans le champ des administrations publiques, l'IA génère des gains de productivité qui peuvent être réutilisés pour améliorer l'efficacité et la qualité du service public et contribuer à la maîtrise de la dépense. La Cour a ainsi caractérisé à différentes reprises le potentiel des systèmes d'IA en termes de gains d'efficacité et d'amélioration du service rendu aux usagers⁶, tout en soulignant cependant les enjeux qui doivent être anticipés pour maîtriser les risques liés à l'IA.

Le recours à ces technologies comporte en effet des enjeux juridiques et éthiques, qui s'attachent notamment à la manipulation des données de masse et aux biais méthodologiques que les systèmes d'IA sont susceptibles de reproduire. Il pose également la question de la transparence de l'action publique, car les utilisateurs de systèmes d'IA ne sont pas toujours capables de rendre compte des réponses proposées par les machines et d'en expliquer la logique. Les systèmes d'IA ont par ailleurs une très forte empreinte environnementale et peuvent être vulnérables aux cybermenaces, présentant ainsi des risques pour la soutenabilité et la continuité de l'action publique.

De plus, si l'IA améliore généralement la productivité du travail, sa mise en œuvre peut déformer la répartition et la nature de l'activité en amont et en aval de la chaîne de production.

³ OCDE, définition révisée à l'automne 2023.

⁴ OCDE, « Paysage économique de l'IA », in *L'intelligence artificielle dans la société*, éditions de l'OCDE, Paris 2019.

⁵ Cour des comptes, *La stratégie nationale de recherche en intelligence artificielle*, avril 2023, page 30. Le volet recherche devait être financé à hauteur de 445 M€, le volet enseignement supérieur pour 128 M€, le volet économie pour 390 M€ et le volet défense et sécurité à hauteur de 410 M€.

⁶ Cf. Cour des comptes, « Le numérique au service de la transformation de l'action publique », in *Rapport public annuel 2020*, tome II, avril 2021, et *La stratégie nationale de recherche en intelligence artificielle : une stratégie à structurer et à pérenniser*, rapport public thématique, avril 2023.

Les systèmes d'IA présentent ainsi de forts enjeux pour l'organisation des équipes et l'évolution des métiers nécessaires à la conduite des politiques publiques.

Enfin, le développement des systèmes d'IA présente un potentiel de mutualisations aux échelles ministérielles et interministérielles, mais aussi un risque de dispersion. Si une gouvernance structurée peut permettre l'échange des communs numériques⁷, la capitalisation des expériences et le partage des infrastructures, une absence de pilotage peut à l'inverse disperser les moyens entre des projets non viables, ou conduire à retenir à une échelle non pertinente des choix technologiques dont la dimension structurante n'aurait pas été identifiée.

Afin de dresser un premier bilan du recours à l'IA pour un ensemble homogène de politiques publiques, la présente enquête analyse les modalités, les forces et les faiblesses du recours à l'IA au sein du ministère de l'économie, des finances et de la souveraineté industrielle et numérique (MEFSIN), et l'utilisation qui y est faite de ressources proposées au niveau interministériel, par les directions du ministère de la transformation et de la fonction publiques (MTFP), pour appuyer le recours à l'IA.

Elle recense ainsi l'ensemble des systèmes d'IA élaborés au sein du MEFSIN, présente les conditions de leur développement et analyse les données disponibles concernant les gains de productivités associés à ces outils.

Elle s'attache ensuite à évaluer la capacité du MEFSIN à piloter le recours à l'intelligence artificielle sur la base d'informations pertinentes et selon une gouvernance d'ensemble cohérente, qu'il s'agisse de la priorisation des projets, de la gestion prévisionnelle des compétences ou de la mutualisation des bases de données.

Elle présente également une méthodologie originale, conçue par la Cour pour cette enquête, qui permet d'examiner les systèmes d'IA de ces ministères à l'aune des critères de l'IA de confiance. Ce référentiel permet en effet d'évaluer les garanties apportées par les ministères économiques et financiers concernant l'explicabilité des résultats, le niveau de transparence envers les usagers, la prévention des discriminations, la robustesse technique, la sécurisation des données et la soutenabilité énergétique des systèmes d'IA qu'ils développent.

Elle examine enfin la contribution des ressources interministérielles au développement des systèmes d'IA au sein du MEFSIN.

⁷ Un commun désigne une ressource produite et entretenue collectivement par une communauté d'acteurs généralement hétérogènes, qui collaborent selon des règles assurant que la production finale soit partagée à un public large, de façon généralement gratuite. Un commun est numérique lorsqu'il s'agit d'une ressource dématérialisée : logiciel, base de données, lignes de code par exemple. Les communs numériques sont souvent produits de façon à être modifiables par les utilisateurs afin de créer des variantes adaptées à leurs besoins.

1 UN RECOURS A L'INTELLIGENCE ARTIFICIELLE INÉGAL ET ENCORE INSUFFISAMMENT PILOTE

Les systèmes d'intelligence artificielle recensés au sein du MEFSIN mobilisent la plupart des technologies d'IA les plus récentes. Ces outils sont utilisés dans des domaines variés pour des tâches concernant aussi bien les relations externes que les fonctions support de l'administration.

Malgré cette relative diversité, certaines directions restent cependant en marge des expérimentations en cours. Si la prépondérance des applications concernant la lutte contre les fraudes et l'exécution de la dépense publique reflète l'importance de ces missions au sein des différentes attributions du MEFSIN, elle révèle aussi la capacité très inégale des différents services à évaluer et à mobiliser le potentiel de l'IA pour leurs missions.

En l'absence d'une démarche pilotée au niveau ministériel, le développement de l'IA risque de se concentrer à terme sur la gestion financière et fiscale et la lutte contre les fraudes, sans que son potentiel puisse être testé dans d'autres fonctions, concernant notamment les missions dévolues aux « directions d'état-major » (DG Trésor, direction du budget, services d'inspection et de contrôle).

Tout en préservant les logiques ascendantes, qui permettent à chaque direction de détecter des cas d'usages et de tester la validité des IA au plus proche de ses métiers, le MEFSIN gagnerait à établir une démarche ministérielle concertée, garante d'un déploiement équilibré de l'IA entre ses différentes missions.

1.1 Un recours à l'IA diversifié mais dominé par la gestion fiscale et la détection des fraudes

Les directions et services du MEFSIN étudient, développent ou exploitent au total 35 systèmes d'intelligence artificielle à fin 2023. Treize systèmes d'IA sont déployés et exploités, huit sont en cours de développement et 14 sont envisagés ou font l'objet d'une étude de faisabilité.

Au total, huit directions et services recourent à l'IA. La plupart de ces directions présentent un panel d'outils variés reposant sur des technologies d'IA diversifiées.

1.1.1 Des projets qui mobilisent l'ensemble des technologies associées aux dernières générations d'IA

L'apprentissage par la machine (*machine learning*), qui permet à une IA de déterminer elle-même des corrélations et des règles après une phase d'entraînement sur un très grand nombre de données, est développé au sein du MEFSIN sous ses différentes formes.

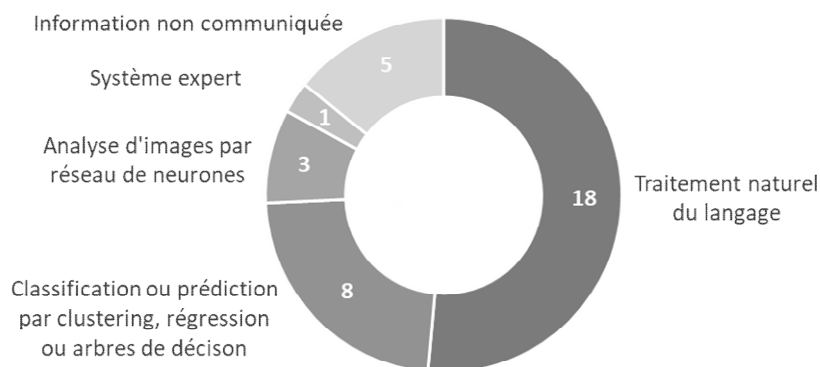
L'apprentissage supervisé ou semi-supervisé, qui consiste à entraîner une IA au moyen de données labellisées et qualifiées, concerne 26 programmes sur 35. Quatre systèmes reposent

cependant sur un apprentissage non supervisé, par lequel la machine apprend à structurer des données non labellisées en identifiant elle-même des différences, des récurrences ou des anomalies. Trois projets mobilisent par ailleurs les technologies d'apprentissage profond (*deep learning*), qui consiste à entraîner un réseau de plus de trois couches de neurones artificiels à réaliser une tâche, telle que la reconnaissance d'objets dans une image.

Les équipes du MEFSIN expérimentent ainsi la plupart des technologies d'IA les plus récentes, qui peuvent être regroupées au sein de trois principales catégories :

- Le traitement naturel du langage, qui permet la création d'IA génératives capables de créer et d'analyser des textes, des images ou des sons, afin par exemple d'orienter des demandes, d'identifier des risques ou de réorganiser des données, et la production d'IA capables de générer des contenus tels que des propositions de réponse aux demandes des usagers ou des agents sous la forme de textes élaborés, ou dans le cadre d'un robot de conversation (*Chatbot*) ;
- La classification et la prédiction par clustering ou par régression, qui permet d'entraîner des IA à détecter des anomalies ou des dossiers à risque, de générer automatiquement des échantillons de dossiers représentatifs d'un ensemble, ou encore de contribuer à la génération automatique de réponses en sélectionnant les données adaptées à la demande ;
- L'analyse d'images par réseaux de neurones : ces réseaux sont composés « d'au moins trois couches de neurones : une couche d'entrée qui reçoit des données brutes, reliée à une couche cachée qui traite ces données, elle-même reliée à une couche de sortie qui produit le résultat »⁸. Le programme est ainsi capable de proposer une interprétation des images permettant par exemple de détecter des constructions non déclarées en vue de leur intégration aux bases fiscales, ou de détecter des marchandises non déclarées à des fins de contrôle douanier.

Graphique n° 1 : Répartition des systèmes d'IA des ministères économiques et financiers selon les principales technologies utilisées



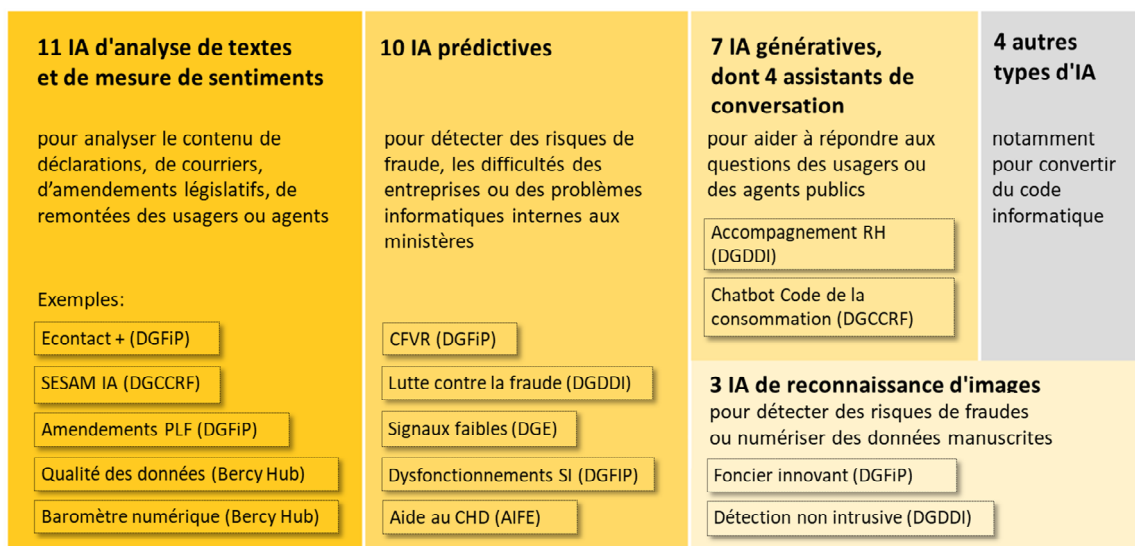
Source : Cour des comptes

⁸ Conseil d'État, *Intelligence artificielle et action publique : construire la confiance, servir la performance*, 2022, *op. cit.*, p 37.

La diversité des technologies utilisées au sein du MEFSIN répond à la variété des cas d'usage identifiés par les services comme propices à l'expérimentation de l'intelligence artificielle. Elle témoigne, à ce stade, de la capacité de ce ministère à mobiliser les différents innovations technologiques associées à l'intelligence artificielle, en s'appuyant de façon significative sur des compétences présentes en interne.

En pratique, les systèmes d'IA reposent souvent sur une combinaison de briques technologiques provenant des différentes familles présentées précédemment. Plusieurs programmes cumulent ainsi des fonctions d'analyse de données et de génération de contenu, ou de prédiction et de traitement naturel du langage. Il est cependant possible de regrouper les systèmes développés au sein du MEFSIN selon leur fonction principale afin d'en proposer un tableau d'ensemble. Ce classement fait apparaître une répartition globalement équilibrée de ces systèmes entre les principaux types d'IA.

Graphique n° 2 : Répartition par type d'IA des programmes des ministères économiques et financiers



Source : Cour des comptes. Lecture : parmi les 35 systèmes d'IA que comportent le MEFSIN, 11 ont ou auront pour fonction principale d'analyser des données textuelles, dont le programme Econtact + développé par la DGFIP.

Le recours à l'IA au sein du MEFSIN se traduit ainsi par un panel d'outils répondant à des tâches diversifiées (Cf. tableau n°1), cette variété étant présente à l'échelle de chaque direction impliquée.

Ainsi, la DGFIP exploite ou développe des IA de reconnaissance d'images (*Foncier innovant*⁹), des IA prédictives (dans le cadre des projets *Ciblage de la Fraude et Valorisation des Requêtes* et *Valorisation des cessions foncières*) et des IA génératives (dans le cadre des projets *Assistance métiers juridiques* et *Econtact GMBI*). De même, la DGDDI exploite un système expert, pour proposer des réponses aux demandes des agents concernant des actes de gestion du personnel, et développe une IA prédictive ainsi qu'un réseau de neurones dans le cadre de la lutte contre les fraudes.

⁹ L'objet des systèmes d'IA, mis en œuvre au MEFSIN entre 2018 et 2023, est décrit en annexe 2, et certains d'entre eux présentés dans le tableau 1.

Tableau n° 1 : Exemples de systèmes d'IA actuellement déployés par le MEFSIN

Nom du programme et direction concernée	Fonction de la composante d'IA au sein du programme
<p><i>ClaudIA</i> Agence pour l'informatique financière de l'État (AIFE)</p>	<p>Robot de conversation pour aider les utilisateurs de <i>Chorus Pro</i> à accomplir leurs démarches. <i>ClaudIA</i> répond à près de 80 % des questions des utilisateurs. Les demandes que <i>ClaudIA</i> ne permet pas de traiter sont transmises à des conseillers en ligne.</p>
<p><i>IA Réponse Conso</i> DGCCRF</p>	<p>Système d'IA générative conçue pour automatiser une partie des réponses aux demandes des consommateurs. Après avoir analysé le contenu des demandes et de leurs pièces jointes, le système d'IA mobilise une banque de paragraphes types pour proposer un courrier de réponse ensuite modifié et validé par un agent.</p>
<p><i>Signaux faibles</i> DGE</p>	<p>Algorithme qui calcule le risque de défaillance à 18 mois des entreprises de plus de 10 salariés, en mobilisant l'ensemble des données détenues par les services de l'État (DGE, DGEFP, Insee), les organismes de sécurité sociale et la Banque de France. Les agents habilités accèdent aux données des entreprises de leur ressort sur une plateforme sécurisée, dans laquelle ils peuvent convenir des mesures à proposer aux entreprises et suivre leur mise en place.</p>
<p><i>Ciblage de la fraude et valorisation des requêtes (CFVR)</i> DGFIP</p>	<p>Recours aux techniques d'analyse prédictive pour alimenter la programmation des contrôles fiscaux. Un service central de la DGFIP développe des algorithmes permettant d'automatiser la détection de dossiers à risques. Les cas proposés par l'IA sont transmis aux directions interrégionales de contrôle fiscal (DIRCOFI), qui vérifient leur pertinence avant l'engagement de tout contrôle, et remontent les anomalies éventuelles pour corriger les algorithmes.</p>
<p><i>Foncier Innovant</i> DGFIP</p>	<p>Recours à l'IA pour suivre l'évolution des bases de la fiscalité foncière (bâtiments, annexes, terrains). Des prises de vue aériennes, réalisées par l'IGN dans le cadre de ses missions habituelles, sont analysées par une IA qui détecte les aménagements et les compare avec les données cadastrales et fiscales. Les anomalies potentielles sont vérifiées par les agents. Le programme comprend deux fonctions : la détection des bâtis non-déclarés (déployée) et la contribution à la mise à jour du plan cadastral (en expérimentation).</p>

Source : Cour des comptes.

1.1.2 Des missions variées mais dominées par la gestion fiscale et la lutte contre les fraudes

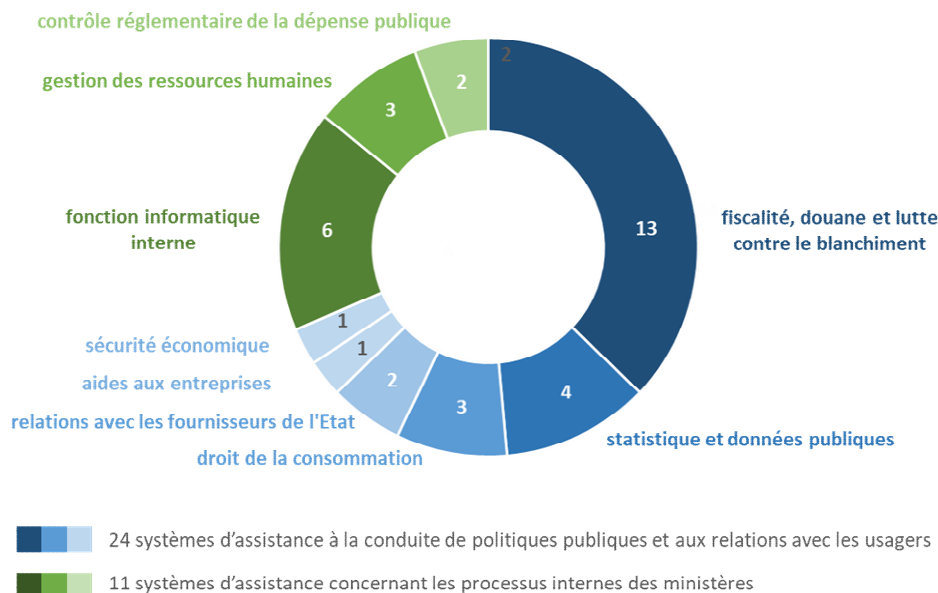
Les systèmes d'IA recensés concernent la fiscalité, les contrôles douaniers, la lutte contre le blanchiment, la statistique publique, le droit de la consommation, les aides aux

entreprises, la sécurité économique, le contrôle réglementaire de la dépense publique ou encore les relations avec les fournisseurs de l'État.

Certaines applications interviennent dans l'action extérieure du MEFSIN et d'autres sont mobilisées pour ses fonctions support : 24 systèmes d'IA concernent la conduite des politiques publiques et les relations avec les usagers, et 11 systèmes concernent ses processus internes.

Ce partage se retrouve à l'échelle de la plupart des directions impliquées : la DGFIP, la DGDDI, la DGCCRF et l'Agence pour l'informatique financière de l'État (AIFE) développent à la fois des systèmes d'aide à la conduite des politiques publiques et des systèmes d'IA concernant la gestion de leurs ressources humaines ou de leurs systèmes d'information.

Graphique n° 3 : Répartition des programmes d'IA des ministères économiques et financiers selon leur domaine d'intervention



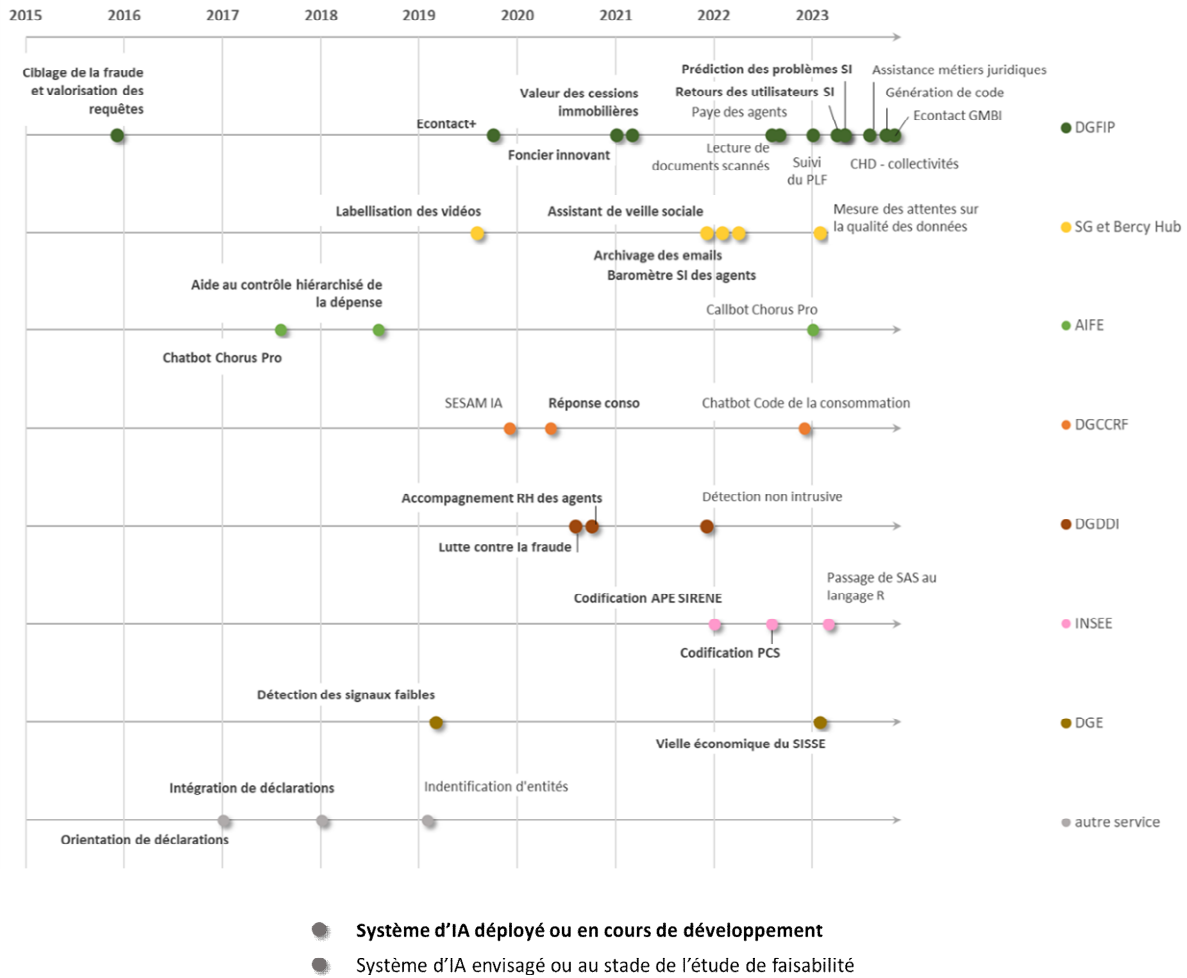
Source : Cour des comptes¹⁰

La lutte contre les fraudes (fiscalité, douanes, droit de la consommation) et le blanchiment regroupent 16 systèmes d'IA sur 35, soit 46 % du total.

Entre 2015 et 2017, les directions en charge de la gestion fiscale et de la lutte contre le blanchiment ont été les premières à initier le développement de systèmes d'IA au sein du MEFSIN, de même que l'AIFE, agence nationale qui a notamment pour mission d'entretenir et de moderniser le système d'information financier et comptable de l'État (*Chorus*).

¹⁰ L'objet des systèmes d'IA, mis en œuvre au MEFSIN entre 2018 et 2023, est décrit en annexe 2.

Graphique n° 4 : Chronologie des systèmes d'IA étudiés, développés ou exploités au sein des ministères économiques et financiers



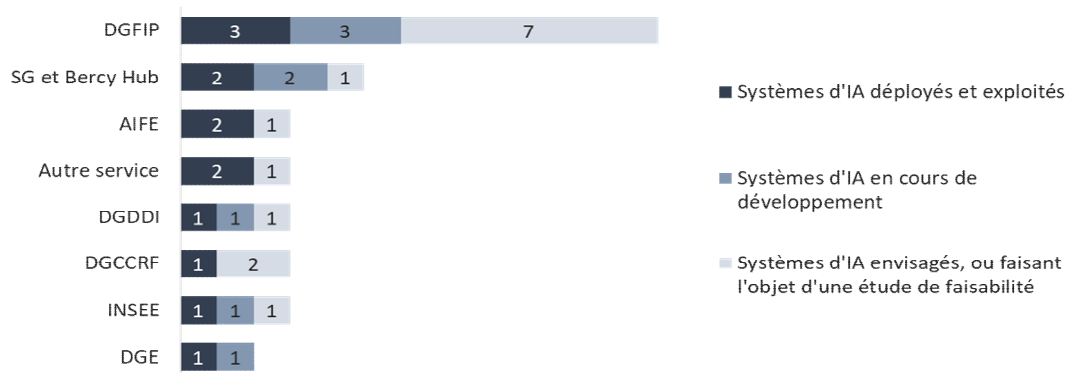
Source : Cour des comptes. Lecture : La DGFIP a commencé le développement du projet « Econtact+ » en 2019. CHD : contrôle hiérarchisé de la dépense. PLF : projet de loi de finances. SISSE : service d'intelligence stratégique et de sécurité économique¹¹.

À partir de 2019 la DGE, la DGCCRF, la DGDDI, l'Insee et le *Bercy Hub*¹², ont commencé à développer des systèmes d'IA. Fin 2023, huit directions et services des ministères économiques et financiers recourent à l'intelligence artificielle.

¹¹ L'objet des systèmes d'IA, mis en œuvre au MEFSIN entre 2018 et 2023, est décrit en annexe 2.

¹² Au sein du service du numérique du MEFSIN, le *Bercy Hub* regroupe une équipe notamment chargée d'aider les directions du MEFSIN à identifier les opportunités liées aux données de masse et de leur proposer des outils pour tester les technologies innovantes.

Graphique n° 5 : Répartition des systèmes d'IA par stade de développement et par direction



Source : Cour des comptes. Lecture : début 2024, la DGFiP exploite trois systèmes intégrant des technologies d'intelligence artificielle, développe 3 systèmes de ce type et étudie le développement de sept autres systèmes¹³.

La DGFiP est à l'origine de 13 systèmes d'IA, soit 37 % du total. Elle comporte le plus grand nombre de systèmes d'IA en service, en cours de développement et en projet.

1.1.3 Des directions d'état-major qui recourent encore insuffisamment à l'IA

Les différents services et directions qui composent le MEFSIN se répartissent entre les directions « à réseau », dont les services sont répartis sur tout le territoire (DGFiP, DGDDI, Insee), et les directions et services dits d'« état-major », qui comprennent la direction du Budget (DB), l'inspection générale des finances (IGF), le Conseil général de l'économie, de l'industrie, de l'énergie et des technologies (CGEIET) ou la direction générale du Trésor (cette dernière possédant toutefois un réseau international).

Le recensement réalisé par la Cour ne fait apparaître aucun projet d'IA concernant les directions d'« état-major » et les services d'inspection et de contrôle des ministères économiques et financiers, à l'exception du système d'IA *LLaMendements* développé par la DGFiP et pouvant intéresser la direction du budget et la direction de la législation fiscale pour le traitement des amendements parlementaires. La DG Trésor indique que certains de ses outils de calcul et de modélisation recourent à l'apprentissage machine, tels que le modèle « Opale », mais n'a signalé aucun système d'intelligence artificielle produit ou en cours de développement dans le cadre du recensement des SIA conduit par la Cour.

Or, les technologies d'IA présentent un potentiel qui mérite d'être évalué concernant certaines de leurs missions. Les intelligences artificielles de classification ou de « prédiction » pourraient par exemple présenter un intérêt pour la DG Trésor, qui a notamment pour mission d'élaborer des prévisions de conjoncture économique (croissance, compétitivité, emploi) ou d'analyser les effets potentiels de projets de réforme dans le domaine des politiques fiscales, sociales ou sectorielles. L'intelligence artificielle est par exemple l'un des trois axes du service *Innovation* de la Banque de France, qui regroupe une trentaine d'agents dont 10 *data-scientists*, et développe des projets de systèmes d'IA selon trois versants : l'exploitation transversale des données, l'outillage des agents et la remédiation des risques de cyber sécurité.

¹³ L'objet des systèmes d'IA, mis en œuvre au MEFSIN entre 2018 et 2023, est décrit en annexe 2.

Les systèmes d'IA pourraient également contribuer à optimiser certaines tâches relevant des fonctions supports de ces directions, telles que des actes de gestion de paie ou des travaux de conversion de code informatique permettant de maintenir certains serveurs en condition opérationnelle.

Enfin, l'Insee a développé trois SIA permettant d'automatiser certaines tâches. Elle a par ailleurs développé le SSP *Cloud* et encadré le développement de son interface *Onyxia* dans un cadre collaboratif. Ce projet vise à mettre à la disposition des utilisateurs professionnels des données de l'Insee des moyens de traitement des données plus diversifiés et plus accessibles grâce à un espace de travail en nuage (*cloud*), afin d'y favoriser le recours aux techniques de *data science* les plus récentes. *Onyxia* favorise ainsi le recours à l'IA, mais ne constitue pas en soi un projet de SIA.

L'Insee accorde une attention croissante aux enjeux liés à la *data science* dans le cadre de son plan « Insee Horizon 2025 ». Cependant, l'institut n'a pas encore arrêté de stratégie spécifique concernant la valorisation des nouvelles techniques de *data science*. Si la multiplication des données de masse et le développement des techniques de *data science* constituent un défi pour l'Insee, qui est également garante des principes de la statistique publique en France, l'Insee dispose cependant de ressources humaines et numériques qui devraient lui permettre de jouer un rôle encore accru dans ce domaine au sein du MEFSIN. L'Insee indique à ce titre que la gestion de ses recrutements s'efforce de tenir compte des enjeux de *data science*.

1.2 Une capacité à expérimenter l'IA hors de portée pour certains services

L'absence de projet d'IA concernant les directions « d'état-major » et l'inégale répartition des projets d'IA au sein des huit directions expérimentatrices s'expliquent en bonne partie par le positionnement dont elles ont hérité en matière de systèmes d'information.

Les directions disposant historiquement d'un parc informatique important, de services informatiques étoffés et d'un grand nombre de données de masse sont plus à même d'identifier des cas d'usages prometteurs pour les projets d'IA et de mener à bien leur développement.

S'il semble logique que la DGFIP, disposant de la plus forte empreinte informatique et administrant des échanges massifs avec les usagers, concentre à ce stade la plupart des systèmes d'IA, cette répartition inégale ne permet pas de tester leur potentiel dans chacune des principales fonctions du MEFSIN.

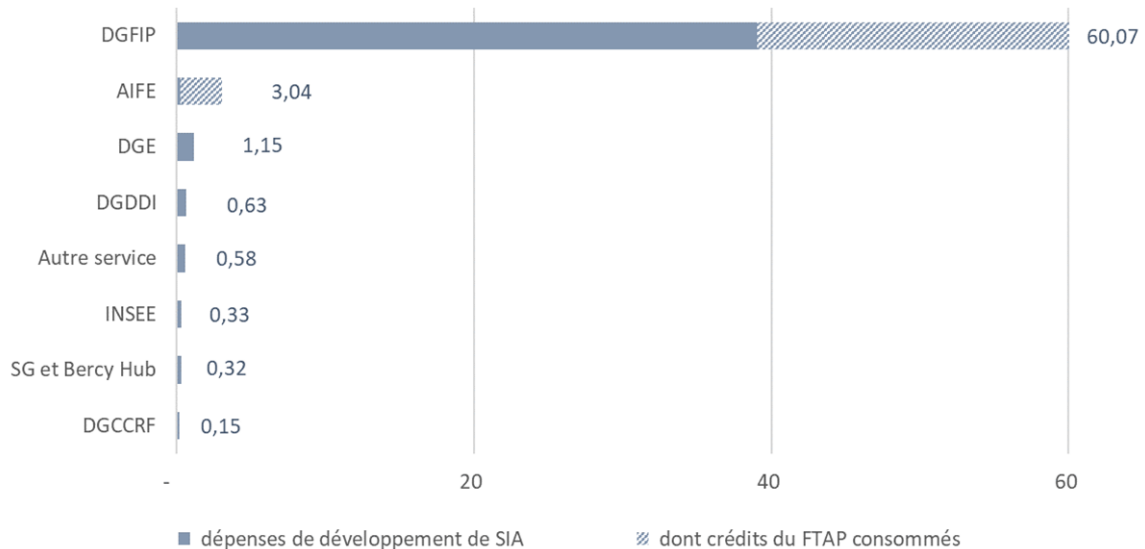
1.2.1 Des coûts de développement principalement portés par les projets de la DGFIP

Si la répartition du nombre de systèmes d'IA entre les directions du MEFSIN montre une prépondérance de la DGFIP, de l'AIFE et du *Bercy Hub*, la répartition des moyens consacrés à leur développement fait apparaître une situation encore plus contrastée.

D'après les éléments recueillis, la Cour estime à 66,3 M€ le montant total des coûts de développement des 35 systèmes d'IA recensés, sur la période 2015-2023. Ce montant inclut les

achats de prestations externes et la valorisation des ressources mobilisées en interne par les directions. Les moyens mis en œuvre par la DGFIP en représentent 91 % (60,1 M€). Les projets de la DGFIP, de l'AIFE (5 %) et de la DGE (2 %) regroupent 98 % des moyens consommés pour le développement de systèmes d'IA au sein de ce ministère depuis 2015.

Graphique n° 6 : Montant total des dépenses de développement des systèmes d'intelligence artificielle par directions pour la période 2015-2023



Source : Cour des comptes

Les coûts présentés par la DGFIP sont eux-mêmes concentrés sur deux projets : la réorganisation du ciblage du contrôle fiscal dans le cadre de *CFVR* (26,5 M€) et le programme *Foncier innovant* (27,3 M€), qui représentent ensemble 81 % des coûts de développement recensés pour la période.

Ces montants incluent les financements apportés par le Fonds pour la transformation de l'action publique (FTAP), qui permet de cofinancer la réalisation de mesures de modernisation des administrations publiques, dont des projets transformation numérique de l'État (Cf. Encadré). Sept projets de recours à l'intelligence artificielle au sein du MEFSIN ont obtenu une décision de financement par le FTAP, dont cinq ont effectivement mobilisé des crédits de ce fonds.

Le FTAP a prévu de soutenir les projets d'IA du MEFSIN à hauteur de 30,2 M€. À ce stade, 23,8 M€ ont déjà été versés, soit 80 % du soutien du FTAP qui était prévu et 40 % du total des coûts de développement (60,1 M€) mis en œuvre pour développer l'IA au sein du MEFSIN. Les financements attribués au MEFSIN représentent 18 % du soutien financier accordé par le FTAP à des projets de systèmes d'IA ministériels entre 2018 et 2022, pour un montant total de 163,5 M€¹⁴. Le MEFSIN est ainsi le troisième ministère le plus soutenu par le FTAP dans ce domaine, après le ministère des solidarités (36 M€) et le ministère de la transition écologique et de la cohésion des territoires (33,9 M€).

¹⁴ Total hors *Datalake* de la DGFIP, et hors *Veille de sécurité économique* du SISSE.

Le fonds pour la transformation de l'action publique¹⁵

Le FTAP permet de cofinancer la réalisation de mesures de modernisation des administrations publiques, dont des projets transformation numérique de l'État.

Tous les deux mois, un comité d'investissement présidé par le ministre de la transformation et de la fonction publiques ou le délégué interministériel à la transformation publique (DITP) examine les projets qui lui sont soumis. Les projets retenant son intérêt se voient désigner un rapporteur de la DITP ou de la Dinum qui accompagner les candidats pour finaliser leur projet.

À l'issue de la phase de sélection, un engagement réciproque est signé entre la DITP et le lauréat, qui définit les coûts et le rythme de déploiement du projet, le calendrier prévisionnel de versement des fonds, le montant des économies générées et les objectifs de résultats attendus. Les crédits sont mis à disposition des lauréats par tranches successives, et un suivi de l'avancement du projet est réalisé par la DITP et la Dinum pour s'assurer que les jalons clés sont respectés.

Le FTAP a été doté de 700 M€ pour la période 2017-2022, et de 330 M€ pour 2023-2025. Fin 2023, 18 projets recourant à l'intelligence artificielle sont cofinancés dans ce cadre, dont six portés par le ministère de l'économie et des finances et cinq par le ministère de la transition écologique. Le coût total de développement de ces 18 projets est estimé à 300 M€, dont 164 M€ (55 %) financés par le FTAP.

Le FTAP a ainsi cofinancé le développement de trois projets d'IA de la DGFIP : 12,1 M€ ont été accordés pour le projet de *Foncier innovant*, 5,2 M€ pour *CFVR*, et 3,7 M€ pour *Econtact+*. L'AIFE a mobilisé pour sa part 2,8 M€ de crédits issus du FTAP pour financer le développement de deux outils : le projet *ClaudIA* et le projet d'*Aide au contrôle hiérarchisé de la dépense*. Un sixième projet du MEFSIN retenu par le FTAP et porté par la DGCCRF n'a pas encore mobilisé les financements prévus.

Les moyens consacrés aux projets d'IA se concentrent donc pour l'essentiel sur la DGFIP. Le FTAP, qui a vocation à financer des projets présentant un potentiel de modernisation important et disposant déjà d'un bon niveau de maturité technique, conduit de fait à renforcer cette concentration des moyens. En effet, les directions disposant de compétences et de budgets informatiques limités sont rarement en mesure de proposer des projets susceptibles de correspondre aux attendus du FTAP.

Toutefois, ces financements n'incluent pas les coûts d'entretien et de maintien en condition opérationnelle (MCO) des systèmes d'IA. À ce stade, ces coûts semblent maîtrisés concernant les SIA les plus importants. Ainsi, le coût de fonctionnement et d'entretien annuel du programme *CFVR* s'établit à 4 M€ dont 3,5 M€ de dépenses de personnel, ce qui est globalement conforme aux prévisions associées au projet. Pour de nombreux SIA, les coûts de fonctionnement sont par ailleurs limités (17 700 € annuels pour le *Chatbot* d'accompagnement RH de la DGDDI, 42 000 € annuels pour le SIA *Réponse Conso* de la DGCCRF).

Cependant, les coûts de MCO n'ont pas été systématiquement renseignés par les services interrogés dans le cadre de cette enquête. La soutenabilité de ces coûts présente des enjeux plus importants pour les directions qui dépendent de prestataires extérieurs pour l'entretien de leurs SIA. Les directions du MEFSIN doivent donc rester attentives aux coûts

¹⁵ Voir également : Cour des comptes, *La modernisation de l'État : des méthodes renouvelées, une ambition limitée*, novembre 2023, disponible en ligne : <https://www.ccomptes.fr>

d'entretien, dont les montants doivent être systématiquement indiqués et fiabilisés avec précision avant le passage en développement de tout projet d'IA.

1.2.2 Des compétences internes sur le champ de l'IA inégalement réparties

La plupart des directions du MEFSIN disposent d'un nombre limité d'agents ayant des compétences requises pour concevoir et mener à bien le développement de systèmes d'IA.

Parmi les services interrogés, la délégation au numérique de la DGFIP rassemble le plus grand nombre d'agents disposant des compétences nécessaires au développement de systèmes d'IA (analystes de données, *data engineer*, *data scientist*, directeur ou directrice des données). La délégation à la transformation numérique (DTNUM) de la DGFIP regroupe ainsi 18 agents sur ce périmètre de compétences, pour une rémunération brute annuelle totale de 1 M€, tandis que le *Bercy Hub* du SG dispose d'effectifs de moitié moindre, avec neuf agents pour une rémunération brute annuelle totale de 0,5 M€.

Cinq agents sont consacrés à des fonctions de *data science* au sein de la DGE, dont trois au sein du pôle « données » du service de l'information stratégique et de la sécurité économiques (Sisse¹⁶). De même, le pôle « données » du service du MEFSIN en charge de la lutte contre le blanchiment comporte également trois agents.

La DG Trésor ne dispose pas d'effectifs pleinement dédiés à des fonctions de *data scientist*. Elle dispose cependant d'agents ayant des compétences analogues, parmi lesquels des statisticiens et des administrateurs de l'Insee, répartis dans ses différents bureaux métiers. La DG Trésor doit à présent organiser une « communauté » pour recenser ces compétences et les mobiliser dans le cadre de la feuille de route directionnelle de la donnée.

Les agents maîtrisant les compétences nécessaires au développement de systèmes d'IA au sein du MEFSIN sont donc concentrées au sein de la DGFIP, de l'Insee et du secrétariat général (*Bercy Hub*). Au sein des autres directions, ces compétences sont présentes en nombre très limité.

À l'exception de la DGFIP, de l'Insee et du *Bercy Hub*, les services du MEFSIN ne sont donc pas en mesure de tester l'intérêt des technologies d'intelligence artificielle à un rythme plus rapide ou sur un panel de métiers plus large que le périmètre des travaux actuellement conduits.

1.2.3 Un accès différencié aux données de masse et à la puissance de calcul

La distribution actuelle des systèmes d'IA reflète également la répartition des données de masse entre les différentes administrations du MEFSIN. En effet, la capacité à développer des systèmes d'IA dépend largement de l'accès à des données de masse (*big data*) présentant un degré de structuration et d'intégrité suffisant pour servir à l'entraînement du système.

¹⁶ Le service de l'information stratégique et de la sécurité économiques (Sisse) élabore et propose la politique de sécurité et de protection des intérêts économiques français.

Or l'organisation des différents jeux de données dont disposent le MEFSIN (données fiscales, économiques, budgétaires, comptables, démographiques, etc.) reste globalement cloisonnée entre ses différents services.

La DGFIP et l'Insee occupent à cet égard une position particulière, car leurs missions les conduisent à collecter, héberger et administrer de très nombreuses séries de données de masse. Elles disposent pour ce faire d'infrastructures conséquentes et ont accès à une puissance de calcul globalement adaptée au traitement de leurs données. À l'inverse, les autres directions du MEFSIN administrent des bases de données nettement plus limitées et ne disposent généralement pas de la puissance de calcul qui leur permettrait de développer de façon autonome des systèmes d'IA.

Pour ces directions, l'accès aux données de la DGFIP et de l'Insee susceptibles de les intéresser est organisé au cas par cas et présente des fonctionnalités limitées.

Cette situation s'explique par plusieurs niveaux de contraintes. D'une part, les données ne sont parfois disponibles qu'au sein de l'application de gestion qui les mobilise pour des missions déterminées. Il est dans ce cas difficile de prévoir l'accès de partenaires extérieurs sans remettre en cause la continuité d'exploitation de l'application. La consultation des données peut être ouverte à travers des « infocentres » attachés à chaque application, qui fonctionnent selon des modalités très diverses. D'autre part, des contraintes juridiques s'appliquent à la protection des données selon leur niveau de sensibilité. Afin de garantir notamment le respect des dispositions du RGPD, la DGFIP ou l'Insee procèdent à des habilitations au cas par cas, dans le cadre de contrats de service, en vérifiant notamment la finalité d'accès aux données formulée par les demandeurs. Enfin, les impératifs de protection face aux cybermenaces conditionnent les modalités de partage des données, en limitant par exemple leur consultation aux postes informatiques inclus dans un périmètre de sécurité offrant des garanties suffisantes.

Si ces contraintes ne font pas obstacle, sous les conditions précitées, au décloisonnement des données au sein du MEFSIN, elles posent cependant des défis techniques qui majorent le coût de ce décloisonnement et ralentissent sa progression.

L'accès aux données repose sur une gestion des accès au cas par cas, selon des modalités peu ergonomiques, qui ne permettent pas la valorisation des données à des fins d'IA.

Ainsi, la DG Trésor accède aux données sociales de l'Insee au moyen d'un boîtier fourni par son centre d'accès sécurisé aux données (CASD) et équipé d'un lecteur de cartes magnétiques. Ce système permet d'accéder aux données sans formuler de demandes ponctuelles, mais il présente de délais de réponse parfois longs et ne peut être utilisé qu'au sein des locaux de la DG Trésor. Par ailleurs, la manipulation des données et des algorithmes ne peut être effectuée qu'au sein du boîtier, les éventuelles exportations faisant l'objet d'une procédure d'autorisation spécifique. Ce système permet donc la mise à disposition des données sociales de l'Insee, mais il ne permet pas aisément de les soumettre à des travaux de *data science*.

L'accès des directions du MEFSIN aux données de la DGFIP est également géré au cas par cas, à travers des contrats de service. Le projet de lac de données conduit par cette direction, qui devait permettre d'« industrialiser » la gestion des accès, n'a pas mis en œuvre cette fonction pour les tiers extérieurs à la DGFIP (Cf. 1.3.1).

L'accès aux données de masse et à la puissance de calcul requise pour entraîner certains systèmes d'IA demeure donc limité pour la plupart des directions du MEFSIN. Dans ce contexte, les services recourent parfois à des alternatives qui permettent de contourner cet

obstacle lors de la phase initiale des projets, mais qui ne sont pas durables et peuvent poser des problèmes de sécurité technique et juridique.

Le Snum évoque ainsi un cas de recours à une offre de puissance de calcul dite « gratuite » pour procéder à l'étude de faisabilité d'un projet, mais qui n'a pas permis la poursuite du développement. La DG Trésor indique par ailleurs que certains agents ont acquis à titre personnel des licences pour recourir au service web d'hébergement et de développement de logiciels *GitHub* de Microsoft pour tester des outils d'intelligence artificielle, ce qui peut poser des problèmes de sécurité et de confidentialité. La DG Trésor a entamé un recensement en vue d'encadrer plus strictement à l'avenir le recours aux solutions proposées notamment par Microsoft et Open AI.

L'accès aux données de masse, aux puissances de calcul, aux outils collaboratifs et aux compétences permettant de développer des systèmes d'IA reste donc très limité pour la plupart des directions du MEFSIN.

1.3 Un pilotage qui doit se structurer pour garantir un développement cohérent du recours au *data-mining* à l'IA

1.3.1 Une répartition des ressources qui n'est pas corrigée par les dispositifs ministériels et interministériels

Certains dispositifs semblent en mesure d'apporter un appui aux directions manquant de ressources pour expérimenter des technologies d'intelligence artificielle. D'une part, au sein du secrétariat général, les équipes du service du numérique (Snum) peuvent être sollicitées par les directions sectorielles pour leur apporter un appui technique aux projets d'IA. D'autre part, le FTAP, qui dispose de ressources financières importantes (330 M€ pour la période 2023-2025), peut être mobilisé pour cofinancer certains d'entre eux. Enfin, les projets de décloisonnement des données de masse entre les différents services du MEFSIN pourraient permettre à diverses directions d'accéder aux *big data* hébergées notamment par la DGFIP et l'Insee.

Cependant, ces dispositifs ne permettent pas encore, dans leur configuration actuelle, de soutenir les expérimentations des directions les moins dotées en ressources pour l'IA.

Ainsi, le Snum dispose d'effectifs trop limités pour apporter un appui technique durable à l'ensemble des directions qui en auraient besoin. De fait, si le *Bercy Hub* du Snum a contribué au développement de cinq systèmes d'IA depuis 2019, ces applications sont destinées aux services du SG et non aux autres directions métier de MEFSIN. Par ailleurs, le *Bercy Hub* conduit des actions de formation et de valorisation transversale des bases de données, mais il dispose de ressources qui ne lui permettent pas d'accompagner dans la durée les projets de systèmes d'IA des directions métier les moins avancées en la matière.

Par ailleurs, le Snum dispose d'attributions limitées et éprouve des difficultés à acquérir une vision complète des chantiers d'IA en cours au sein du MEFSIN et à proposer une synthèse des enjeux qui leur sont liés en termes d'infrastructures ou de commande publique.

De plus, au niveau interministériel, le FTAP tend à financer des projets à fort impact, qui sont par nature complexes et proposés par des directions disposant déjà d'une capacité de portage significative. En effet, ce fonds sélectionne des projets dont le potentiel de modernisation et les gains de productivité attendus sont étayés par les porteurs de projet et évalués par la DITP et la Dinum. Leur calendrier technique est documenté avec précision et ses jalons sont suivis attentivement par la DITP. La progression du projet doit être régulière et les crédits versés peuvent être repris s'ils ne sont pas exécutés après un délai de six mois.

La mobilisation du FTAP implique de documenter les gains attendus, de s'engager au respect d'un calendrier précis et de disposer d'équipes en capacité d'absorber les travaux de développement et de conduite du changement. Les directions le plus à même de mobiliser les financements du FTAP sont donc généralement celles qui concentrent déjà les données, les infrastructures et les compétences nécessaires au développement des IA. Ce fonds ne semble donc pas pouvoir contribuer à rééquilibrer l'inégale répartition des ressources de développement d'IA entre les différents services.

Enfin, les projets portés par la DGFIP et l'Insee pour faciliter l'exploitation des données de masse par les différents services du ne permettent pas encore d'accélérer significativement le recours à l'IA et aux techniques de *data science*.

La DGFIP a entamé un programme de décloisonnement de ses données en créant un lac de données (*datalake*) et en modernisant sa gestion des API (Cf. Encadré). Ce projet complexe a marqué un progrès, mais ce décloisonnement reste essentiellement interne à la DGFIP et ne profite quasiment pas aux autres services du MEFSIN.

Le projet « Datalake et API Management » de la DGFIP et la *data science*

Selon le contrat de transformation¹⁷ conclu dans le cadre du FTAP, le projet de « Valorisation et de mise à disposition des données de la DGFIP » repose sur deux volets :

- *Un volet portant sur le stockage des données de la DGFIP dans un « lac de données »* : celui-ci doit permettre de stocker ces données en les libérant des contraintes auxquelles elles sont soumises lorsqu'elles se trouvent au sein des applications de gestion. Le stockage des données dans le *datalake* doit ainsi faciliter leur utilisation pour différents usages (croisement avec d'autres données, analyse prospective, *data science*). À cet égard, le contrat prévoit que le *datalake* propose des outils de croisement et de visualisation des données.

- *Un volet portant sur l'accessibilité des données* : un composant doit permettre « d'exposer les données de la DGFIP à l'ensemble de ses partenaires », sous réserve des exigences du RGPD. Ce composant de « management d'API » doit permettre « d'industrialiser » la gestion des accès des différents services au sein et à l'extérieur de la DGFIP, pour leur offrir un accès plus fluide tout en garantissant le respect des impératifs de protection et de sécurité des données.

Le projet de valorisation des données de la DGFIP devait ainsi profiter à ses partenaires extérieurs et le contrat de transformation mentionnait des cas d'usage au profit de l'IGF, du GiP Union Retraite ou du ministère de la justice¹⁸. Le projet de *datalake* pourrait ainsi constituer un atout pour l'expérimentation des systèmes d'IA, en permettant aux différents services d'accéder plus facilement à

¹⁷ Ministère de l'action et des comptes publics, Contrat de transformation du 14 janvier 2020 relatif au projet « Datalake et API Management : valorisation et mise à disposition des données de la DGFIP », p. 2-5.

¹⁸ *Idem*, p. 4.

ces données de masse et donc d'envisager de nouveaux cas d'usages pour le développement de systèmes d'IA.

Le management d'API a permis d'optimiser la valorisation des données pour certains métiers de la DGFIP. Le *datalake* contribue ainsi au fonctionnement du système d'IA *Foncier Innovant* : les prises de vue aériennes sont croisées avec les informations fiscales au sein du lac de données. À moyen terme, le système d'IA *CFVR* pourrait également mobiliser les données du *datalake*, de même que le bureau des études statistiques de la DLF. Enfin, le lac de données permet à la DGFIP de répondre plus facilement aux demandes de documents adressées à par les particuliers et les collectivités territoriales.

Il n'est cependant pas accessible aux autres directions du MEFSIN selon les modalités d'API management décrites dans le contrat de transformation. Le lac de données leur est ouvert ponctuellement, selon une procédure qui reste lourde du fait des enjeux de sécurité. Le projet apparaît à cet égard en deçà des engagements pris dans le cadre du contrat conclu entre la DGFIP et la DITP.

Dès lors, le lac de données ne constitue pas, à ce stade, un élément de réponse au cloisonnement des données qui limite le recours à l'intelligence artificielle au sein du MEFSIN. La délégation au numérique de la DGFIP indique que plusieurs jalons supplémentaires seraient à franchir pour lui permettre de contribuer plus largement à des projets d'IA. Ces limites concernant notamment le rythme d'actualisation des données par les directions métier qui alimentent le lac¹⁹, le niveau de sécurité des réseaux et des postes informatiques, très variable selon les directions du MEF, et les fonctionnalités de traitement limitées qu'offre le *datalake*. En effet, ses fonctions reposent sur le modèle de « l'infrastructure en tant que service » (*infrastructure as a service, IaaS*) : le lac de données fournit un environnement de consultation virtuelle des données, mais les utilisateurs doivent souvent utiliser leur propre système d'exploitation et leurs propres logiciels pour valoriser ces données.

En outre, le SSP *Cloud* et son interface *Onyxia*, initiés par l'Insee, visent également à améliorer l'accès des directions du MEFSIN aux données de masse et à leur traitement par des outils de *data science*. Destiné à ce stade aux statisticiens publics²⁰, il pourrait offrir des perspectives pour le développement ultérieur de prototypes de SIA, à condition d'être accessible à davantage de services du MEFSIN, dans le cadre d'une stratégie spécifique de valorisation des techniques de *data science* qu'il reviendrait à l'Insee de définir, en articulation avec une stratégie ministérielle du développement des SIA qui reste à construire.

1.3.2 Un pilotage ministériel à établir pour garantir la cohérence du recours à l'IA au sein du MEFSIN

L'inégale répartition des moyens de développement d'IA au sein du MEFSIN n'a pas empêché une relative diversité des politiques publiques et des missions dans lesquelles l'assistance de l'IA a été testée jusqu'à présent. En outre, la concentration des moyens sur la DGFIP est en cohérence avec le volume et la nature de ses missions. Elle reflète l'importance

¹⁹ L'alimentation du lac de données dépend de la disponibilité des services métiers et n'est pas prioritaire par rapport aux missions de gestion qui mobilisent leurs systèmes d'information.

²⁰ Cf. [Le SSPCloud : une fabrique créative pour accompagner les expérimentations des statisticiens publics – Courrier des statistiques N7 - 2022 | Insee](#), p. 68.

de son système d'information et de ses effectifs, qui représentent 75 % des effectifs du MEFSIN. Le recours à l'IA par la DGFIP est par ailleurs pertinent compte tenu des enjeux portés par la modernisation d'une direction assurant de nombreux services publics concernant entreprises, les contribuables particuliers et professionnels ou encore les collectivités territoriales.

Cependant, certaines directions sont restées à l'écart des expérimentations de l'IA, alors que ces technologies peuvent apporter des gains pour la qualité et l'efficacité de leur action. Or, aucun des dispositifs identifiés au cours de l'enquête ne permet de compenser cette lacune.

Il semble ainsi nécessaire d'identifier l'instance ministérielle en capacité de garantir l'équilibre et la cohérence de développement des systèmes d'IA au sein du MEFSIN, et d'accompagner certaines directions métier par une offre de services en *data science*. D'autres enjeux plaident pour la définition d'un pilotage ministériel de l'IA. D'une part, il est nécessaire de garantir une mutualisation efficace de l'expérience et des communs numériques développés par chaque équipe au sein des différentes directions. D'autre part, il est indispensable de veiller à la cohérence entre le développement des systèmes d'IA et la feuille de route ministérielle de la donnée. L'interdépendance entre les IA et les données de masse implique une programmation interministérielle concertée dans laquelle le MEFSIN doit pouvoir jouer un rôle moteur, afin de mutualiser davantage les infrastructures disponibles et de rationaliser l'identification des infrastructures nécessaires à l'avenir. Par ailleurs, il est nécessaire de disposer d'un interlocuteur de niveau ministériel pour jouer le rôle d'interface entre l'action de transformation numérique interministérielle portée par la DITP et la DINUM, et les directions métiers du MEFSIN. Enfin, l'accompagnement des administrations du MEFSIN dans la mise en œuvre de l'IA nécessite d'avoir recours à une activité de veille et d'anticipation ministérielles en la matière, tant ses usages évoluent.

Pour répondre à ces différents enjeux, la Cour recommande d'identifier l'instance ministérielle le pilotage stratégique de la donnée et de l'IA, tout en préservant les initiatives ascendantes et les compétences techniques associées qui permettent à chaque direction de détecter des cas d'usages et de tester la validité des IA au plus proches de ses métiers.

Cette entité serait notamment chargée de sécuriser la présence des trois prérequis à la mise en œuvre de l'IA, à savoir la qualité et la disponibilité des données, des infrastructures et des compétences, et de garantir un développement équilibré et mutualisé des systèmes d'IA entre les différents métiers et les différents publics. Elle pourrait s'appuyer sur les nombreux rapports²¹ et mobiliser les ressources et compétences numériques du MEFSIN dans le cadre d'un travail en réseau.

À titre d'exemple, le ministère de l'Europe et des affaires étrangères s'est récemment doté d'une direction du numérique qui comporte un incubateur pour les « intrapreneurs » et un pôle ministériel dédié à l'intelligence artificielle. Le ministère de la justice a créé en janvier 2023 une direction du numérique²² au sein de son secrétariat général, qui a notamment en charge

²¹ Entre autres : IGF et conseil général de l'économie de l'industrie, de l'énergie et des technologies, Les ressources humaines de l'État dans le numérique, janvier 2023 ; Insee et DINUM, Évaluation des besoins de l'État en compétences et expertises en matière de donnée, septembre 2021.

²² Arrêté du 29 janvier 2024 modifiant l'arrêté du 30 décembre 2019 relatif à l'organisation du secrétariat général et des directions du ministère de la justice

de développer et d'entretenir des systèmes d'intelligence artificielle²³. Le ministère de l'intérieur a placé sa filière numérique sous la responsabilité d'un secrétaire général adjoint (SGA) en charge du numérique, qui assure également les fonctions de directeur de la transformation du numérique (DTNUM)²⁴.

L'organisation actuelle du MEFSIN semble en retrait par rapport à ces exemples. Sur le plan technique, le Snum présente des ressources et des attributions limitées (cf. 1.3.1). Sur le plan du pilotage ministériel, le MEFSIN dispose d'un comité stratégique du système d'information qui réunit les directeurs et directrices généraux et les chefs des services du MEFSIN tout au plus deux fois par an.

Cette instance dispose d'un champ de compétences étendu, puisqu'elle peut examiner « la planification budgétaire des directions, les projets informatiques les plus importants, les choix technologiques, la sécurité des systèmes d'information, le développement et l'usage du numérique et toute question commune relative aux systèmes d'information »²⁵. Toutefois, l'arrêté définissant ses attributions ne mentionne pas spécifiquement la gestion des données de masse, ni le recours à l'intelligence artificielle, ni l'articulation du MEFSIN avec la stratégie interministérielle du numérique. L'enquête conduite par la Cour n'a pas fait apparaître d'action spécifique de ce comité dans les domaines de l'IA et de la *data science*.

Recommandation n° 1. (SG des MEF) : Identifier l'instance ministérielle en charge du pilotage stratégique de l'intelligence artificielle et des données numériques.
--

1.3.3 Une offre de services à définir en faveur des directions métier en mobilisant la DINUM et les instruments financés par la stratégie nationale d'IA

La mise en œuvre de l'IA nécessite un appui méthodologique dans la gestion de projets et la commande publique des capacités de calcul²⁶. Elle doit donc être mutualisée. Intégrer des standards de performance, de robustesse et d'évaluation dans les cahiers des charges des marchés publics liés à l'IA nécessite des connaissances techniques approfondies. La commande publique dans le domaine de l'IA, indispensable au développement des entreprises du secteur, serait plus fluide en assistant les responsables de la commande publique dans la rédaction des cahiers des charges.

Parallèlement à l'identification d'une instance en charge du pilotage ministériel de l'IA, la Cour recommande la mise en place d'un incubateur d'IA au sein du MEFSIN. Un incubateur d'IA a vocation à internaliser et à mutualiser les compétences en gestion de projet, structuration des données, calcul et conception algorithmique. Il doit permettre de créer des produits qui favorisent la souveraineté, l'accessibilité et la sécurité numérique au sein de l'État.

²³ Art. 4 de l'arrêté précité.

²⁴ Décret n° 2023-582 du 5 juillet 2023 modifiant le décret n° 2013-728 du 12 août 2013 modifié portant organisation de l'administration centrale du ministère de l'intérieur et du ministère des outre-mer

²⁵ Art. 1^{er} de l'arrêté du 22 août 2017 relatif à la gouvernance du système d'information des ministères économiques et financiers.

²⁶ Les ministères peuvent d'ores et déjà faire appel à la DINUM, en contactant son équipe *Datalab*, pour bénéficier des capacités interministérielles et partager avec elle tout projet d'investissement en capacités de calcul

Au sein du MEFSIN, l'incubateur apporterait un appui technique aux directions dont les moyens de *data science* sont limités et diffuserait dans l'ensemble des directions les lignes directrices du pilotage ministériel, ainsi que les bonnes pratiques et les outils proposés au niveau interministériel notamment par la Dinum. Il assurerait un contrôle-qualité sur les projets selon le crible du *produit minimum viable* (MVP²⁷), permettant d'éviter le risque d'une dispersion des financements sur des projets ne seraient pas viables *in fine*.

Le ministère de l'agriculture et de la souveraineté alimentaire (MASA) est ainsi accompagné par la Dinum pour créer un incubateur, afin de susciter des projets et accompagner des intrapreneurs dans les phases de création ou de développement.

L'incubateur du MEFSIN serait constitué de l'équipe actuelle du *Bercy Hub*, dont les moyens seraient renforcés par la mobilisation des compétences de *data-science* présentes dans les directions métiers, sous la forme d'une communauté de travail structurée en réseau et bien identifiée au sein du ministère. L'administrateur ministériel des données, algorithmes et codes sources (AMDAC) en serait membre.

Recommandation n° 2. (SG des MEF, Dinum) : Créer un incubateur d'IA au sein du MEFSIN avec l'aide de la Dinum.

La création d'une équipe ministérielle (« *task force OlympIA* »), amorcée par le secrétariat général des MEFSIN au printemps 2024, constitue à ce titre une initiative à saluer. Constituée d'agents issus de quinze directions du ministère travaillant en réseau, cette équipe devra faciliter l'accès des directions d'état-major aux technologies d'IA, contribuer au décloisonnement des données de masse et garantir une meilleure appropriation des outils disponibles au niveau interministériel.

En effet, les outils d'accompagnement de la Dinum sont rarement identifiés par les porteurs de projet de systèmes d'IA du MEFSIN.

Or, la Dinum propose une méthode de conduite de projet pour promouvoir l'innovation (les start-ups d'État²⁸). Son offre de services en matière d'IA a été étoffée²⁹ à la suite de la médiatisation des solutions d'IA générative (LLM) en fin d'année 2022. Elle met à disposition des outils (comme les entrepreneurs d'intérêts généraux et la brigade d'intervention numérique³⁰, une IA générative souveraine, dénommée « Albert » et un incubateur de projets IA interministériel, appelé « *alliance* », pour accompagner les administrations dans le déploiement de projets en lien avec l'intelligence artificielle. En plus de proposer du mentorat

²⁷ En anglais, *minimum viable product* (MVP).

²⁸ Les start-ups d'État renvoient à des produits numériques développés par de petites équipes autonomes, de manière itérative et très proche des utilisateurs finaux. Cette méthode est vue par ses promoteurs comme une manière d'améliorer la qualité du service public et d'accroître l'acculturation des administrations à l'innovation.

²⁹ La feuille de route intitulée « Une stratégie numérique au service de l'efficacité de l'action publique », publiée en 2023, prévoit que la Dinum anime et professionnalise la filière des ressources numériques de l'État et assure la souveraineté numérique de l'État.

³⁰ Le programme « Entrepreneurs d'Intérêt Général » (EIG) vise à faciliter le recrutement d'experts numériques sur des projets d'amélioration du service public proposés par les administrations. Le dispositif des commandos UX facilite la mise à disposition auprès des administrations de designers et développeurs recrutés par l'État pour simplifier les démarches en ligne et améliorer l'expérience utilisateur.

et de l'aide à la conduite de projet, l'incubateur IA offre à ces derniers une connexion avec des représentants du secteur privé, du monde académique (INRIA, CNRS, Centrale-Supelec), de la commission nationale Informatique et Libertés (Cnil) et de l'Agence nationale de sécurité des systèmes d'information.

Les coopérations du MEFSIN avec la recherche publique devraient cependant être facilitées alors que celles-ci sont rares. Sur les 35 systèmes d'IA du MEFSIN, seul le système d'IA de la DGCCRF, *Réponse Conso*, a fait l'objet d'une coopération avec un laboratoire de recherche.

Par ailleurs, le MEFSIN gagnerait à utiliser directement les instruments financés depuis 2018 dans le cadre de la stratégie nationale d'IA, lancée le 28 mars 2018 pour positionner la France comme un acteur majeur de l'IA³¹.

La stratégie nationale d'intelligence artificielle (SNIA)

La SNIA a été élaborée à la suite des avancées de l'apprentissage profond et dans le contexte de la publication des rapports de France Stratégie en 2017, de celui de M. Cédric Villani en 2018, des travaux conduisant à la stratégie européenne pour l'IA publiée par la Commission européenne en avril 2018³², et de l'adoption par l'OCDE en mai 2019³³ des principes sur l'intelligence artificielle.

Dans sa première phase qui a duré de 2018 à 2022, la SNIA avait identifié cinq domaines à soutenir avec 1 527 M€ de financements publics. La transformation publique constituait un de ces cinq volets, avec la recherche, l'enseignement supérieur, la diffusion dans l'économie et la défense et la sécurité. La diffusion de l'intelligence artificielle et des sciences des données dans l'économie et dans l'administration et la promotion d'un modèle éthique équilibré entre innovation et protection des droits fondamentaux constituait deux des trois priorités de la SNIA³⁴.

Le volet transformation publique de la SNIA et la diffusion de l'IA et des sciences des données dans l'administration manifestaient l'obligation d'adaptabilité ou de mutabilité des services publics³⁵. Le bon fonctionnement du service public suppose en effet une adaptation aux changements de la société, aux besoins des usagers et aux évolutions techniques. Le financement prévisionnel de la transformation publique par l'Etat représentait 10 % du budget initial de la SNIA³⁶. Depuis 2022,

³¹ Concernant la SNIA, voir Cour des comptes, *La stratégie nationale de recherche en intelligence artificielle une stratégie à structurer et à pérenniser*, rapport public thématique, avril 2023.

³² Un plan de coordination européenne dans le domaine de l'IA a été publié en décembre 2018.

³³ OCDE, *Recommandations du Conseil sur l'intelligence artificielle*, adoptées en mai 2019, modifiées en novembre 2023

³⁴ La 3e priorité était de développer l'attractivité des talents.

³⁵ Le juriste Louis Rolland a théorisé au début du XXe siècle trois grands principes régissant le bon fonctionnement du service public et s'appliquant à toutes ses activités : la continuité, l'adaptabilité et l'égalité. Le principe de continuité prescrit un fonctionnement régulier des services publics, sans interruption autre que celles autorisées par la législation. Le principe d'adaptabilité impose au gestionnaire d'un service public de faire évoluer son organisation en fonction des progrès techniques, des besoins des usagers et de toute circonstance nouvelle. Le principe d'égalité et son corollaire le principe de neutralité soulignent l'importance d'assurer à tous, sans aucune discrimination, un égal accès aux services publics. Ils correspondent à des principes généraux du droit administratif français. Les principes de continuité et d'égalité sont des objectifs à valeur constitutionnelle.

³⁶ Cour des comptes, *la stratégie nationale de recherche en intelligence artificielle*, avril 2023, page 30. Le volet recherche devait être financé à hauteur de 445 M€, le volet défense et sécurité, 410 M€, le volet économie, 390 M€, le volet transformation publique, 130 M€ et le volet enseignement supérieur, 128 M€.

la SNIA se déploie dans une nouvelle phase, dite d'"accélération". La transformation publique par l'IA n'y apparaît plus nommément.

Grâce à la SNIA, le supercalculateur Jean Zay³⁷ du CNRS met par exemple à disposition plus de 3100 GPUs ainsi que du stockage de masse à haut débit souverain. Une équipe de treize experts accueille les utilisateurs sur le centre, les aide à transférer leurs données, installer leurs outils, les dépanner, optimiser leurs applications et se former. Selon l'équipe en charge de la coordination de la SNIA, « Jean Zay » a soutenu plus de 1200 projets académiques et industriels en IA. Le MEFSIN a eu recours au supercalculateur Jean Zay pour développer deux projets d'IA (Foncier Innovant et un projet de LLM de la DGFIP) sur trente-cinq SIA recensés.

L'offre de service de la SNIA doit être adaptée pour permettre un accès aisé des administrations publiques à ses différentes ressources, dont le supercalculateur, les appels à projet en faveur des communs numériques, l'observatoire des cas d'usages résolus par l'IA³⁸, la recherche publique Grand défi « IA de confiance », le programme « Confiance.ai » et les évaluations lancées dans le cadre du programme *Digital Europe*³⁹.

Dans ces conditions, la Cour invite les administrations du MEFSIN à se saisir plus régulièrement des ressources de la Dinum et des instruments financés par la SNIA. La Cour incite la Dinum à mener une action de sensibilisation permettant aux directions du MEFSIN de mieux connaître cette offre. C'est à cette condition que la transformation publique par l'IA deviendrait manifestement un objectif prioritaire de la SNIA⁴⁰, comme initialement prévu lors de sa première phase.

CONCLUSION INTERMÉDIAIRE

Les systèmes d'intelligence artificielle développés depuis 2015 au sein du MEFSIN mobilisent la plupart des technologies d'IA les plus récentes et témoignent de la capacité de ses services à développer des programmes concernant aussi bien les relations externes que les fonctions support de l'administration.

Cependant, cette relative diversité masque le fait que certaines directions restent en marge des expérimentations en cours. En l'absence d'une démarche pilotée au niveau

³⁷ Selon l'équipe en charge de la coordination de la SNIA, Jean Zay est : « l'un des supercalculateurs les plus écoresponsable du monde de par son refroidissement interne mais aussi le mécanisme de récupération de la chaleur fatale de la machine pour chauffer plus de 1000 logements sur le plateau de Saclay grâce à un investissement du CNRS et de l'EPAPS, une première en Europe à cette échelle ».

³⁸ Déclaration du Président de la République, sur les efforts en faveur du développement de l'intelligence artificielle, Paris le 21 mai 2024. Piloté conjointement par le SGPI et la DGE, l'observatoire des cas d'usage doit permettre aux entreprises qui veulent intégrer une solution d'IA de savoir si le cas d'usage a déjà été résolu et donc acheter une solution sur étagère, ou bien si elles doivent passer par un projet de R&D. Le montant et l'origine du financement n'ont pas été définis. L'observatoire est également appelé référentiel des cas d'usage testé et approuvé.

³⁹ Le programme Digital Europe, également nommé programme pour une Europe numérique, finance des projets en matière de supercalcul, d'intelligence artificielle, de cybersécurité, de compétences numériques avancées. Doté d'un budget global prévu de 7,5 milliards d'euros (en prix courants), il fait partie du cadre financier pluriannuel 2021-2027.

⁴⁰ La deuxième phase de la SNIA vise à accélérer la diffusion de l'IA dans l'économie, notamment en soutenant l'offre « deep tech » et en rapprochant l'offre et la demande en IA. Elle était financée initialement à hauteur de 1,2 Md€ puis a été réduite à 1 Md€.

ministériel, le développement de l'IA risque de se concentrer sur la gestion financière et fiscale et la lutte contre les fraudes, sans que son potentiel puisse être testé dans d'autres fonctions, notamment celles qui sont dévolues à la DG Trésor ou à la direction du budget.

D'autres enjeux plaident pour la définition d'un pilotage ministériel de l'IA. Il est ainsi nécessaire de garantir une mutualisation efficace de l'expérience et des algorithmes développés par les différentes directions, des données de masse et des infrastructures prévues pour leur hébergement. Enfin, le MEFSIN a besoin d'une interface ministérielle efficace avec l'action de transformation numérique interministérielle portée par la DITP et la Dinum et les instruments financés par la SNIA.

La Cour recommande au MEFSIN d'identifier une instance à même de piloter le recours à l'IA à l'échelle ministérielle, tout en préservant les logiques ascendantes qui permettent à chaque direction de tester la validité des IA au plus proche de ses métiers.

La Cour recommande par ailleurs la création d'un incubateur IA fonctionnant en réseau et rattaché au secrétariat général des ministères économiques et financiers. Amorcée au printemps 2024, la mise en place de cette équipe permettrait d'éviter au MEFSIN la dispersion des financements sur des projets qui ne peuvent finalement pas être étendus à l'échelle souhaitée.

2 IA ET GAINS DE PRODUCTIVITÉ : UN ENJEU D'EFFICACITÉ, DES EFFETS AMBIVALENTS

Le recours à l'intelligence artificielle permet de produire des prévisions aussi fiables, voire plus fiables que les prévisions reposant sur le seul facteur humain, tout en diminuant très fortement leur coût⁴¹. L'IA permet dès lors de générer des gains de productivité, la productivité étant définie comme le rapport entre une production et les ressources mises en œuvre pour l'obtenir.

Lorsque ces gains de productivité permettent de libérer des emplois au sein des administrations publiques, celles-ci doivent décider de leur affectation. Les emplois peuvent être maintenus sur la même mission, redéployés sur d'autres fonctions, ou restitués dans une optique de maîtrise des dépenses publiques. Ces décisions ont donc des conséquences sur l'efficacité socio-économique, la qualité du service et l'efficacité budgétaire de l'action publique. Elles concernent ainsi les agents et les décideurs publics, mais également les citoyens à travers le point de vue de l'utilisateur et du contribuable⁴².

⁴¹ OCDE, « Paysage économique de l'IA », in *L'intelligence artificielle dans la société*, éditions de l'OCDE, Paris 2019.

⁴² Ces « points de vue » sont les sous-jacents utilisés pour la construction des indicateurs de performance budgétaire des documents annexés aux projets de lois de finances. Cf. Direction du budget, *Guide de la performance*, mars 2015.

Par ailleurs, le recours à l'IA a des effets ambivalents sur la nature des missions confiées aux agents et sur leur cadence de travail. L'IA peut ainsi libérer les agents de certaines tâches à faible valeur ajoutée, mais accroître à l'inverse des tâches répétitives telles que la relecture des propositions de la machine. En rendant certains services plus performants ou plus accessibles, elle peut par ailleurs susciter un effet de rebond sur l'activité des services.

L'allocation des gains de productivité permis par l'IA doit donc reposer sur des informations fiables, qui permettent d'anticiper les conséquences de ces technologies en termes de ressources humaines. Or, l'IA est très faiblement abordée dans le cadre du dialogue social ministériel, et la conduite des projets réserve une place trop limitée aux implications quantitatives et qualitative des systèmes d'IA pour le travail des agents concernés.

2.1 Les systèmes d'IA améliorent la productivité mais ne permettent pas nécessairement de libérer des emplois

2.1.1 Un nombre limité de systèmes d'IA réduisent le coût d'une action à périmètre constant

Sur 35 SIA, quatre systèmes ont été conçus pour réduire le coût de l'action de l'administration à périmètre constant, et un système d'IA (*ClaudIA*, cf. 2.2.1) a permis de diminuer les dépenses liées au traitement des demandes tout en augmentant le nombre de demandes traitées.

Ainsi, *CFVR* a diminué les coûts de programmation du contrôle fiscal sans modifier le volume ou les objectifs de cette programmation. Le nombre de dossiers proposés au contrôle reste stable, mais ce résultat est obtenu en mobilisant 427 ETP de moins en 2023 par rapport à 2018 : 50 % de la programmation concernant les professionnels et 30 % de la programmation concernant les particuliers reposent désormais sur une équipe de 32 personnes assistées par le *data-mining*, contre environ 500 personnes en 2018.

De même, les programmes *Chatbot Econtact*, *Foncier innovant* et *IA CHD* ont été associés, dès leur programmation, à une trajectoire de restitution des moyens ainsi économisés.

Les gains de productivité liés à ces systèmes d'IA ont permis de libérer des emplois qui ont été restitués par les directions concernées. L'IA permet alors une baisse de dépense par rapport aux coûts passés (Cf. 2.1.2). Toutefois, le recours aux systèmes d'IA peut accroître dans certains cas le volume d'activité des services, même lorsque leur introduction a été envisagée à charge constante. Ainsi, l'administration peut maintenir inchangé le nombre de contrôles fiscaux avant et après l'introduction de *CFVR*, mais, dans le cas d'*Econtact* et de *Foncier innovant*, le recours aux systèmes d'IA conduit à une augmentation de l'activité des services, au moins de façon temporaire (Cf. 2.2.).

2.1.2 Une majorité de systèmes d'IA stabilisent le coût d'une action tout en augmentant ses résultats

La majorité des systèmes d'IA recensés sont développés pour améliorer les résultats des services ou absorber une augmentation du volume d'activité, sans augmentation des dépenses. Dans ce cas, les emplois dégagés par le recours à l'IA sont redéployés, de façon programmée ou spontanée, pour améliorer la productivité globale à moyens constants.

L'IA permet dans certains cas d'améliorer la qualité et les délais de réponse à effectifs constants. Ainsi, le système d'IA *LLaMmendements*⁴³ libère les agents des tâches consistant à classer et à résumer les amendements du Parlement⁴⁴, mais ce gain est immédiatement réemployé pour améliorer la qualité et les délais des réponses apportées. *LLaMendements* permet donc aux équipes d'absorber la charge de travail (85 345 amendements déposés entre juin 2022 et janvier 2024) sans recourir à des moyens humains supplémentaires. De même, les projets de systèmes d'IA concernant la veille de sécurité économique du SISSE, ou la valorisation des données d'enquêtes de la DGCCRF, visent à améliorer la détection des alertes de sécurité économique et le ciblage des enquêtes à moyens humains constants.

D'autres services ont mobilisé l'IA pour augmenter leur volume d'activité sans augmenter leur coût. La labellisation et l'orientation des remontées des professionnels soumis à des obligations déclaratives, dans le cadre de la lutte contre la fraude fiscale et le blanchiment, ont ainsi été automatisées. Ce recours à l'IA a permis de redéployer vingt-cinq agents sur des fonctions d'enquête pour faire face à l'augmentation des signalements à moyens constants.

Dans ces cas de figure, l'IA permet une baisse de dépenses par rapport à un scénario alternatif, dans lequel ces processus auraient été améliorés en recourant à des effectifs supplémentaires.

De même, le robot conversationnel *ClaudIA* a été conçu pour accompagner la généralisation de l'interface *Chorus Pro* à l'ensemble des fournisseurs de l'État. Entre 2018 et 2020, *Chorus pro* est devenu obligatoire pour les ETI, les PME et les TPE. Le nombre de sollicitations par les usagers est ainsi passé de 20 000 à 100 000 par mois et le robot *ClaudIA* a permis d'absorber cette charge sans y consacrer d'effectifs supplémentaires. Dans ce dernier cas, les gains de productivité réalisés ont été supérieurs aux besoins générés par l'extension de *Chorus Pro*, permettant la restitution d'emplois et une baisse des dépenses par rapport à leur niveau antérieur à l'implémentation du système d'IA. Ainsi, le contrat de transformation conclu avec la DITP au titre de *ClaudIA* prévoyait la restitution de vingt-huit ETP pour une économie annuelle de 1,3 M€. L'AIFE indique que 55% des économies prévues ont été réalisées à ce stade.

⁴³ L'objet de ce système d'IA est décrit en annexe 2.

⁴⁴ À titre d'exemple, le projet de loi de finances pour 2024 a été accompagné de plus de 3000 amendements. Sous la XV^e législature (2017-2022), les députés ont déposé 200 000 amendements en séance, en forte augmentation par rapport à la dernière législature (115 200 amendements entre 2012 et 2017).

2.1.3 Certains systèmes d'IA peuvent prendre en charge des missions nouvelles

Certains systèmes d'IA effectuent des tâches qui n'étaient pas assurées auparavant. Dans ce cas, l'IA ne permet pas de dégager des emplois pour assurer de nouvelles fonctions ou contribuer à la maîtrise des dépenses publiques.

Ainsi, le recours à l'IA pour assurer l'archivage de la correspondance électronique des cabinets ministériels du MEFSIN ou automatiser l'écriture de lignes de code informatique, permet d'agir sur des volumes de données qui n'étaient pas traités. De même, la détection précoce des difficultés des entreprises par le programme *Signaux Faibles* (DGE) et le système d'IA projeté par la DGDDI pour détecter des objets suspects dans des colis sans les ouvrir, permettent d'accomplir des tâches qui ne seraient pas réalisables à une échelle significative sans recourir à l'IA.

S'ils ne libèrent pas d'emplois par rapport à la situation antérieure, ces systèmes d'IA peuvent générer des gains de productivité indirects et des économies budgétaires à moyen terme. Ainsi, la détection précoce des difficultés des entreprises peut rendre l'action de l'État plus efficace et globalement moins coûteuse en la matière.

2.1.4 Le recours à l'IA peut entraîner un effet de rebond sur l'activité des services

Lorsque la productivité associée à un service augmente, son rapport qualité-prix s'améliore. Les consommateurs tendent alors à privilégier ce service, devenu plus accessible et plus avantageux. Si le producteur augmente son offre pour répondre à cette demande, il mobilise finalement davantage de facteurs de production que durant la période précédente.

Certains économistes ont ainsi qualifié d'« effet rebond »⁴⁵ la tendance par laquelle les économies liées à l'usage d'une nouvelle technologie sont en partie annulées par la demande additionnelle que ce progrès entraîne. Cette tendance peut se manifester lorsque l'intelligence artificielle est introduite dans un processus de production⁴⁶.

Ainsi, les *Chatbots* permettent aux usagers de solliciter l'administration plus facilement et sans contrainte horaire, ce qui peut provoquer une augmentation du volume global des demandes, et en leur sein, du nombre de demandes complexes pour lesquelles l'IA n'est pas en mesure de proposer de réponse. L'administration peut alors être amenée à consacrer au total plus de moyens humains au traitement des demandes qu'avant l'introduction de l'IA.

L'amélioration d'une étape de la production grâce à l'IA peut aussi avoir des conséquences sur les étapes situées en amont et en aval dans le processus d'ensemble.

⁴⁵ La notion d'effet rebond est initialement mise en évidence dans les rapports entre consommation d'énergie et production économique (William Stanley Jevons, 1865). Pour une définition plus récente, voir Steve Sorrell, « The Rebound Effect : definition and estimation », in *International handbook on the economics of energy*, 2009 ; Taoyuan Wei and Yang Liu, « Estimation of global re-bound effect caused by energy efficiency improvement » in *Energy Economics* 66, 2020, et David Stern, « How Large is the Economy-Wide Rebound Effect ? » in *CAMA Working Paper*, n°70, 2020.

⁴⁶ Cf. notamment Martina Willenbacher et al. « Rebound Effects in Methods of Artificial Intelligence », in *Advances and New Trends in Environmental Informatics*, 2021.

En amont, un travail d'acquisition et de structuration des données est nécessaire au fonctionnement de l'IA. Une partie de ce travail peut être reporté sur les usagers, lorsque ceux-ci répartissent par exemple des informations sur les différents champs d'un formulaire en ligne, mais cette charge échoit dans certains cas à l'administration. Par ailleurs, l'entretien du système d'IA et sa mise à jour régulière nécessitent un travail supplémentaire de l'administration.

En aval, un système d'IA qui permet une meilleure détection des suspicions de fraudes peut augmenter le travail d'enquête et de régularisation de la part des services. Ainsi, *Foncier innovant* permet d'augmenter la détection de bâtis potentiellement non-déclarés à l'administration fiscale, tout en présentant une marge d'erreur plus importante que lorsque cette détection est assurée par les géomètres du cadastre. Le recours à un SIA, en diminuant le coût de la détection, conduit la DGFIP à disposer d'un plus grand nombre de biens dont la situation doit être vérifiée puis régularisée par les agents, ce qui augmente *in fine* la charge de travail des équipes.

De même, les algorithmes utilisés par le programme *CFVR* pourraient se perfectionner à l'avenir et proposer une plus grande proportion de dossiers en situation de fraude au sein de la programmation du contrôle fiscal. *CFVR* conduirait alors à augmenter l'activité des équipes, à nombre de contrôles inchangé, car les contrôles se traduiraient plus souvent par la détection de fraudes, le rappel de droits et les procédures contentieuses et de recouvrement qui s'y attachent. Cette logique pourrait s'appliquer au projet *Valorisation des cessions immobilières* (VCI), qui doit permettre à l'administration fiscale d'estimer de façon plus précise et mieux étayée la valeur vénale des biens immobiliers afin de détecter certains types de fraudes.

L'introduction d'un système d'IA peut donc modifier la structure de l'activité globale d'un service à l'échelle d'une mission donnée. Leur mise en œuvre peut modifier la répartition de certaines tâches entre les agents et les usagers, et alléger ou au contraire alourdir les besoins en amont et en aval de l'étape traitée par l'IA. Dans certains cas, l'effet rebond est susceptible de consommer tout ou partie des emplois libérés par l'IA.

2.2 Des économies avérées, mais moins importantes qu'estimées initialement

2.2.1 Une cible totale de 61 M€ d'économies annuelles pour cinq systèmes d'IA

Les baisses de dépenses budgétaires permises par les systèmes d'IA du MEFSIN se concentrent sur cinq projets conduits par la DGFIP et l'AIFE entre 2015 et 2021, sur les 35 projets recensés au sein du MEFSIN.

Ces projets ont sollicité et obtenu un cofinancement du fonds de transformation de l'action publique (FTAP, Cf. 1.2.1). Dans ce cadre, les directions responsables des projets ont estimé les économies permises par le recours aux systèmes d'IA et se sont engagées à les réaliser selon un calendrier déterminé.

Les contrats de transformation conclus au titre du FTAP indiquent une estimation du temps de travail économisé par le recours à l'IA, sous la forme d'ETP valorisés en euros. Des

frais de fonctionnement par agent sont parfois comptabilisés. Les contrats concernant ces systèmes d'IA présentent également le cadencement annuel des diminutions d'ETP ainsi permises jusqu'à l'achèvement complet des projets en 2024.

À cette date, les économies prévues devaient permettre d'accomplir les missions concernées en mobilisant au total 1 013 ETP de moins qu'en 2019, soit une économie de 60,7 M€ en année pleine, dont 500 ETP rendus en lien avec *CFVR*, 300 ETP concernant *Foncier innovant*, 138 ETP en lien avec *Econtact* et 75 ETP concernant les projets de l'AIFE (*ClaudIA* et *IA CHD*).

Des gains de recettes fiscales sont par ailleurs attendus de certains SIA. *Foncier innovant* devait ainsi permettre 130 M€ de recettes annuelles additionnelles « pour les collectivités locales à horizon 2022 »⁴⁷, tandis que *CFVR* devait contribuer à l'amélioration du recouvrement des droits rappelés à l'issue des contrôles fiscaux⁴⁸, le montant de ce gain n'étant toutefois pas évalué.

2.2.2 Des économies constatées pour un montant de 20 M€ en 2022

Pour assurer le suivi des économies réalisées, les contrats prévoient que les porteurs de projet transmettent chaque année à la DITP le montant des économies réalisées et l'explication des écarts éventuels avec les cibles initialement retenues.

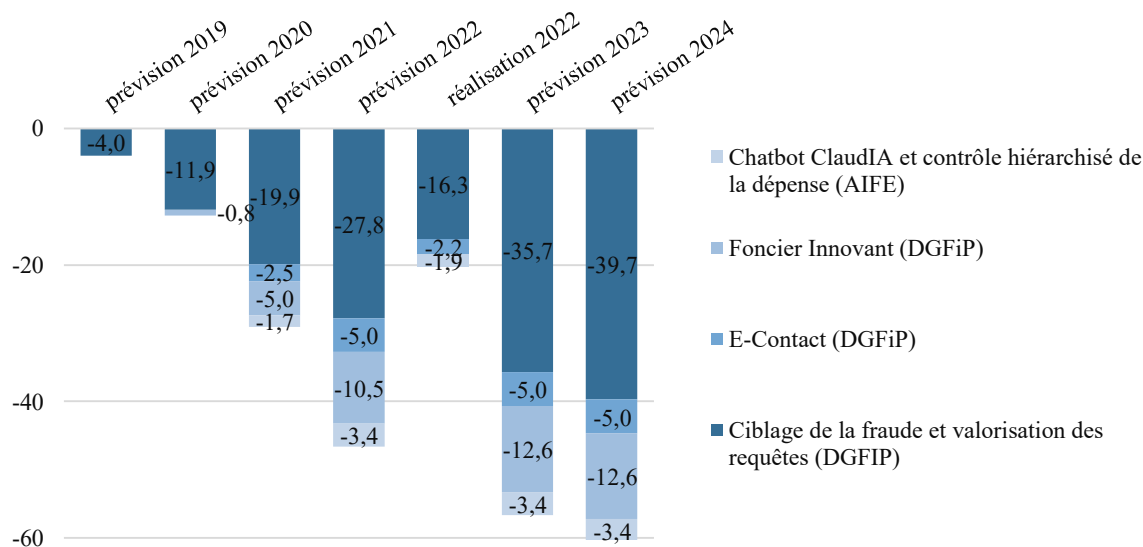
D'après les dernières données transmises par la DGFIP et l'AIFE, qui concernent l'année 2022, le recours aux cinq systèmes d'IA concernés a généré 20,4 M€ d'économies contre 46,6 M€ attendus pour cet exercice.

Cet écart est amplifié par l'absence de données concernant le projet *Foncier Innovant* en 2022. Les données transmises à la DITP pour ce système d'IA ne font apparaître aucune restitution d'ETP, sans qu'aucune explication ne soit apportée.

⁴⁷ Contrat de transformation relatif au projet *Foncier Innovant* du 14 octobre 2019, p 6.

⁴⁸ Contrat de transformation relatif au projet *CFVR* du 10 décembre 2018, p 5.

Graphique n° 7 : Économies prévues par les contrats de transformation des cinq projets d'IA cofinancés par le FTAP et montants constatés en 2022



Source : Cour des comptes, contrats de transformation conclus au titre du FTAP.

Pour les quatre systèmes d'IA dont les données ont été transmises, les économies réalisées en 2022 peuvent être estimées à 20,4 M€ pour une cible de 36,1 M€, soit une cible atteinte à 56 %.

2.2.3 Un écart qui témoigne d'un suivi incomplet et d'hypothèses budgétaires parfois optimistes

L'écart entre les économies constatées et les économies visées résulte en premier lieu d'un suivi qui demeure incomplet. Cela concerne en premier lieu *Foncier Innovant*, pour lequel la DITP, en charge du FTAP, n'a pas obtenu de données chiffrées. Par ailleurs, si les données concernant *CFVR* contiennent une description précise et fidèle des jalons réalisés sur le plan opérationnel, elles ne donnent aucune explication de l'écart entre les économies prévues (27,8 M€) et les économies réalisées (16,3 M€). Il en va de même pour les projets de l'AIFE.

De plus, certaines hypothèses retenues par la DGFIP pour évaluer les effets budgétaires des systèmes d'IA apparaissent élevées et peu justifiées. Ainsi, le coût moyen des ETP économisés en lien avec *CFVR* était estimé à 79 400 € annuels lors de la présentation du projet (l'économie portant à 70 % sur des emplois de catégorie A) tandis que le coût constaté a été de 36 000 € en 2022. Par ailleurs, la DGFIP a retenu un coût annuel de 43 000 € pour les emplois de catégorie A dans le contrat de transformation du système d'IA *Econtact*, conclu un an plus tard.

De même, les gains de recettes fiscales attendus au titre du projet *Foncier Innovant* ont été estimés à « 130 M€ à horizon 2022 » par la DGFIP dans le contrat de transformation, sans que ce montant ne soit expliqué. En 2022, ces gains se sont élevés à 4,4 M€ de recettes pérennes et 5,7 M€ concernant la rectification d'impositions antérieures, concernant les neuf départements ayant expérimenté le dispositif. La DGFIP estime à présent que 40 M€ de recettes pérennes devraient être apportées par le dispositif après sa généralisation en 2023.

La DGFIP semble avoir recouru dans certains cas à des hypothèses optimistes dont les sous-jacents auraient mérité un examen plus attentif de la DITP.

2.2.4 Le FTAP, un dispositif structurant dont les indicateurs peuvent être enrichis

Le FTAP met en œuvre une évaluation prévisionnelle des gains de productivité qui permet de prévoir et d'organiser leur affectation. Si cette affectation a pris la forme d'une restitution nette des emplois dans le cas des systèmes d'IA cofinancés par le FTAP concernant le MEFSIN, ce dispositif peut aussi être utilisé pour des projets prévoyant des redéploiements, dès lors que le projet permet d'améliorer l'efficacité globale et la qualité du service rendu par l'administration.

Il constitue à ce titre un cadre pertinent visant à assurer un emploi efficace des ressources publiques. Cependant, dans le cas des projets d'IA, l'évaluation de la pertinence du projet devrait être enrichie pour mieux tenir compte des effets de ces outils sur les étapes situées en amont et en aval de la chaîne de traitement.

Ainsi, les conséquences des systèmes d'IA sur les besoins en amont et en aval de la chaîne de traitement n'ont pas été documentés dans le cadre du FTAP. De fait, le recours au système d'IA *Foncier innovant* accroît initialement la charge de travail des agents du fait des capacités de traitement du SIA, qui a permis d'augmenter de 5,7 % (+ 20 356) le nombre de piscines imposées à la taxe foncière dans les départements expérimentateurs. La charge de travail est également accrue du fait de ses limites, car *Foncier Innovant* génère de nombreux faux positifs qu'il revient aux agents d'exclure du processus. Le recours à un système d'IA pourrait également modifier les habitudes des usagers et augmenter sensiblement la charge des points de contacts téléphoniques et des services en contact avec les usagers.

Cependant, les contrats de transformation n'indiquent pas si les cadencements des diminutions d'ETP intègrent ou non ce type d'effets⁴⁹. De même, ils ne précisent pas s'ils incluent les besoins nouveaux liés à l'entretien et à la mise à jour des systèmes d'IA (sauf dans le cas de *CFVR*).

Par ailleurs, l'évaluation prévisionnelle des projets de systèmes d'IA devrait prendre en compte de façon plus systématique la qualité des réponses et des résultats attendus de l'IA.

Ainsi, sur l'ensemble des indicateurs des cinq systèmes d'IA concernés, un seul est relatif au niveau de satisfaction des utilisateurs (*Chatbot Econtact*), et un seul sur la qualité des résultats (taux de précision de l'algorithme de *Foncier Innovant*). Or la qualité des réponses attendues est un critère essentiel pour évaluer la pertinence du recours à un SIA, et peut constituer une limite pour certains d'entre eux. Ainsi, une cible de 65 % de taux de satisfaction des utilisateurs était fixée pour le *Chatbot Econtact* dans son contrat de transformation, mais le niveau de satisfaction était de 50% en 2023 (49 % en 2022).

La présentation des projets de systèmes d'IA gagnerait donc à s'enrichir d'indicateurs concernant la qualité des réponses des systèmes d'IA et du service rendu aux usagers. Ces indicateurs doivent être systématiques et leur réalisme doit être évalué à l'aune des résultats issus de la phase préalable au développement (*proof of concept*) et en comparant les résultats

⁴⁹ Les notes d'exécution budgétaire de la mission transformation et fonction publiques 2021, 2022, 2023 recommandent de compléter les indicateurs de performance du FTAP.

annoncés aux résultats effectivement atteints par les systèmes d'IA du même type développés par le passé.

2.3 Une affectation des gains de productivité qui doit gagner en visibilité

2.3.1 L'affectation des économies réalisées grâce aux IA soulève des enjeux particuliers pour l'administration

Lorsque les systèmes d'IA permettent de libérer des emplois, l'administration doit décider de leur affectation. Les emplois peuvent être maintenus sur la même mission, redéployés sur d'autres fonctions, ou restitués dans une optique de diminution des dépenses. Ces arbitrages peuvent ainsi contribuer à l'extension du champ d'intervention de l'administration, à l'amélioration de la qualité du service et à la maîtrise des dépenses publiques.

Certains systèmes d'IA permettent de contribuer simultanément à plusieurs de ces objectifs. Le *Chatbot ClaudIA* permet par exemple d'augmenter le nombre de réponses aux usagers tout en diminuant les dépenses qui y sont consacrées. Dans d'autres cas, l'allocation des emplois libérés par l'IA peut être plus complexe, notamment si l'introduction d'un système d'IA génère des attentes supplémentaires de la part des usagers ou un surcroît d'activité pour l'administration.

En fonction de leur stratégie de développement, les entreprises choisissent de répondre à la demande additionnelle générée par l'amélioration de leur productivité, ou de maintenir leur production à un niveau constant pour diminuer leurs charges. Ce choix est cependant plus contraint pour l'État, car des impératifs particuliers s'imposent à son action.

L'administration fiscale est par exemple chargée de veiller au respect des obligations des redevables et à l'égalité des contribuables face à la loi fiscale. Dès lors qu'une IA permet d'augmenter la détection de biens non déclarés, l'administration est tenue de régulariser leur situation fiscale.

Dans certains cas, l'administration plafonne les résultats des systèmes d'IA afin de maîtriser l'effet rebond qui peut en résulter. Le programme *CFVR* est géré de façon à maintenir constant le nombre de dossiers proposés au contrôle fiscal, alors que l'algorithme pourrait générer davantage de propositions. De même, le traitement automatisé des données relatives à la détection de cas de blanchiment est calibré pour identifier une quantité de dossiers adaptée aux capacités de traitement des équipes d'enquête.

Dans d'autres cas, l'administration ne contrôle pas tous les leviers permettant de maîtriser l'effet rebond éventuel. C'est notamment le cas lorsque les systèmes d'IA sont en interaction directe avec le public. Ainsi, les *Chatbots* permettent à un plus grand nombre d'usagers d'interagir simultanément avec l'administration, mais chaque usager a le droit d'obtenir, dans des délais raisonnables, une réponse de qualité, sécurisée sur le plan juridique et vérifiée par un agent si nécessaire.

Ces interrogations doivent être traitées lors des phases préalables au développement et à la mise en production des SIA. Or ces enjeux semblent faiblement documentés à l'échelle des projets et au niveau ministériel.

2.3.2 Les conséquences des IA en termes de ressources humaines ne sont pas suffisamment anticipées

Les phases d'études et de développement des projets de systèmes d'IA du MEFSIN se concentrent sur leurs paramètres techniques et intègrent de façon limitée leurs incidences en termes de gestion des ressources humaines.

D'après les éléments recueillis durant l'enquête, les conséquences des systèmes d'IA sur la trajectoire d'emplois, le positionnement des agents et leur charge de travail font très rarement l'objet d'une concertation avec les agents et les organisations syndicales.

Entre 2018 et 2020, la DGFIP et l'AIFE avaient documenté précisément l'affectation des emplois économisés par le recours à l'IA, dans le cadre des cinq projets cofinancés par le FTAP. De par leur ampleur et en l'absence de tout redéploiement, ces projets ont pu contribuer à ce que l'IA soit exclusivement perçue comme un moyen de réduction de l'emploi public par les agents et les organisations syndicales, alors même que la majorité des projets recensés au sein du MEFSIN ne sont pas spécifiquement motivés par la réduction des effectifs.

Cette évolution a contribué à ce que les implications RH de l'intelligence artificielle ne soient plus abordées dans le cadre du dialogue social. Dès lors, la prévision et l'affectation des gains de productivité liées aux grands projets d'IA ne font pas l'objet d'une présentation transparente et partagée, ce qui ne permet pas d'anticiper convenablement les mesures d'accompagnement nécessaires.

Enfin, dans ce contexte, les directions sont amenées à présenter les projets de systèmes d'IA exclusivement sous l'angle de l'amélioration de la qualité de service et de vie au travail. Or cette stratégie oblitère une partie de l'information nécessaire pour évaluer la pertinence des projets et décider de l'affectation des gains de productivité, sans écarter pour autant la défiance des partenaires sociaux. Elle conduit par ailleurs les porteurs de projets à insister sur des améliorations qualitatives qu'ils ne sont pas toujours en capacité de prouver.

En effet, le principe général selon lequel l'introduction d'un système d'IA permet de repositionner les agents sur des tâches à plus forte valeur ajoutée n'est pas toujours applicable, et semble être assez rarement concrétisé. La réalisation d'un tel repositionnement est documentée précisément dans un cas sur les treize systèmes d'IA actuellement en service au sein du MEFSIN, et a porté sur vingt-cinq agents.

Une enquête de satisfaction sur l'usage des SIA, réalisée auprès de 7 800 agents de la DGFIP par l'une des organisations syndicales en 2023 indique que 85 % d'entre eux ne sont pas satisfaits du recours à l'IA, considérant notamment que « les systèmes d'IA employés actuellement ne conduisent pas à une valeur ajoutée plus forte que les systèmes qu'ils remplacent »⁵⁰. De fait, l'implémentation des systèmes d'IA peut nécessiter des tâches répétitives consistant à mettre en forme les données en amont, ou à trier les réponses en aval de la machine. De même, l'intégration d'un *Chatbot* ne permet pas nécessairement de libérer les agents de tâches répétitives pour les concentrer sur les demandes les plus complexes, car les réponses proposées par l'IA doivent être relues et contrôlées par les agents. Selon le contexte et les modalités du projet, il n'est donc pas exclu que les agents passent d'une position de

⁵⁰ Enquête menée par l'organisation syndicale *Solidaires*.

rédaction à une position de relecture, sans que la valeur ajoutée ou la diversité de leur travail ne soient significativement améliorés.

À l'inverse, si une application d'IA permet d'automatiser l'ensemble des tâches simples d'un agent, celui-ci devra accomplir le même volume de travail qu'auparavant, « sans le répit que pouvaient lui procurer l'alternance » entre tâches faciles et missions complexes⁵¹.

Les systèmes d'IA ont donc des effets ambivalents sur la nature, la cadence et la diversité des missions confiées aux agents, comme l'ont souligné plusieurs enquêtes⁵² conduites par l'OCDE sur ce sujet. Si l'IA permet une amélioration de la santé des actifs dans de nombreux cas, elle débouche fréquemment sur une accélération des cadences et une intensification du travail. Ces enjeux doivent être pris en compte et documentés dès la phase d'étude préalable au développement des systèmes d'IA au sein du MEFSIN.

2.3.3 Une prospective à construire à l'échelle des projets et à l'échelle ministérielle

Qu'il s'agisse de systèmes d'IA structurants pour l'évolution des effectifs, ou de projets aux implications plutôt qualitatives, l'anticipation des enjeux de ressources humaines et la qualité des échanges entre les décideurs, les équipes métier, les équipes techniques et les partenaires sociaux sont insuffisantes. Or, comme l'a déjà souligné la Cour⁵³, le succès d'une transformation numérique ambitieuse dépend en grande partie de l'anticipation et de l'accompagnement des implications en termes de ressources humaines.

À l'échelle de chaque projet, les implications en termes de productivité, d'emplois, de formation et de conditions de travail devraient être envisagées et documentées dès le stade de la preuve de concept, selon des modalités adaptées à la taille du projet. La preuve de concept (POC) technique pourrait ainsi être accompagnée d'une preuve de concept relative aux ressources humaines. Ce POC « ressources humaines » indiquerait les gains de productivité nets attendus, les affectations envisagées, les effets du système d'IA sur la nature des missions des agents et les actions envisagées en termes d'accompagnement et de formation. Pour les plus grands projets, ces données permettraient d'enrichir les études MAREVA⁵⁴ pour disposer d'une vision plus complète de l'impact du système d'IA envisagé.

À l'échelle ministérielle, ces données devraient faire l'objet d'une synthèse mise à jour régulièrement, pour alimenter le dialogue social relatif à l'évolution des effectifs et des compétences.

⁵¹ OCDE, « Intelligence artificielle, qualité des emplois et inclusivité », in *Perspectives de l'emploi de l'OCDE 2023*, Éditions OCDE, Paris, 2023.

⁵² Voir notamment Marguerita Lane, Morgan Williams et Stijn Broecke, *The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers*, OCDE, Paris, 2023, et Bill Condie et Leigh Dayton, « Four AI technologies that could transform the way we live and work », in *Nature*, vol. 588/7837, 2020.

⁵³ Cour des Comptes, « Le numérique au service de la transformation de l'action publique », in *Rapport public annuel 2020*, février 2020, p 12.

⁵⁴ Les études dites « MAREVA » (Méthode d'Analyse et de Remontée de la Valeur) estiment la valeur prévisionnelle des projets de système d'information, en rapportant leur coût estimé et leurs bénéfices attendus, afin de produire un indice de retour sur investissement.

Concernant ce second aspect, le MEFSIN présente un fort manque de visibilité. Ainsi, alors que le recours à l'IA et aux processus d'automatisation numérique pourrait avoir des effets importants à l'avenir sur la répartition des compétences nécessaires aux missions du MEFSIN, le travail de prospective ministérielle a été lancé tardivement. Le SG des MEF⁵⁵ a ainsi commandé un rapport relatif aux conséquences de l'IA sur les métiers du contrôle fiscal et douanier, confié en avril 2023 à deux chercheurs du *LaborIA* au sein de l'INRIA, suivi par un comité scientifique du *Labor IA* et un comité de pilotage administratif.

Cette démarche devrait être complétée par une cartographie interne des principaux processus pour lesquels le recours à l'IA pourrait apporter des gains significatifs pour les prochaines années. Le rapport commandé par le MEFSIN à INRIA concernant l'IA et les métiers du contrôle fiscal et douanier, constitue une étape pour orienter la stratégie relative à l'IA et aux ressources humaines. Piloté par le service des ressources humaines du secrétariat général du MEFSIN, le recensement préconisé serait partagé avec les partenaires sociaux et comparé aux conclusions du rapport de l'INRIA, lorsque celles-ci seront disponibles. Le MEFSIN disposerait ainsi dès 2025 d'une cartographie à l'échelle ministérielle, complétée par une analyse plus précise concernant les métiers du contrôle fiscal et douanier, afin de réintégrer l'intelligence artificielle dans le champ du dialogue social sur des bases objectivées.

Recommandation n° 3. (SG et directions du MEFSIN) Identifier en 2025 les missions et les processus pour lesquels l'IA est susceptible d'apporter des gains d'efficacité et de productivité significatifs.

CONCLUSION INTERMÉDIAIRE

La mise en œuvre de l'IA permet la plupart du temps de générer des gains de productivité, et, dans certains cas, de libérer des emplois. L'administration doit décider de leur affectation : les emplois peuvent être maintenus sur la même mission, redéployés sur d'autres fonctions, ou restitués dans une optique de diminution des dépenses. Ces arbitrages peuvent ainsi contribuer à l'extension du champ d'intervention de l'administration, à l'amélioration de la qualité du service et à la maîtrise des dépenses publiques.

Sur les cinq projets sur 35 systèmes d'IA développés ou étudiés par le MEFSIN, la mobilisation de ces outils selon une stratégie de diminution des dépenses a permis de réaliser de l'ordre de 20 M€ d'économies annuelles à compter de 2022, confirmant le potentiel de ces outils en termes de contribution à la maîtrise des finances publiques

Cependant, l'intégration d'un système d'IA au sein d'un processus déforme généralement la nature de l'activité en amont et en aval de la chaîne de travail. Ces nouveaux outils peuvent aussi augmenter la charge de travail des services du fait de leurs performances, en détectant par exemple d'avantage d'irrégularités à traiter, ou du fait de leurs limites, en

⁵⁵ Le secrétariat général des ministères économiques et financiers (SG des MEF) agit de manière transversale en exerçant tout à la fois des missions de pilotage du ministère de l'économie, des finances, de la souveraineté industrielle et numérique dans le domaine de la modernisation, des fonctions support et des missions de gestion des services centraux (ressources humaines, logistique, immobilier, informatique, etc.). Il assure par ailleurs le soutien du ministre de la Transformation et de la Fonction publiques qui en dispose. Cf. Arrêté du 18 décembre 2019 portant organisation du secrétariat général des ministères économiques et financiers.

généralisant par exemple des réponses erronées qu'il revient aux humains de détecter. Les systèmes d'IA présentent ainsi de forts enjeux pour l'organisation du travail et l'évolution des métiers nécessaires à la conduite des politiques publiques du MEFSIN.

Or ces enjeux ne sont pas suffisamment documentés à l'échelle des projets et sont quasiment pas traités par les instances de dialogue social à l'échelle ministérielle. L'évaluation des conséquences de l'IA sur ses métiers a commencé tardivement et demeure très limitée.

La Cour recommande donc au MEFSIN d'établir une carte des processus pour lesquels le recours à l'IA pourrait apporter des gains significatifs durant les prochaines.

À l'échelle de chaque projet, la Cour recommande au MEFSIN de documenter les implications en termes de productivité, d'emplois, de formation et de conditions de travail dès la phase d'études préalables, par une preuve de concept relative aux ressources humaines. Pour les plus grands projets, ces données enrichiraient les études MAREVA.

3 UN DÉVELOPPEMENT DES SYSTÈMES D'IA AU MEFSIN QUI NE TRAITE QU'IMPARFAITEMENT LES ENJEUX DE CONFIANCE ET LES DÉFIS ENVIRONNEMENTAUX

En parallèle du déploiement des capacités de l'IA au MEFSIN et de l'allocation de ses gains de productivité, de nombreuses alertes ont été émises quant à l'IA. Elles sont émises par des organisations internationales. La notion d'IA de confiance s'est imposée à cet égard, renvoyant à la fois à l'ensemble des questionnements concernant les sujets éthiques associés à l'IA et aux conditions nécessaires au développement d'une IA répondant aux inquiétudes et aux dangers du déploiement des systèmes d'IA.

Ces craintes sont spécifiques à l'IA. L'absence de traitement de ces questionnements porte un risque important d'amointrissement de la performance des systèmes d'IA.

La grille de maturité d'IA de confiance développée par la Cour montre une disparité des actions pour sécuriser les systèmes d'IA du point de vue de la confiance au sein du MEFSIN. Dans le contexte de *l'AI Act*, la Cour recommande de consolider ces approches au niveau ministériel et de piloter la mise en œuvre des systèmes d'IA par les risques.

L'AI Act

L'AI Act est le règlement de l'Union européenne sur l'intelligence artificielle pour renforcer la confiance dans cette technologie. Il vise à encadrer le développement et l'utilisation de l'IA dans les États membres, en protégeant les droits fondamentaux des citoyens. Il a été voté par le Parlement européen le 13 mars 2024.

Il interdit certains systèmes d'IA qui portent atteinte aux valeurs et aux droits de l'UE, comme la notation sociale, la reconnaissance faciale à des fins répressives, ou la manipulation subliminale.

Il impose des exigences strictes aux systèmes d'IA à haut risque, comme les machines médicales, les voitures autonomes, ou les systèmes de recrutement. Ces systèmes doivent être transparents, fiables, et respecter la protection des données.

Il crée un système de gouvernance et de supervision de l'IA, avec des autorités nationales compétentes, un comité européen de l'IA, et des codes de conduite volontaires.

Il prévoit des sanctions dissuasives en cas de non-respect des règles, pouvant aller jusqu'à 30 millions d'euros ou 4% du chiffre d'affaires.

Ce diagnostic se reproduit lorsque l'on étudie la prise en compte par les systèmes d'IA de leur impact sur le changement climatique.

3.1 Une prise en compte peu structurée des considérations de confiance

3.1.1 Des principes partagés mais non règlementés, un besoin d'outil évaluatif

3.1.1.1 Un appel partagé pour lutter contre les craintes vis-à-vis de l'IA, la non-explicabilité et le biais d'automatisation

Le concept d'IA de confiance a été développé par les secteurs privé⁵⁶ et public pour guider les actions de régulation des capacités et des usages de l'IA. Les recommandations et avis des organisations internationales se sont succédés depuis la fin des années 2010. En 2019, les pays membres de l'OCDE ont adopté les principes sur l'intelligence artificielle de confiance⁵⁷.

En France, en 2018, le rapport Villani « *donner un sens à l'IA, pour une stratégie nationale et européenne* » alertait sur les problèmes posés par les “boîtes noires” (algorithmes dont on ne parvient pas à expliquer le fonctionnement interne) et recommande la création d'un comité d'éthique de l'IA. En 2022, le Conseil d'État, dans son étude « *Intelligence artificielle et action publique : construire la confiance, servir la performance* », insistait sur le devoir d'exemplarité du secteur public pour répondre à la « méfiance naturelle qu'inspire le recours à des machines et à des technologies souvent mal comprises » et « favoriser un déploiement fluide des systèmes d'IA dans l'action publique »⁵⁸.

La stratégie de confiance doit répondre aux craintes des utilisateurs et aux dangers identifiés. La recherche a modélisé certains défis à relever. Les chercheurs Junesoo Lee et Jaehyuk Park ont ainsi distingué quatre types de « peurs de l'IA » identifiables parmi les personnes qui sont appelées à travailler avec des systèmes d'IA sans y avoir été préparées : l'incompréhension vis-à-vis du fonctionnement l'IA (« *je ne sais pas comment l'IA fonctionne, ni comment m'en servir* »), la crainte de la substituabilité par l'IA, le flou concernant la responsabilité de l'IA et l'obligation de mise en œuvre de l'IA⁵⁹. À chacune de ces craintes, des dangers potentiels sont déduits.

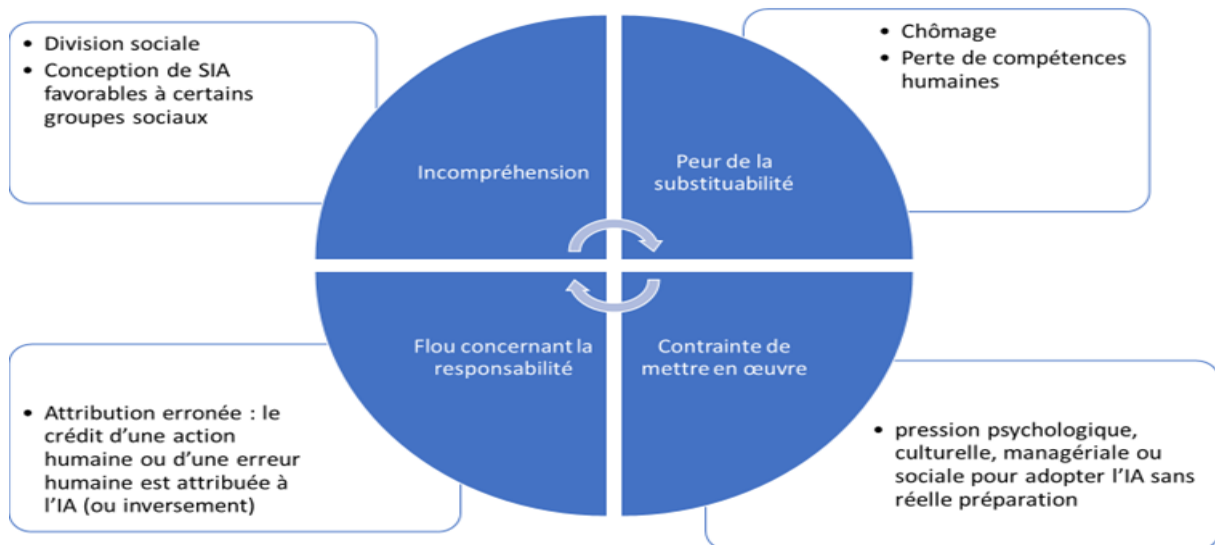
⁵⁶ Cf. « *Pause Giant AI Experiments: An Open Letter* », <https://futureoflife.org/open-letter/pause-giant-ai-experiments> (consulté le 07/07/2023)

⁵⁷ Anne Bellon, Julia Velkovska, « L'intelligence artificielle dans l'espace public : du domaine scientifique au problème public », in *Réseaux*, 2023/4 (N° 240), pages 31 à 70.

⁵⁸ Conseil d'État, étude adoptée en assemblée générale plénière du 31/03/2022

⁵⁹ Centre national de recherche de Corée du Sud, Junesoo Lee, Jaehyuk Park, « AI as Another I : Journey map of working with artificial intelligence from AI-phobia to AI-preparedness, 2023

Graphique n° 8 : Des craintes vis-à-vis de l'IA et la préparation à l'IA : les défis à relever



Source : D'après de « *AI as Another I : Journey map of working with artificial intelligence from AI-phobia to AI-preparedness*, Junesoo Lee, Jaehyuk Park, 2023, centre national de recherche de Corée du Sud ; traduction : Cour des comptes

Par ailleurs, la recherche en sciences comportementales a mis en évidence l'existence d'un biais d'automatisation, qui correspond au fait d'accorder un crédit excessif aux machines, au détriment de son propre jugement. Ce biais est d'autant plus présent que le risque d'erreur propre à un système d'IA est mal connu de ses utilisateurs. Faire confiance à une machine qui a un faible taux d'erreur pour prendre des décisions est judicieux mais peut conduire les utilisateurs à minimiser les risques d'erreur qui sont associés à ces machines. Ce biais survient aussi lorsqu'une omission de la part d'une machine conduit l'utilisateur à ne pas agir lorsqu'il devrait le faire. La personnalité, la complexité et la difficulté de la tâche à accomplir peuvent influencer la probabilité d'apparition de ce biais⁶⁰. Ce biais est plus présent pour les personnes qui débutent la réalisation d'une tâche en question, de même que chez les personnes qui sont très habituées à la machine qui les assiste.

Pour recourir à l'IA en maîtrisant ce biais, les chercheurs en sciences comportementales recommandent trois actions. En premier lieu, les acteurs de l'IA doivent être informés sur les algorithmes et les processus qui gouvernent les systèmes d'IA. En deuxième lieu, les organisations doivent augmenter la sensibilisation et la responsabilisation des professionnels qui utilisent l'IA. En troisième lieu, les concepteurs de modèles d'IA doivent choisir par défaut un *design* d'aide à la décision et non pas de décision.

⁶⁰ Cf. Site recommandé par le pôle des sciences comportementales de la DITP, [Raccourcis | Guide pratique des biais cognitifs | Shortcuts \(shortcogs.com\)](#)

Un cas d'école du biais d'automatisation

Les biais d'automatisation sont souvent mis à jour dans le domaine médical.

Des chercheurs ont identifié en 2016 une erreur dans l'algorithme d'une technologie appelée « Soutien à la décision clinique » (*Clinical Decision Support ou CDS*), permettant au personnel médical d'obtenir des suggestions de diagnostic sur la base de symptômes. L'algorithme exigeait que le personnel soignant teste la présence de plomb dans le sang de jeunes enfants alors que ce n'était pas nécessaire. Même si cette commande allait dans le sens contraire des connaissances du personnel soignant, certains d'entre eux se sont exécutés, entraînant, selon les auteurs de l'étude, des coûts de services de santé inutiles et des inquiétudes chez les parents.

Enfin, le déploiement récent et massif d'IA génératives (LLM pour *Large Language Model*), accessibles en ligne, a mis en lumière le risque dit « d'hallucination » ou d'inexactitude. Il pousse de plus en plus d'utilisateurs, éventuellement agents du MEFSIN, à envisager leur usage dans le cadre professionnel. Cela conduit à des risques en matière de la confidentialité et des secrets. Pour fonctionner, ces IA « grand public » se nourrissent de toutes les données fournies par ses utilisateurs (documents, requêtes, etc.) et les stockent dans des *clouds*, non soumis à la réglementation européenne, pour enrichir leur base d'apprentissage. De même, cela conduit aussi à des risques de sécurité : la méthode dite de « *prompt injection* » permet à n'importe quel pirate de faire produire à l'IA génératives grand public des contenus malveillants en lui faisant contourner sciemment les règles que lui ont donné ses créateurs. Cette manipulation des IA peut se faire soit de manière directe (en la convaincant de ce qu'on demande n'est pas contraire à ses règles) soit de manière indirecte en préparant des contenus malveillants qui seront ensuite envoyés aux utilisateurs qui font des recherches sur ces thèmes.

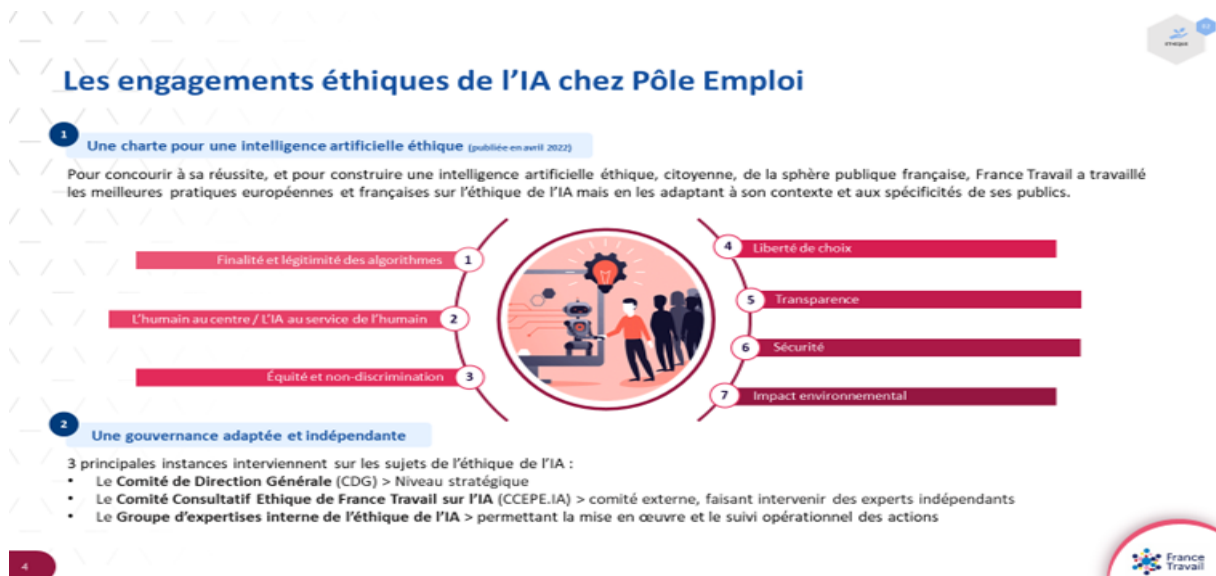
L'ensemble de ces risques et biais incite à la prudence dans la mise en œuvre des systèmes d'IA. Les résultats des systèmes d'IA, la performance du système, la bonne gestion des deniers public ne sont pas sécurisés dès lors que ces risques ne sont pas traités.

Dans ces conditions, les administrations publiques ont bénéficié d'une grande autonomie pour créer de la confiance autour de leurs systèmes d'IA. À l'exception du règlement général de protection des données, qui soutient certains principes de l'IA de confiance, aucune règle technique n'a imposé, durant la période 2018-2023, aux administrations des réglementations auxquelles elles pourraient se conformer pour respecter les principes d'IA de confiance. Un mouvement de normalisation s'est déployé *in itinere*, conduisant à proposer des outils de mesure sur certains aspects de la confiance, sans pour autant proposer, à ce stade, des règles techniques à suivre pour se conformer aux principes de l'IA de confiance. Il s'agit par exemple de la norme ISO/IEC 22989:2022 Concepts et terminologie relatifs à l'intelligence artificielle, de la norme ISO/IEC 23894:2023 Recommandations relatives au management du risque IA et de la norme ISO/IEC 42001 Intelligence artificielle — Systèmes de management. Ces normes ne sont pas imposables aux administrations. Parallèlement, divers labels ont été créés dans le secteur privé sans que la logique soit reprise par les administrations.

À titre d'exemple, le déploiement depuis 2017 de l'IA à Pôle Emploi, aujourd'hui France Travail, témoigne d'une stratégie claire et de choix explicites sur les conditions préalables au recours à ces technologies. Lors de l'appel à projets réalisé par le Fonds pour la Transformation de l'Action Publique (FTAP), France Travail a proposé dès 2018 le projet « Intelligence Emploi », voué à installer durablement l'IA au sein de Pôle Emploi. Entre 2019 et 2022, différents cas d'usage ont été développés autour de trois figures, le collaborateur, le demandeur

d'emploi et l'entreprise. Chaque système d'IA a été développé en respectant une méthode commune nécessitant de formaliser le besoin avec les conseillers ou les usagers, d'étudier les biais et les risques, d'évaluer la pertinence scientifique de la solution envisagée dans le cadre d'un comité éthique, d'expérimenter avant de déployer. Selon France Travail, cette méthode a permis de traiter au préalable les craintes des usagers et des agents.

Graphique n° 9 : Charte et gouvernance de l'IA au sein de France Travail



Source : France Travail

Une gouvernance a été établie. Une charte pour une intelligence artificielle éthique a été publiée en avril 2022, en se fondant sur les pratiques européennes et françaises.

Le programme Intelligence Emploi s'est achevé en 2022. France Travail a mené des travaux afin de capitaliser sur les expériences acquises et pérenniser le développement de l'intelligence artificielle au sein de France Travail. En parallèle, un audit de la stratégie d'IA a été réalisée par un cabinet indépendant.

3.1.1.2 Maturité des systèmes d'IA en termes de confiance au moyen d'une grille d'analyse élaborée par la Cour

En l'absence de réglementation sur le sujet et faute d'instance où ces sujets auraient été discutés formellement durant la période d'instruction, il est difficile pour les acteurs d'évaluer les effets de leurs actions et la maturité de la stratégie, chacun d'entre eux étant libre de choisir de mettre en œuvre l'un ou l'autre des principes de confiance, selon des formes arbitrées à leur niveau.

La Cour a choisi d'évaluer la maturité des systèmes d'IA du MEFSIN en termes de confiance en étudiant les modalités choisies et le rythme suivi par les administrations pour construire la stratégie adaptée.

A cette fin, une grille de maturité de l'IA de confiance a été créée à l'occasion de cette enquête de la Cour. Le postulat pour l'établir est que la stratégie d'IA de confiance est d'autant

plus robuste que les principes de confiance⁶¹ sont connus au sein de l'organisation, qu'ils font l'objet d'un pilotage par les risques et qu'ils sont testés par des tiers, de façon régulière ou récurrente. La grille démontre le caractère dynamique, évolutif, de la démarche d'IA de confiance. La maturation de l'IA de confiance s'appréhende comme un mouvement en cinq phases, même si celles-ci ne sont pas forcément séquentielles et peuvent être menées conjointement. Ces phases sont identifiées par cinq niveaux dans la grille de maturité de l'IA. La grille de maturité permet de traiter les différents aspects de l'IA de confiance en analysant la maturité des sept principes d'IA de confiance.

Les grilles de maturité

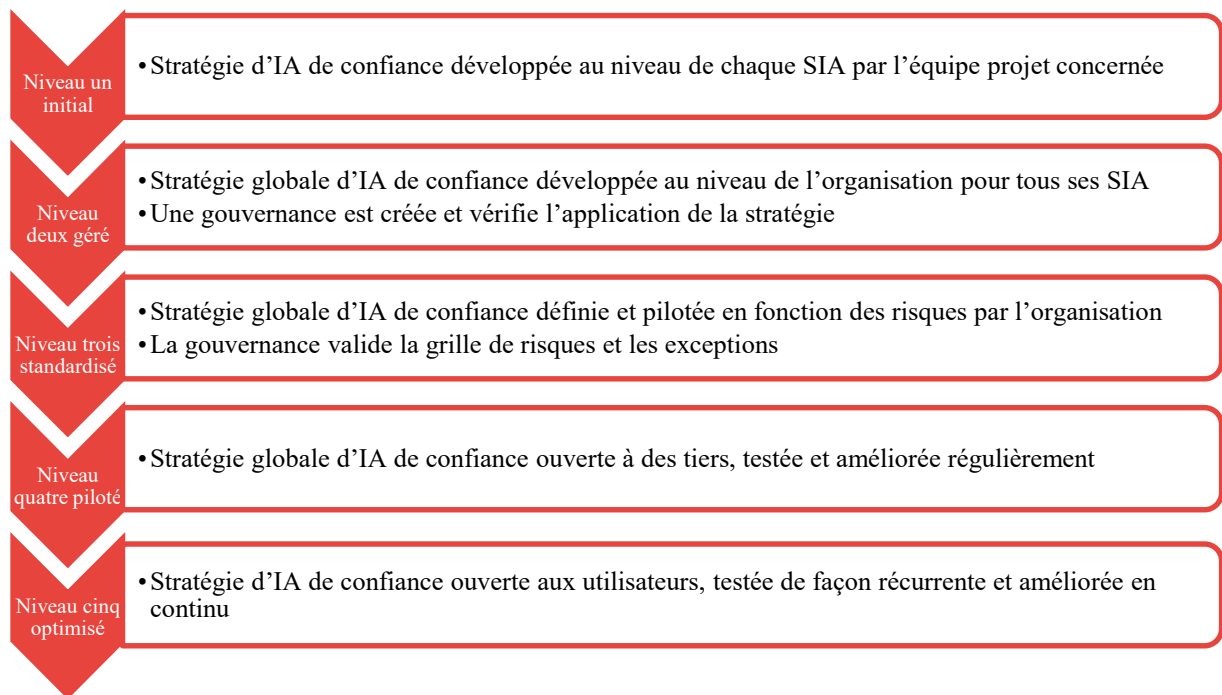
La maturité évalue la capacité de l'organisation à effectuer sa mission de façon efficace, selon des critères définis préalablement. Les grilles de maturité des processus constituent un référentiel d'analyse et de progression permettant au pilote ou à l'évaluateur de situer le processus étudié sur une échelle d'un à cinq, de le comparer et d'engager un plan de progrès adapté. Elles supposent une évolution étape par étape.

La Cour des comptes a déjà procédé à la création de grilles de maturité. À titre d'exemple, cette méthode a été choisie pour évaluer la conduite des grands projets numériques de l'État, qui a conduit à dégager des principes de bonne gestion. La grille de maturité qui en est déduite a été utilisée ultérieurement pour analyser la réforme du prélèvement à la source⁶².

⁶¹ Les principes d'IA de confiance sont détaillés en annexe.

⁶² Cf. Cour des comptes, la conduite des grands projets numériques de l'État, communication à la commission des finances du Sénat, juillet 2020 et Cour des comptes, la mise en œuvre du prélèvement à la source, janvier 2022.

Graphique n° 10 : Grille de maturité de l'IA de confiance



Source : Cour des comptes

Selon cette approche, le « niveau 1 » de maturité en termes d'IA de confiance est atteint dès lors que l'équipe porteuse du système d'IA est consciente des enjeux, définit une stratégie, et la respecte tout en étant capable de l'expliquer. La maturité reste toutefois limitée au système d'IA en question. Elle est dépendante des personnes qui composent l'équipe et est vulnérable au *turn-over* des agents. Le résultat dépend de l'effort d'individus plus que de celui de l'organisation publique. Il n'y a pas de maturité organisationnelle autour de l'IA de confiance. Ce niveau 1 est appelé « niveau de maturité d'IA de confiance initial ».

Le « niveau 2 » de maturité en termes d'IA de confiance est atteint dès lors que l'organisation publique porteuse du système d'IA a commencé à élaborer une stratégie globale d'IA de confiance, laquelle prend éventuellement la forme d'une charte. Des exceptions ne sont pas prévues. L'organisation prévoit de créer ou a créé une gouvernance pour vérifier l'application de cette stratégie (comité stratégique, comité consultatif éthique, groupe d'expertises internes de l'éthique, etc.). L'organisation prévoit de s'engager ou s'engage sur des indicateurs de conformité. Le niveau 2 est appelé « niveau de maturité d'IA de confiance géré ».

Le « niveau 3 » est atteint dès lors que l'organisation publique porteuse du système d'IA a affiné sa stratégie globale d'IA de confiance en la pilotant par les risques. Des exceptions à la stratégie globale d'IA de confiance sont autorisés dès lors que l'organisation estime et justifie d'un risque maîtrisé. Elle garantit être en mesure d'explicitier ses choix. Le niveau 3 est appelé « niveau de maturité d'IA de confiance standardisé ».

Le « niveau 4 » de maturité, appelé « niveau de maturité d'IA de confiance piloté », est atteint dès lors que l'organisation publique porteuse du système d'IA a créé les conditions d'une évaluation régulière par des experts, par audit externe et par comité scientifique. À ce niveau, les systèmes d'IA sont pilotés par les risques et au moyen d'indicateurs. Les éventuelles dérives sont identifiées et corrigées systématiquement.

Le « niveau 5 » est atteint lorsque l'organisation publique porteuse du système d'IA a créé les conditions d'une évaluation récurrente de sa stratégie par des tiers et d'une ouverture complète de celles-ci aux utilisateurs. L'organisation est dans une boucle permanente d'amélioration qui permet de gérer la performance organisationnelle et d'analyser les causes de dysfonctionnement et de les résoudre. Le niveau 5 est appelé « niveau de maturité d'IA de confiance optimisé ».

Le « niveau 0 » correspond à la situation selon laquelle l'équipe porteuse du projet n'est pas en mesure de présenter les choix effectués pour respecter les principes d'IA de confiance. Le système d'IA est immature en termes d'IA de confiance.

La maturité globale est calculée selon le niveau de maturité déclaré par le porteur de système d'IA pour l'application d'une stratégie en faveur de chacun des principes de confiance, à savoir de la transparence, de la responsabilisation, du contrôle humain, de la gouvernance des données, de l'équité et non-discrimination, du bien-être sociétal et environnemental et de la sécurité et de la robustesse techniques⁶³.

Pour chaque principe, un degré de maturité peut être établi selon certains jalons. En l'état de l'art et sous réserve du contexte propre à chaque SIA, cela peut être les jalons ci-dessous :

Pour le principe de transparence, en l'état de l'art et sous réserve du contexte propre à chaque SIA, il s'agit par exemple des jalons suivants, même si ceux-ci ne sont pas les seuls. Le niveau 1 est atteint à partir du moment où la finalité est définie ; le niveau 2 est abordé quand une charte de l'IA est définie. Le système d'IA passe au niveau 3 de maturité quand le cahier des charges et le code (ou équivalents) sont publiés ; au niveau 4 quand l'organisation a mis le système d'IA en capacité d'être challengé par des tiers et au niveau 5 si le code est publié en open source.

Pour le principe de responsabilité et d'explicabilité, en l'état de l'art et sous réserve du contexte propre à chaque SIA, il s'agit par exemple des jalons suivants, même si ceux-ci ne sont pas les seuls. Le niveau 1 est atteint à partir du moment où la conformité au cadre juridique est vérifiée ; le niveau 2 est abordé quand un comité d'IA est créé. Le système d'IA passe au niveau 3 de maturité quand les responsabilités sont contrôlées ; au niveau 4 quand l'organisation a demandé à des tiers de vérifier la réalité des besoins et de la conformité au cahier des charges et au niveau 5 si les audits sont systématiques.

Pour le principe de primauté humaine, en l'état de l'art et sous réserve du contexte propre à chaque SIA, il s'agit par exemple des jalons suivants, même si ceux-ci ne sont pas les seuls. Le niveau 1 est atteint à partir du moment où l'explicabilité aux acteurs est assurée ; le niveau 2 est abordé quand un programme de formation systématique des acteurs de l'IA est établi. Le système d'IA passe au niveau 3 de maturité quand l'autonomisation du système d'IA est autorisée seulement en cas de risque maîtrisé ; au niveau 4 quand l'organisation a demandé à des tiers de contrôler les responsabilités et au niveau 5 si des audits récurrents de la fiabilité du résultat sont confiés à un tiers.

Pour le principe de respect des données, en l'état de l'art et sous réserve du contexte propre à chaque SIA, il s'agit par exemple des jalons suivants, même si ceux-ci ne sont pas les seuls. Le niveau 1 est atteint à partir du moment où la gestion interne des données est conforme

⁶³ Les principes d'IA de confiance sont détaillés en annexe.

au cadre normatif ; le niveau 2 est abordé quand Comité de gouvernance des données existe. Le système d'IA passe au niveau 3 de maturité quand les dérogations sont maîtrisées par une grille risque ; au niveau 4 quand les systèmes d'IA sensibles au sens de l'IA Act sont audités par des tiers et au niveau 5 en cas d'audits récurrents de la gouvernance des données

Pour le principe d'équité et non-discrimination, en l'état de l'art et sous réserve du contexte propre à chaque SIA, il s'agit par exemple des jalons suivants, même si ceux-ci ne sont pas les seuls. Le niveau 1 est atteint à partir du moment où une stratégie anti-biais est déployée par l'équipe projets ; le niveau 2 est abordé quand un comité d'IA vérifie l'existence de biais et leur traitement. Le système d'IA passe au niveau 3 de maturité quand un comité éthique et scientifique valide le système d'IA et au niveau 5 si les analyses du comité d'éthique sont accessibles aux parties prenantes de l'IA.

Pour le principe de bien-être sociétal et environnemental, en l'état de l'art et sous réserve du contexte propre à chaque SIA, il s'agit par exemple des jalons suivants, même si ceux-ci ne sont pas les seuls. Le niveau 1 est atteint à partir du moment où la proportionnalité des techniques à la finalité est vérifiée ; le niveau 2 est abordé quand une éco-responsabilité du système d'IA est définie. Le système d'IA passe au niveau 3 de maturité quand les risques sont homologués ; au niveau 4 quand l'organisation vérifie la conformité du système d'IA aux engagement d'éco-responsabilité et au niveau 5 si les contrôles sont effectués de façon récurrente et par des tiers.

Pour le principe de sécurité et robustesse techniques, en l'état de l'art et sous réserve du contexte propre à chaque SIA, il s'agit par exemple des jalons suivants, même si ceux-ci ne sont pas les seuls. Le niveau 1 est atteint à partir du moment où l'accès aux données est restreint ; le niveau 2 est abordé quand le système d'IA respecte *SecNunCloud* ou est établi sur des serveurs sécurisés. Le système d'IA passe au niveau 3 de maturité quand les risques sont homologués ; au niveau 4 quand l'organisation a demandé à des tiers de vérifier la conformité aux règles de sécurité et au niveau 5 si des alertes d'intrusion existent.

Tableau n° 2 : Exemples de jalons de maturité

Principes	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Transparence	Finalité définie	Charte de l'IA	Cahier des charges et code publiés	Challenge par des tiers	code publié en open source
Responsabilité et explicabilité	conformité au cadre juridique vérifié	comité d'IA	responsabilités contrôles	audit de la réalité des besoins et de la conformité au cahier des charges	audits systématiques
Contrôle humain	test à grande échelle avant mise en production, explicabilité assurée	formation des acteurs à l'IA	autonomisation du SIA autorités en cas de risque maîtrisé	contrôle du résultat et des responsabilités	contrôle récurrent de la fiabilité du résultat et des responsabilités confié à un tiers
Respect des données	gestion interne des données conforme au cadre normatif	comité de gouvernance des données	dérogation au cadre normatif maîtrisé par une grille de risque	SIA "sensibles" audités	audit récurrent de la gouvernance des données des SIA
Equité et non-discrimination	Stratégie anti-biais déployée par l'équipe projet	gouvernance vérifie l'existence de biais	comité éthique et scientifique valide le SIA	Contrôles réguliers sur les biais	analyses du comité éthique accessibles
Sécurité et robustesse techniques	accès aux données restreint	respect SecNumCloud ou serveurs sécurisés	Risques homologués	conformité aux règles et alertes	mécanismes d'alerte d'intrusion
Bien-être sociétal et environnemental	techniques proportionnées à la finalité	écoresponsabilité définie	risques homologués et coûts identifiés	contrôle de la conformité aux engagements d'écoresponsabilité	contrôles récurrents par des tiers

Source : Cour des comptes

3.1.2 Un niveau de maturité non standardisé, non piloté, non optimisé

3.1.2.1 Un niveau de maturité ministériel estimé à 2 sur 5, soutenu par des pratiques fortes en matière de gouvernance des données

Pour évaluer la maturité des systèmes d'IA du MEFSIN et les comparer entre eux, un questionnaire a été transmis à la majorité des porteurs de projets. La Cour a reçu 21 réponses

durant l'instruction. La Cour a évalué ces réponses selon une méthode présentée en annexe⁶⁴ du présent rapport. La notation obtenue par chaque système d'IA figure également en annexe du présent rapport.

En application de ce canevas, à l'échelle ministérielle, la maturité globale moyenne en termes d'IA de confiance se situe au niveau 2, c'est-à-dire qu'elle commence à être gérée au-delà des seules équipes projet. La stratégie de confiance dans les systèmes d'IA n'est toutefois pas standardisée au sein des directions et encore moins à l'échelle ministérielle. Elle n'est pas gérée par les risques. L'ouverture à des tiers de la stratégie d'IA de confiance reste rare et les tests sont rarement établis selon une programmation récurrente.

Tableau n° 3 : Niveau de maturité du MEFSIN en termes d'IA de confiance établi par la Cour

Principes de confiance	Moyenne au MEFSIN
Transparence	2,1
Responsabilisation	2,3
Facteur et contrôle humain	2,7
Respect de la vie privée et gouvernance des données	3,2
Équité et non-discrimination	2,0
Sécurité et robustesse technique	2,0
Bien être sociétal et environnemental	2,3

Source : Cour des comptes. La méthode employée est développée supra et en annexe. Note de lecture : l'Insee atteint le niveau 3 pour le principe de transparence dans l'emploi de l'IA, c'est-à-dire qu'il a créé des standards en la matière.

Le principe d'IA de confiance le plus observé et mature au sein du MEFSIN est celui qui concerne le respect de la vie privée et la gouvernance des données. Cette situation s'explique par le fait qu'il reflète la conformité au règlement général de protection des données⁶⁵ entré en application en 2018 et qu'il est surveillé par un réseau formalisé, celui des responsables de la sécurité des systèmes d'information (RSSI). Ce principe est également en adéquation avec les différentes feuilles de route ministérielles de la donnée.

En revanche, le principe d'équité et de non-discrimination est le moins poursuivi, avec celui concernant la sécurité et la robustesse technique, en dépit des inquiétudes fortes sur ce point. Cette situation est assez homogène au sein des directions du MEFSIN.

En termes de transparence, la DGFIP se démarque mais un comité d'éthique est en cours de constitution au sein de sa délégation à la transformation numérique, qui pourrait traiter ce sujet.

Il est plausible que la nouveauté de la préoccupation du biais algorithmique explique cela. Les systèmes d'IA pâtissent de l'absence de référentiel défini sur l'équité et la lutte contre

⁶⁴ Cf. Annexe 3

⁶⁵ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

le biais algorithmique et n'ont pas eu l'opportunité d'être conseillé par un réseau de responsables chargés d'assurer la conformité de ce référentiel.

Le fait que les systèmes d'IA génératifs sont les moins matures en termes de confiance pourrait conforter l'analyse selon laquelle les technologies nouvelles pâtissent du manque d'outillage et de référentiels.

3.1.2.2 La maturité des systèmes d'IA n'est pas corrélée au stade de développement mais semble plus liée à l'effet « organisation »

Il n'y a pas de corrélation entre la maturité des systèmes d'IA en termes de confiance et le stade de développement. Autrement dit, ce n'est pas parce qu'un système d'IA est plus récent, qu'il est en cours de conception ou au stade *POC*, qu'il prend plus en compte les enjeux de confiance que les autres systèmes d'IA. De même, ce n'est pas parce qu'un système d'IA est développé qu'il est plus mature en termes d'IA de confiance par rapport aux autres systèmes d'IA du MEFSIN.

La prise de conscience des enjeux de confiance dans les systèmes d'IA ne se fait pas au cours du cycle de vie du système d'IA.

En revanche, un effet-levier paraît se jouer au niveau de l'organisation. Cet effet est à nuancer dans la mesure où le nombre de systèmes d'IA diffère entre les directions du MEFSIN. Pour parvenir à ces résultats, la moyenne des résultats de huit systèmes d'IA portés par la DGFIP a été établie. En revanche, la DGE et la DGCCRF ne portent respectivement qu'un seul système d'IA évalué en termes de confiance.

Tableau n° 4 : Niveau de maturité du MEFSIN par direction en termes d'IA de confiance établi par la Cour

Principes de confiance	DGDDI	DGFIP	DGE	Insee	SG	AIFE	DGCCRF
Transparence	2,3	1,6	4,5	3,0	2,3	3,0	0,0
Responsabilisation	3,7	1,6	4,0	3,3	1,0	3,5	2,0
Facteur et contrôle humain	2,4	3,4	4,0	3,0	2,0	1,3	0,7
Respect de la vie privée et gouvernance des données	3,5	2,6	4,5	3,5	4,0	4,0	1,0
Équité et non-discrimination	1,2	2,3	3,5	2,5	2,0	1,5	0,0
Sécurité et robustesse technique	2,4	2,5	3,2	2,0	0,8	1,4	0,8
Bien-être sociétal et environnemental	3,8	2,5	3,5	1,5	1,5	0,0	1,0
Moyenne	2,8	2,4	3,9	2,7	1,9	2,1	0,8

Source : Cour des comptes. La méthode employée est développée supra et en annexe. Note de lecture : l'Insee atteint le niveau trois pour le principe de transparence dans l'emploi de l'IA, c'est-à-dire qu'il a créé des standards en la matière.

De même, le niveau atteint peut indiquer un choix délibéré. Ainsi, le niveau des systèmes d'IA de la DGFIP en termes de transparence pourrait s'expliquer par une volonté de préserver la discrétion sur les méthodes de contrôle et de protection du secret fiscal. Néanmoins, le niveau de maturité de la DGFIP pourrait être amélioré en explicitant les raisons d'une moindre transparence.

Le niveau de maturité de la DGCCRF établi ci-dessous ne tient pas compte du système d'IA Réponse Conso qui a été mis en place opérationnellement. Selon la DGCCRF, l'analyse de Réponse Conso en termes d'IA de confiance devrait lui permettre de présenter un niveau de maturité supérieur.

3.1.2.3 « Signaux Faibles », un système d'IA au niveau de confiance 4 sur 5

Les résultats de la DGE sont portés par la maturité en termes de confiance du système d'IA nommé « *Signaux faibles* » qui se distingue par sa stratégie formalisée et sa gouvernance spécifique⁶⁶

Signaux faibles bénéficie en effet d'une convention-cadre qui, depuis avril 2019, définit la finalité et la gouvernance du programme, et d'un comité scientifique et technique.

⁶⁶ Ce jugement ne porte pas sur l'efficacité du dispositif en tant que tel, qui fait d'une évaluation dans le cadre d'un contrôle en cours de la Cour des Comptes sur la détection et le traitement des difficultés des TPE et PME.

De plus, *Signaux faibles* présente un très beau niveau d'explicabilité. Sa convention rend compte des partages de données, des principes généraux de sécurité et des droits d'accès (restrictions) aux données et résultats. Les résultats du système d'IA sont expliqués via un radar qui met en lumière les variables les plus influentes de la détection, assorti de phrases explicatives complémentaires et de fiches mettant en valeur les multiples signaux et data exploitées. Plusieurs mises en valeur graphiques ont été testées pour rendre intelligibles les fondamentaux de la détection.

La finalité est présentée sur une page publique, précisée dans un guide d'utilisateurs et les travaux de l'ensemble du système d'IA sont publiquement documentés sur un espace de dépôt GitHub public⁶⁷. Le code du système d'IA est expliqué à tous les niveaux (du code source jusqu'au résultat métier) avec un travail spécifique de vulgarisation auprès des utilisateurs.

Le guide d'usage donne des conseils pour bien interpréter la détection proposée par le SIA. Des formations sont dispensées régulièrement aux utilisateurs pour la compréhension du système d'IA et ses évolutions sont discutées avec les partenaires au sein d'un comité scientifique.

Par ailleurs, les équipes de *Signaux faibles* audient leur système via une exploration des besoins des utilisateurs et du métier en continu. Un comité scientifique et technique a été créé pour corriger des biais, s'adapter aux besoins des utilisateurs et discuter des impacts organisationnels. Des clubs utilisateurs sont organisés régulièrement pour évaluer la satisfaction et évoquer les impacts du système d'IA dans leur métier.

Le système d'IA est hébergé sur une infrastructure ayant obtenu le label *SecNumCloud* délivré par l'ANSSI. Les données fiscales et le modèle d'IA sont exploités sur le lac de données de la DGFIP.

Signaux Faibles a fait l'objet d'une homologation au regard du Référentiel général de sécurité (RGS), précédée de tests de pénétration, d'une analyse de risques et d'un plan de remédiation. En matière de conformité au RGPD, les traitements de données à caractère personnel font l'objet d'une fiche de traitement, élaborée en lien avec le service du Délégué à la protection des données ministériel et versée au registre de traitement ministériel.

Les responsabilités des hébergeurs sont prévues dans des conventions de service et les responsabilités des fournisseurs de données sont établies dans la convention de partenariat et dans des conventions de partage de données spécifiques si cela a été jugées nécessaires. Les responsabilités sont donc clairement établies et contrôlées.

3.1.3 Une stratégie d'IA de confiance à consolider pour assurer la sécurité juridique des systèmes d'IA

Pour atteindre un niveau de maturité supérieur (piloter la stratégie d'IA de confiance par les risques, ouvrir cette stratégie aux contrôles par des tiers, de façon régulière ou récurrente), la Cour recommande au MEFSIN de piloter la stratégie d'IA de confiance au niveau ministériel. La formalisation d'une charte de l'IA au sein du MEFSIN en se fondant sur les pratiques

⁶⁷ Page publique : <https://beta.gouv.fr/startups/signaux-faibles.html> . Guide d'utilisation : <https://signaux-faibles.gitbook.io/guide-dutilisation-et-f.a.q.-de-signaux-faibles>

européennes et françaises sur l'éthique de l'IA et en les adaptant au contexte du MEFSIN et aux spécificités de ses publics, serait un marqueur fort de ce mouvement.

3.2 Une intégration affichée des enjeux environnementaux mais une démarche non coordonnée

La majorité des administrations de Bercy déclarent se soucier de l'impact environnemental de leurs systèmes d'IA mais elles ne suivent aucune méthode commune d'évaluation et ne partagent pas entre elles les études faites sur le sujet ou les arbitrages réalisés.

L'impact environnemental des systèmes d'IA va s'accroître mécaniquement avec l'intensification du recours à la technologie de traitement du langage naturel. Au regard des investissements déjà consentis pour créer des systèmes d'IA, de la courbe de croissance de ces systèmes d'IA, et de la réglementation relative au numérique responsable, le MEFSIN a intérêt à intégrer la démarche de normalisation entreprise par l'AFNOR et pilotée par le ministère de l'écologie en 2024.

3.2.1 L'impact environnemental croissant de l'IA

La plupart des modèles de systèmes d'IA doivent être entraînés sur de nombreux processeurs graphiques (ou *graphics processing unit*, GPU). Ces GPU réalisent des calculs rapides mais avec une consommation d'énergie et une production de CO₂. En 2019, la revue du MIT a relayé les résultats d'une étude faite sur les coûts d'un modèle d'IA. La phase d'entraînement d'un seul modèle d'IA peut émettre plus de 313 tonnes d'équivalent CO₂⁶⁸. À titre d'exemple, l'entraînement de GPT-3 a nécessité 190 000 kWh, équivalant à un aller-retour terre-lune en avion⁶⁹.

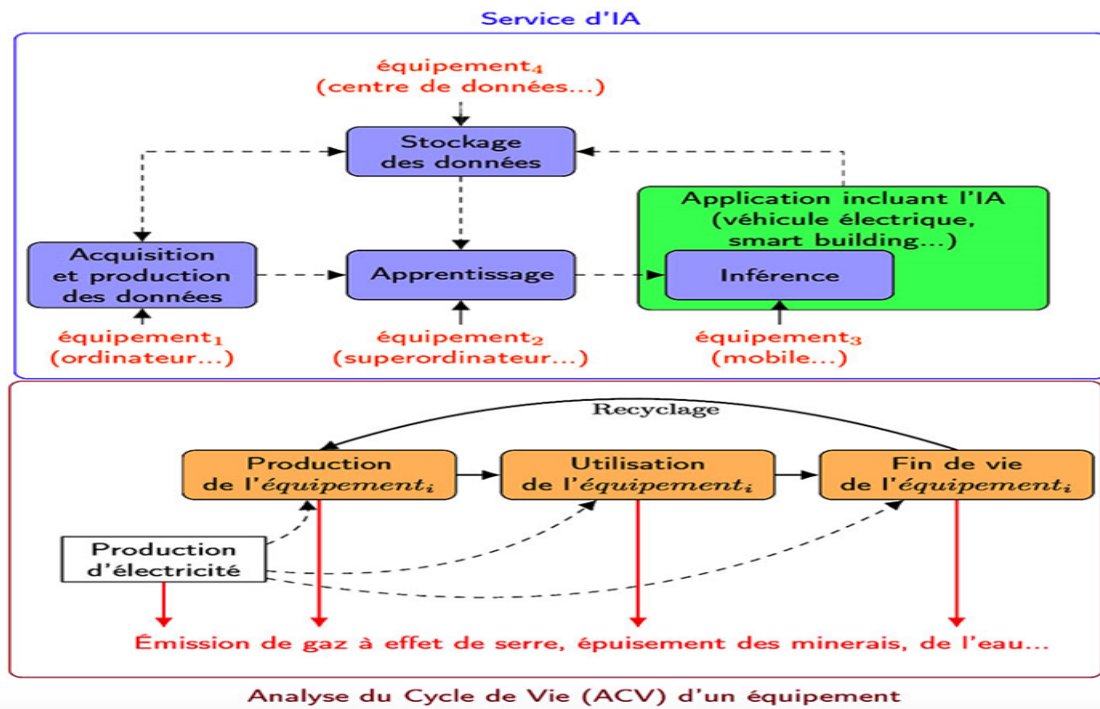
L'avènement de l'IA générative modifie la répartition de la consommation d'un modèle d'IA. Il est communément admis que la phase d'entraînement d'un système d'IA non génératif représente la majorité de la consommation du système. Or, une IA générative peut conduire à une consommation de la phase d'utilisation supérieure à celle de la phase d'entraînement. Une étude publiée par Meta montre que pour un modèle de langage utilisé plusieurs milliards de milliards de fois par jour pendant deux ans, la phase d'apprentissage représente 35 % de l'impact environnemental tandis que la phase d'inférence⁷⁰ en représente 65 %.

⁶⁸ MIT, Karen HAO, juin 2019, [L'entraînement d'un seul modèle d'IA peut émettre autant de carbone que cinq voitures au cours de leur vie | Revue de technologie du MIT \(technologyreview.com\)](#)

⁶⁹ La phase d'entraînement d'un système d'IA ne dépend toutefois pas entièrement de la combustion d'énergies fossiles mais le calcul permet d'attirer l'attention sur la grande quantité d'énergie nécessaire pour former le système d'IA.

⁷⁰ La phase d'inférence renvoie à l'exécution d'un modèle d'IA une fois celui-ci entraîné sur un data set d'apprentissage puis testé sur un data set de validation. Elle fait par conséquent référence au déploiement du modèle, et à la mise en application à partir de données de terrain.

Graphique n° 11 : Cycle de vie d'un Système d'IA



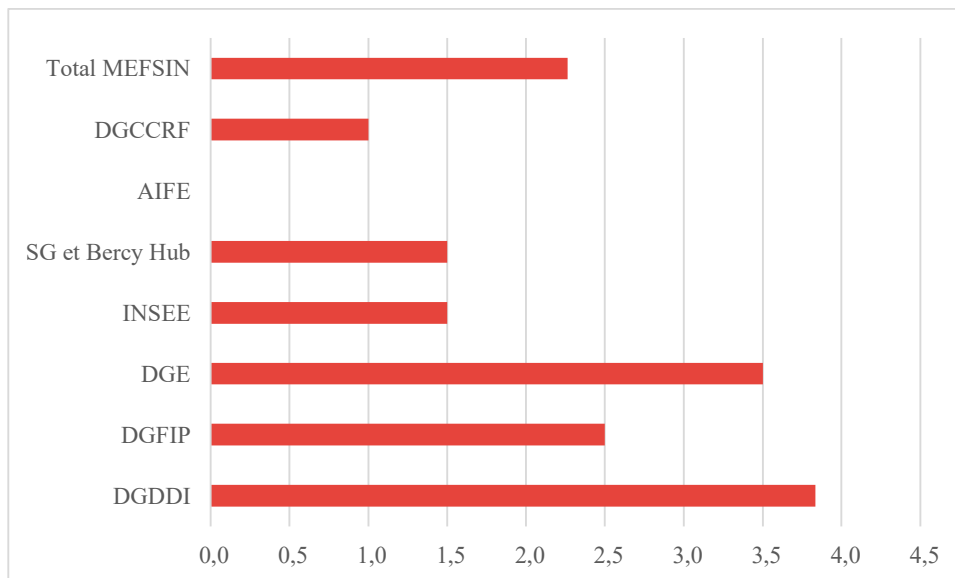
Source : *Interstices Info*⁷¹, revue de l'INRIA

3.2.2 Une prise en compte affichée de l'impact environnemental des systèmes d'IA du MEFSIN

À l'exception de ceux portés par l'AIFE, les porteurs de projet déclarent que leurs systèmes d'IA sont alignés avec les objectifs d'un État exemplaire en matière de transition écologique. Ils indiquent ainsi que leurs critères de conception et de réalisation respectent les engagements de l'État en matière de services publics écoresponsables.

⁷¹ Interstices est une revue de culture scientifique en ligne, créée par des scientifiques pour aider à explorer les sciences du numérique.

**Graphique n° 12 : maturité des systèmes d'IA du MEFSIN concernant les critères de bien-être
environnemental**



Source : Cour des comptes, d'après les réponses apportées par les porteurs de projet de 21 systèmes d'IA du MEFSIN durant l'instruction. L'AIFE n'a pas répondu.

Au-delà de cette déclaration, la difficulté est de savoir quelle méthode d'évaluation de l'impact environnemental des systèmes d'IA a été choisie et quels arbitrages ont été pris pour diminuer cet impact. L'impact environnemental de l'IA est très souvent mesuré en empreinte carbone et par rapport à la phase d'apprentissage du modèle d'IA. Cependant, la prise en compte de critères comme la consommation d'eau et de ressources naturelles, la pollution de l'eau et de l'air est nécessaire afin de mesurer correctement l'impact environnemental de l'IA.

Aussi, l'ARCEP et l'ADEME préconisent une évaluation multi-critères, multi-étapes et multi-composants des impacts environnementaux du numérique⁷². Cette approche peut être déclinée pour les systèmes d'IA.

Les impacts environnementaux du numérique ne se réduisent pas aux émissions de gaz à effet de serre. L'ARCEP et l'ADEME ont calculé que le domaine numérique sécrète 16,9 Mt CO₂ eq., soit 2,5 % de l'empreinte carbone de la France, légèrement supérieure au secteur des déchets (14 MT CO₂ eq. 1). Mais, selon cette étude, d'autres indicateurs présentent des impacts importants, en particulier l'utilisation de ressources abiotiques (minérales et fossiles) ainsi que les radiations ionisantes, même après normalisation et pondération. L'ARCEP et l'ADEME recommandent que ces indicateurs soient pris en compte en priorité lors de stratégies de réduction des impacts environnementaux, afin d'éviter les transferts de pollution.

⁷² Mandatées en août 2020 par le Gouvernement, l'ADEME et l'Arcep ont mené une étude pour préciser l'impact environnemental du numérique en France. D'après les rapports publiés le 19 janvier 2022, les terminaux (et en particulier les écrans et téléviseurs) génèrent l'essentiel des impacts environnementaux (de 65 à 92 %), suivi des centres de données (de 4 à 20%) puis des réseaux (de 4 à 13 %). Le troisième volet, publié le 6 mars 2023, fait une évaluation prospective de l'impact du numérique en France, à horizon 2030 et 2050 .

La dimension multi-étapes permet d'intégrer les impacts générés lors de toutes les étapes du cycle de vie des équipements⁷³ : phase d'extraction des matières premières, qui englobe notamment tous les processus industriels impliqués dans la transformation du minerai en métaux ; phase de fabrication, qui comprend les processus permettant de produire l'équipement à partir de la matière première ; phase de transport, qui comprend tous les processus de transport impliqués, y compris la distribution des produits ; phase d'utilisation, qui comprend principalement la consommation d'énergie de l'équipement pendant son entraînement et son inférence ; phase de fin de vie, qui fait référence aux processus de démantèlement, de recyclage et/ou d'élimination de l'équipement.

La dimension multi-composants permet d'appréhender ces systèmes complexes que sont l'association des terminaux utilisateurs, datacenters et réseaux de télécommunications, tous composés d'une multitude d'équipements ayant chacun des cycles de vie propre. Les évaluations ne doivent pas conduire à des transferts de pollution. Certains risques (maîtrisés ou non) peuvent avoir un effet en cascade, aggravant ou diminuant l'impact environnemental global du système.

La circulaire de la Première ministre du 21 novembre 2023 en faveur d'une transition écologique juste et efficace, vise à réduire les émissions de gaz à effet de serre de l'État de 22 % en 2027 par rapport à 2022 (-5 % par an⁷⁴). Les ministères sont chargés du déploiement du dispositif auprès de leurs administrations centrales et déconcentrées ainsi que des établissements publics et opérateurs sous leur tutelle.

Parmi les quinze engagements figurant dans cette circulaire, la mesure n°8 prévoit de développer le numérique écoresponsable, en augmentant la durée d'usage des matériels les terminaux représentent 79 % de l'empreinte carbone du numérique) et en réduisant la consommation énergétique (10 % de la consommation électrique française est liée aux services numériques). L'action 8.3 prévoit spécifiquement une veille performance énergétique et carbone des centres de données et serveurs et un développement des projets de récupération de la chaleur fatale de la Dinum en 2024⁷⁵.

La mesure n°15 concerne la préservation de la ressource en eau. L'action 15.1 incite spécifiquement les ministères à réduire la consommation d'eau potable en veillant à optimiser les consommations des systèmes de refroidissement des centres de données.

⁷³ L'ACV (Analyse de cycle de vie) est une méthodologie reconnue pour l'évaluation des impacts environnementaux, avec des normes ISO (ISO 14040 et 14044) et une norme méthodologique spécifique pour les technologies de l'information et de la communication.

⁷⁴ Une première estimation du bilan d'émissions de gaz à effet de serre de l'Etat a été établie à partir des données d'achats, précisée dans la stratégie de décarbonation de l'Etat. Les émissions GES de l'Etat sont estimées à environ 10 millions de tonnes équivalent CO₂ (Mteq CO₂) en 2019. Cf. Circulaire PM du 21 novembre 2024.

⁷⁵ Pilote : Dinum Cible 2024 : notification par la DAE d'un marché interministériel d'acquisition de serveurs très performants (X86 certifiés "80 plus" niveau Titanium)

3.2.3 Une possibilité d'intervention sur les données, les modèles d'IA, les infrastructures et l'énergie utilisée tout au long du cycle de vie d'un SIA

Quatre niveaux d'intervention sont identifiés communément pour réduire l'impact environnemental des systèmes d'IA : les données, les modèles, les infrastructures et l'énergie utilisée.

3.2.3.1 Agir sur les données

La réduction des données exploitables (c'est-à-dire correctement annotées), du volume de stockage des données disponible ou du temps d'apprentissage réduit l'impact environnemental d'un système d'IA. De même, le *compressive sensing* qui consiste à reconstruire une information à partir d'un échantillonnage restreint, permet de réduire l'utilisation, la transformation ou le transfert de données. La DGDDI a limité le nombre de données utilisées pour l'entraînement de son système d'IA d'identification de colis suspects.

Cette approche est compatible avec une utilisation locale de l'IA, c'est-à-dire une IA apprenant et s'exécutant sur des infrastructures non déportées vers le *cloud*, avec des ressources limitées, tels que des appareils mobiles, des objets connectés ou des systèmes embarqués.

3.2.3.2 Agir sur les modèles

La consommation énergétique intrinsèque du modèle peut être réduite en privilégiant les modèles simples et compacts. Il faut pour cela créer de nouveaux types d'algorithmes ou optimiser les algorithmes existants, par des méthodes de compression, par exemple en utilisant des réseaux de neurones moins profonds ou moins larges. Sous ces termes se regroupent comme le *pruning*⁷⁶ et le *few-shot learning*⁷⁷.

La volonté de recycler et réutiliser des modèles existants réduit l'impact environnemental. C'est le cas du *transfer learning*, qui réutilise les modèles ayant déjà appris en les transposant sur un nouveau problème. Cette démarche vise à minimiser les efforts et l'énergie consommés lors du développement et de l'apprentissage de nouvelles IA. Elle est promue par le département SJCF 1-D en charge du système d'IA *Ciblage de la fraude et valorisation des requêtes (CFVR)*.

⁷⁶ Le *pruning*, ou élagage en français consiste à réduire la taille d'un réseau de neurones, tout en minimisant la perte de performance.

⁷⁷ Le *few-shot learning*, ou réduction du jeu de données en français, est un modèle d'apprentissage en IA qui tente de résoudre des problèmes d'apprentissage machine avec peu d'exemples.

3.2.3.3 Agir sur les infrastructures

La consommation énergétique globale des modèles d'IA est également fortement impactée par le choix des circuits, composants utilisés ou types de processeurs. Cet impact peut par exemple se caractériser par la quantité d'énergie utilisée (en kWh), ou l'efficacité énergétique de l'infrastructure utilisée (mesurée selon différents critères comme par exemple le temps d'exécution). Il est aujourd'hui possible de faire appel à des outils développés par des groupes de recherche visant à évaluer l'empreinte carbone d'un modèle⁷⁸.

Pour répondre à cette problématique, de nouveaux composants font leur apparition, comme les puces neuromorphiques qui s'inspirent du fonctionnement des neurones humains. Ces puces rapprochent les unités de calcul et la mémoire, limitant ainsi les transferts de données et donc la consommation énergétique par rapport aux puces classiques.

Les administrations du MEFSIN n'emploient pas ces nouveaux composants.

3.2.3.4 Agir sur l'énergie utilisée

La consommation énergétique des infrastructures est un facteur critique en termes d'impact environnemental de l'IA. Pour cela, le choix de la localisation des infrastructures est d'une importance majeure. À titre d'exemple, concernant l'empreinte carbone d'un centre de données (*datacenter*), peu d'opérateurs possèdent assez de ressources pour produire leur propre énergie renouvelable sur le site de leurs infrastructures. La plupart du temps, la consommation énergétique du *datacenter* dépend du mix énergétique de la région dans laquelle il est implanté. La localisation d'un *datacenter* est aussi déterminante pour le système de refroidissement et de ventilation, qui représente jusqu'à 40 % de sa consommation d'énergie. Ainsi un climat adapté et l'abondance d'eau peuvent constituer des facteurs favorables.

La DGDDI et la DGFIP cherchent à améliorer la performance énergétique et carbone de leurs centres de données et serveurs et en cherchant à développer des projets de récupération de la chaleur fatale, sans pour autant que cela conduise à un calcul de l'impact environnemental des systèmes d'IA utilisés sur leurs serveurs.

Il n'y a pas d'approche commune à l'échelle ministérielle sur le sujet, obligeant par exemple à traiter le sujet pour les modèles, mais aussi pour les données, le hardware et l'énergie mobilisée.

3.2.4 **Un risque de dilution des responsabilités**

En n'établissant pas de standard d'évaluation de l'impact environnemental des systèmes d'IA et en laissant ce sujet au niveau des administrations porteuses de systèmes d'IA, le risque

⁷⁸ *Carbon footprint* de Google, permet de mesurer l'empreinte carbone du cloud Google, et donc du développement de ses produits. Code Carbon, un outil intégré au langage de programmation Python, estime la quantité de dioxyde de carbone produite à la fois par le cloud et par les ressources informatiques personnelle utilisées pour exécuter un code.

est qu'il soit sous-traité car les porteurs de projet n'ont souvent la main que sur une partie des facteurs (données, modèles) et pas sur les autres (les serveurs sont à la DGFIP et la DGDDI). Or, l'impact d'un travail de frugalité sur les données ou les modèles pourrait paraître marginal pour les administrations, qui du point de vue du numérique responsable de l'État (et en application de la circulaire de novembre 2023), soutiennent d'abord l'amélioration du poste de travail et des data centers vertueux.

Dans ce contexte, afin de respecter les engagements ministériels concernant les services publics éco-responsables, la Cour recommande au SG des MEF et à la Dinum de promouvoir et de suivre l'évaluation de l'impact environnemental direct et indirect des systèmes d'IA du MEFSIN tout au long de leur cycle de vie. Cette évaluation comprend, notamment, leur empreinte carbone, leur consommation d'énergie et l'impact environnemental de l'extraction des matières premières nécessaires à la fabrication des technologies de l'IA. Le suivi mis en place doit permettre de réduire l'impact environnemental des systèmes d'IA et des infrastructures de données. La Dinum doit mettre à disposition des outils d'évaluation pour aider les administrations à évaluer cet impact environnemental.

La mesure de l'impact environnemental des systèmes d'IA tout au long de leur cycle de vie en vue de les réduire gagnerait à être insérée dans le plan de transformation écologique ministériel et dans ceux des directions.

Recommandation n° 4. (SG des MEF, Dinum) : Promouvoir auprès des directions des ministères économiques et financiers la mesure de l'impact environnemental des systèmes d'IA tout au long de leur cycle de vie en vue de le réduire.

En 2024, l'AFNOR a été mandatée par le ministère de la transition écologique (MTE) pour réunir en 2024 les parties prenantes désireuses de contribuer à la rédaction d'un futur référentiel pour calculer l'impact environnemental de l'IA⁷⁹. Le groupe de travail doit choisir les meilleurs indicateurs et méthodes de calcul pour évaluer l'impact environnemental de l'IA, définir les modalités de redevabilité et de communication en la matière.

Au regard de l'impact potentiel de la norme en cours d'élaboration, le SG des MEF pourrait utilement participer à ce groupe de travail.

3.3 L'IA frugale, le point de convergence entre l'efficacité, la confiance et les intérêts environnementaux et de souveraineté

Le paradigme d'IA frugale soutient qu'au-delà des seuls résultats pour l'environnement, la sobriété en matière de données, de modèles et d'infrastructure offre aussi des avantages concernant l'efficacité de la dépense, la robustesse des systèmes d'IA et la capacité des organisations à contrôler ces systèmes de façon souveraine.

⁷⁹ Le référentiel sera nommé AFNOR Spec 2314

Cette approche peut exister au sein des équipes portant les systèmes d'IA au sein du MEFSIN mais n'est pas animée à l'échelle ministérielle, au détriment du partage de bonnes pratiques et des gains d'efficacité.

3.3.1 L'IA frugale, une efficacité de la dépense

Bien qu'il n'existe pas de définition internationale standardisée pour l'IA frugale, une définition commune est la recherche de l'équilibre entre la performance des modèles d'IA et les contraintes de ressources, visant à maximiser les performances tout en minimisant les coûts directs et les coûts masqués. Cette démarche s'élabore au moment de la conception des systèmes d'IA, mais peut être traitée au stade du POC et du déploiement.

La frugalité conduit les acteurs impliqués à s'interroger sur les choix effectués lors du stockage, de l'annotation de données, du calcul (puissance du calcul, longueur des algorithmes) et de l'achat des composants et de consommation d'énergies.

En utilisant moins de données, les coûts en infrastructures et en stockage se réduisent. Pour mémoire, le coût d'un *datacenter*, même de petite taille, s'élève en moyenne à 4,3 MUSD. La réduction du jeu de données (*few shot learning*) ou l'automatisation de l'annotation⁸⁰, peuvent permettre d'éviter l'annotation de gros jeux de données, une étape très onéreuse en ressources financières et humaines. Dans les applications spécifiques d'IA pouvant être embarquées sur des appareils, l'*Edge AI*, ou calcul en périphérie, permet un gain de temps et de nombreux autres avantages. En *Edge AI*, les algorithmes sont exécutés localement. Cela réduit drastiquement l'usage du cloud et les transferts de données jusqu'à des *datacenter*, mais aussi la latence et les coûts. L'*Edge AI* offre aussi une meilleure confidentialité et disponibilité en raison de sa capacité à fonctionner hors ligne. La démarche d'IA frugale incite à utiliser des composants à l'image des puces neuromorphiques, permettent de consommer jusqu'à mille fois moins d'énergie qu'un processeur conventionnel de même taille, en réduisant les transferts de données.

Pour garantir une indépendance face aux pénuries de matériel, les approches frugales recommandent l'utilisation de solutions nécessitant le moins de matériel possible ou utilisant du matériel déjà disponible. La pénurie de GPU (processeurs graphiques) et des CPI (unités centrales) depuis 2020, issue des difficultés d'approvisionnement en semi-conducteurs et en matériaux nécessaires à la fabrication de GPU durant l'épidémie de covid-19 et de la hausse de la demande, en partie issue du *gaming* et des cryptomonnaies, souligne l'importance de ces approches, incitant à réduire la dépendance aux GPU dernier cri. Des pratiques frugales, telles que l'adoption de modèles d'IA performants sur des architectures réduites, illustrent des solutions pratiques pour lutter contre l'obsolescence des composants. Le consulat français à San Francisco cite ainsi d'entrepreneurs développant des modèles d'IA performants en inférence en utilisant des architectures réduites et spécialisées (transformeurs intégrant du *prompting*, de l'*in-context learning* ou du *zero-shot learning*), ce qui permet une bonne performance en exécution sur des GPUs vieux de cinq ans⁸¹. Les algorithmes plus fractionnés d'IA frugale

⁸⁰ Le machine learning automatisé (AutoML) est un ensemble de méthodes visant à simplifier et automatiser toutes les étapes du machine learning pour le rendre plus accessible sans compromis sur la qualité des modèles.

⁸¹ Consulat français de San Francisco, adopter l'IA frugale, septembre 2023.

constituent une solution partielle à la difficulté d'accès aux GPU nécessaires au test des modèles, en ce qu'ils permettent à la fois d'utiliser moins de GPU et de réduire leur temps d'utilisation.

3.3.2 La frugalité peut contribuer à l'IA de confiance

L'IA frugale contribue à la stratégie d'IA de confiance par la réduction des biais, discrimination et une amélioration des conditions de travail.

Utiliser moins de données offre la possibilité d'en améliorer la qualité. Il est en effet plus économe, en termes de temps et d'effort, de garantir un bon degré de correspondance avec la réalité sur un petit jeu de données que sur un grand jeu de données. Dans les gros ensembles de données non triés, certains types de données sont surreprésentés.

Utiliser moins de données facilite l'explicabilité. Des techniques frugales telles qu'AutoML peuvent représenter des solutions intéressantes, puisqu'elles fournissent des résultats explicables et reproductibles. Par ailleurs, des études ont montré que l'alliance du raisonnement bayésien et de l'apprentissage automatique pouvait créer des IA explicables et capables d'estimer la certitude de leurs prédictions. Cette approche est qualifiée de frugale puisqu'elle fonctionne sur des quantités réduites de données et présente une grande efficacité énergétique.

La démarche d'IA frugale nécessite une approche holistique, dans laquelle chaque métier de l'IA se sent investi d'une responsabilité en la matière afin de faire progresser la technologie de manière impactante, et plus acceptable du point de vue du grand public. Il est donc essentiel d'apprendre à créer, implémenter, acheter et utiliser l'IA de manière raisonnée, en évaluant attentivement les compromis entre la performance, la complexité et la consommation de ressources naturelles précieuses, afin d'exploiter au mieux cet outil en constante évolution.

3.3.3 Classifier les IA selon leur frugalité

Chaque acteur de l'IA peut en effet réaliser des améliorations de portée significative dans son champ de spécialité, tout en sachant que l'impact de ces actions peut être démultiplié dès lors qu'elles sont menées de manière coordonnée.

L'IA frugale nécessite de répondre aux questions ci-dessous et de faire des choix, sur le temps de calcul et sur la performance de l'IA.

- Un système d'IA est-il nécessaire ?
- Quelle est l'empreinte environnementale de chaque étape du cycle de vie du système d'IA ?
- Peut-on reproduire des résultats similaires avec moins de ressources ?

Les réponses à ces questions peuvent être traduites en classant les IA en deux grands types, sur la base des méthodes utilisées et de leur impact énergétique. L'IA rouge, ou « *Red*

AI »⁸², recherche à tout prix l'amélioration de la précision et de la puissance des modèles en les complexifiant et en utilisant une puissance de calcul massive, quitte à augmenter son impact environnemental et social. L'IA verte, ou « *Green AI* », se distingue par son efficacité, elle vise à atteindre une performance égale à celle de *Red AI* tout en réduisant au maximum les coûts de calculs. Son impact environnemental et social est un critère d'évaluation à part entière de sa performance⁸³.

Une classification de cette nature permet une analyse rigoureuse de l'efficacité attendue et réalisée pour chaque système d'IA et d'objectiver les enjeux propres à chaque système d'IA. Les administrations du MEFSIN ne classent pas leurs systèmes d'IA en termes d'IA frugale. Certaines administrations, comme la DGE pour Signaux Faibles, indiquent ne pas avoir utilisé la technique des réseaux de neurones, en considérant que la technologie d'IA utilisée, moins sophistiquée mais plus frugale, suffit à atteindre leur objectif. Il n'existe pas de diaporama ministériel sur le sujet.

Dans ce contexte, la Cour plaide pour l'instauration d'une analyse rigoureuse de l'efficacité de chaque système d'IA, fondée sur une démarche de frugalité et de confiance menée par l'instance de pilotage ministérielle. Sans ralentir l'appropriation de l'IA par les administrations, l'analyse de l'efficacité doit inciter les porteurs de projets d'IA à inscrire leurs projets dans une stratégie d'IA de confiance et à expliquer les améliorations envisagées ainsi que les expérimentations et arbitrages effectués pour favoriser la mise en œuvre d'une IA frugale. Dans cette perspective, il serait utile que l'instance ministérielle soit en capacité de labelliser les systèmes d'IA du MEFSIN en IA rouge ou IA verte.

<p>Recommandation n° 5. (SG des MEF) : Promouvoir auprès des directions une démarche de frugalité ministérielle lors de la mise en œuvre de systèmes d'IA au MEFSIN.</p>

CONCLUSION INTERMÉDIAIRE

La prise en compte des enjeux de confiance et des défis environnementaux lors de la mise en œuvre des systèmes d'IA est en phase d'amorçage au sein du MEFSIN. L'étape à franchir est celle du pilotage de ces enjeux.

La compatibilité de ces différents facteurs, de confiance, environnementaux, frugaux, est une opportunité rare pour l'adoption d'une IA publique robuste, puisqu'elle offre des leviers pour relever les défis actuels de l'IA à tous les niveaux de la chaîne de valeur (des concepteurs jusqu'aux utilisateurs). Cet alignement permet une approche holistique : chaque acteur peut réaliser des améliorations de portée significative dans son champ de spécialité, tout en sachant

⁸² Roy Schwartz et son équipe du Allen Institute for AI ont théorisé en 2019 ces deux grands concepts en IA.

⁸³ L'IA frugale proche de Green AI ne se confond pas avec l'IA au service de l'environnement (« AI for Green »). L'IA frugale consiste à réduire l'impact direct de l'IA sur l'environnement et l'IA au service de l'environnement consiste à utiliser l'IA comme un outil de travail pour gagner en efficacité dans une démarche de préservation de l'environnement.

que l'impact de ces actions peut être démultiplié dès lors qu'elles sont menées de manière coordonnée.

Cliquez ou appuyez ici pour entrer du texte.

ANNEXES

Annexe n° 1.	Glossaire.....	73
Annexe n° 2.	Description des systèmes d'IA mis en œuvre ou projetés par le MEFSIN entre 2018 et 2023	77
Annexe n° 3.	Présentation de la méthode d'évaluation de la maturité des systèmes d'IA en termes de confiance	82
Annexe n° 4.	Sept principes d'IA de confiance	87

Annexe n° 1. Glossaire

Agents : Les agents sont l'ensemble des personnels employés par le MEFSIN.

Algorithme : Un algorithme décrit une séquence d'étapes à suivre pour résoudre un problème, de façon suffisamment précise pour être transcrite en un programme pouvant être mis en œuvre par un ordinateur. Selon la nature du problème, le résultat peut être unique (effectuer une multiplication par exemple) ou comporter plusieurs réponses (proposer plusieurs pages web en réponse à une recherche par exemple). Le mot algorithme vient du nom d'un mathématicien perse du IX^e siècle, Al-Khwârizmî. La science qui étudie les propriétés des algorithmes est l'algorithmique.

Anssi : Agence nationale de la sécurité des systèmes d'information

API : interface de programmation applicative

Biais algorithmiques : Appliqué au contexte d'un algorithme, un biais peut advenir dans le cas où les données fournies, leur entraînement ou la construction de l'algorithme lui-même sont susceptibles de générer des erreurs, des approximations ou des discriminations. Exemple : Un algorithme cherchant le candidat attendu pour un poste donné est entraîné avec un échantillon de CV reflétant les effectifs majoritairement masculins associés à ce type de poste. Il est possible, sans opération de débiaisement, que l'algorithme favorise en reflet les candidatures masculines.

CITP : comité interministériel à la transformation publique

Connaissances en matière d'IA : Les connaissances en matière d'IA désignent les compétences et les ressources, à l'instar des données, du code, des algorithmes, des modèles, de la recherche, du savoir-faire, des programmes de formation, de la gouvernance, des processus et des meilleures pratiques nécessaires pour comprendre le cycle de vie des systèmes d'IA et y prendre part. – **Acteurs de l'IA** : Les acteurs de l'IA sont les parties jouant un rôle actif dans le cycle de vie d'un système d'IA, y compris les organisations et les individus qui déploient ou exploitent l'IA.

Cycle de vie d'un système d'IA : Le cycle de vie d'un système d'IA comporte plusieurs phases : i) la phase de « conception, données et modèles », séquence dépendante du contexte comprenant la planification et la conception, la collecte et le traitement des données, ainsi que la construction du modèle ; ii) la phase de « vérification et validation » ; iii) la phase de « déploiement » ; et iv) la phase d'« exploitation et (de) suivi ». Ces phases présentent souvent un caractère itératif et ne suivent pas nécessairement un ordre séquentiel. La décision de mettre un terme à l'utilisation d'un système d'IA peut intervenir à n'importe quel stade de la phase d'exploitation et de suivi. OCDE. [OECD-LEGAL-0449-fr.pdf](#)

DAE : direction des achats de l'État

DB : direction du budget

DGAFP : direction générale de l'administration et de la fonction publique

DGCCRF : direction générale de la concurrence, de la consommation et de la répression des fraudes

DGFIP : direction générale des finances publiques

DINUM : direction interministérielle du numérique

DITP : direction interministérielle de la transformation publique

Dnum : direction du numérique ministérielle

EIG : entrepreneur d'intérêt général

Équité de traitement ou traitement équitable : L'équité de traitement vise à prendre en compte la situation singulière de chacun afin de proposer les décisions les plus justes. Exemple : Un demandeur d'emploi très éloigné de l'emploi se verra proposer un accompagnement renforcé et différencié en accord avec sa situation pour faciliter son retour à l'emploi.

Éthique : L'éthique est la discipline philosophique de réflexion sur les fondements des règles de conduite et leur mise en œuvre selon les contextes. Elle se distingue de la réglementation, qui établit des normes, et du droit qui édicte des lois et aspire à les faire respecter.

ETP(T) : équivalent temps plein (travaillé)

FTAP : fonds pour la transformation de l'action publique

IGF : inspection générale des finances

IGPDE : institut de la gestion publique et du développement économique

Intelligence artificielle : ensemble de théories et de techniques mises en œuvre pour réaliser des logiciels capables de simuler des fonctions cognitives, utilisées à des fins de résolution de problèmes, d'assistance ou de substitution à des activités humaines. Exemple : L'intelligence artificielle peut être utilisée pour « lire » des documents et en extraire les informations clés afin de pré-remplir les champs d'un formulaire qu'un humain devra ensuite contrôler et valider.

LFI : loi de finances initiale

MAREVA : méthode interministérielle d'analyse et de remontée de la valeur. Les études dites « MAREVA » estiment la valeur prévisionnelle des projets de système d'information, en rapportant leur coût estimé et leurs bénéfices attendus, afin de produire un indice de retour sur investissement.

MEFSIN : ministère de l'économie, des finances et de la souveraineté industrielle et numérique

MTFP : ministère de la transformation et de la fonction publiques

Parties prenantes de l'IA : elles englobent toutes les organisations et personnes impliquées ou affectées par les systèmes d'IA, directement ou indirectement. Les acteurs de l'IA sont un sous-ensemble de parties prenantes. Définition OCDE.

RG2A : Pour faciliter la mise en accessibilité des sites et services numériques, la direction interministérielle du numérique (DINUM) édite le référentiel général d'amélioration de l'accessibilité,

RPA : (Robotic Automatisation Process), est une technologie utilisant des logiques métiers et des intrants structurés, dont l'objet est d'automatiser les processus métier basés sur des règles structurées. Grâce aux outils de RPA, une organisation peut configurer un logiciel, ou un "robot", pour capturer et interpréter des applications afin de traiter une transaction, manipuler des données, déclencher des réponses et communiquer avec d'autres systèmes

RSSI : responsable de la sécurité des systèmes d'information

Système expert : Dans le domaine de l'intelligence artificielle, un système expert est une application conçue pour modéliser le raisonnement d'un spécialiste humain, en reproduisant

ses mécanismes cognitifs. Le logiciel applique un raisonnement à partir de faits connus, et de règles implémentées au départ. Il peut réaliser des déductions ou conclusions, et expliquer la manière dont les résultats ont été obtenus. Le premier système expert a été inventé en 1965. Nommé Dendral, il était capable d'identifier, à partir de résultats de spectrométrie de masse et de résonance magnétique nucléaire, les constituants chimiques d'un matériau.

Systèmes d'IA: Un système d'IA est un système qui fonctionne grâce à une machine et capable d'influencer son environnement en produisant des résultats (tels que des prédictions, des recommandations ou des décisions) pour répondre à un ensemble donné d'objectifs. Il utilise les données et les intrants générés par la machine et/ou apportés par l'homme afin de (i) percevoir des environnements réels et/ou virtuels ; (ii) produire une représentation abstraite de ces perceptions sous forme de modèles issus d'une analyse automatisée (ex. l'apprentissage automatisé) ou manuelle ; et (iii) utiliser les déductions du modèle pour formuler différentes options de résultats. Les systèmes d'IA sont conçus pour fonctionner de façon plus ou moins autonome. Définition OCDE.

SNIA : La stratégie nationale d'intelligence artificielle (SNIA) a été lancée le 28 mars 2018 pour positionner la France comme un acteur majeur de l'IA. La SNIA a été élaborée à la suite des avancées de l'apprentissage profond et dans le contexte de la publication des rapports de France Stratégie en 2017, de celui de M. Cédric Villani en 2018, des travaux conduisant à la stratégie européenne pour l'IA publiée par la Commission européenne en avril 2018⁸⁴, et de l'adoption par l'OCDE en mai 2019⁸⁵ des principes sur l'intelligence artificielle.

Dans sa première phase qui a duré de 2018 à 2022, la SNIA avait identifié cinq domaines à soutenir avec 1527 M€ de financements publics. La transformation publique constituait un de ces cinq volets, avec la recherche, l'enseignement supérieur, la diffusion dans l'économie et la défense et la sécurité. Le financement prévisionnel de la transformation publique par l'État représentait 10 % du budget initial de la SNIA⁸⁶. La diffusion de l'intelligence artificielle et des sciences des données dans l'économie et dans l'administration et la promotion d'un modèle éthique équilibré entre innovation et protection des droits fondamentaux constituait deux des trois priorités de la SNIA⁸⁷.

Le volet transformation publique de la SNIA et la diffusion de l'IA et des sciences des données dans l'administration manifestaient l'obligation d'adaptabilité ou de mutabilité des services publics⁸⁸. Le bon fonctionnement du service public suppose en effet une adaptation aux changements de la société, aux besoins des usagers et aux évolutions techniques.

⁸⁴ Un plan de coordination européenne dans le domaine de l'IA a été publié en décembre 2018.

⁸⁵ OCDE, Recommandations du Conseil sur l'intelligence artificielle, adoptées en mai 2019, modifiées en novembre 2023

⁸⁶ Cour des comptes, la stratégie nationale de recherche en intelligence artificielle, avril 2023, page 30. Le volet recherche devait être financé à hauteur de 445 M€, le volet enseignement supérieur, 128 M€, le volet économie, 390 M€ et le volet défense et sécurité, 410 M€.

⁸⁷ La 3e priorité était de développer l'attractivité des talents.

⁸⁸ Le juriste Louis Rolland a théorisé au début du XXe siècle trois grands principes régissant le bon fonctionnement du service public et s'appliquant à toutes ses activités : la continuité, l'adaptabilité et l'égalité. Le principe de continuité prescrit un fonctionnement régulier des services publics, sans interruption autre que celles autorisées par la législation. Le principe d'adaptabilité impose au gestionnaire d'un service public de faire évoluer son organisation en fonction des progrès techniques, des besoins des usagers et de toute circonstance nouvelle. Le principe d'égalité et son corollaire le principe de neutralité soulignent l'importance d'assurer à tous, sans aucune

Mais contrairement aux volets recherche et enseignement supérieur de la SNIA, la transformation publique par l'IA n'a pas fait l'objet d'une programmation particulière.

Depuis 2022, la SNIA se déploie aujourd'hui dans une nouvelle phase dite d'"accélération". La transformation publique par l'IA n'apparaît plus nommément.

Usagers : ils sont l'ensemble des bénéficiaires des services du MEFSIN.

Utilisateurs : Ensemble des agents et usagers du MEFSIN susceptibles d'être en présence d'une solution d'intelligence artificielle ou de l'utiliser.

UX : expérience utilisateur

discrimination, un égal accès aux services publics. Ils correspondent à des principes généraux du droit administratif français. Les principes de continuité et d'égalité sont des objectifs à valeur constitutionnelle.

**Annexe n° 2. Description des systèmes d'IA mis en œuvre ou projetés par le
MEFSIN entre 2018 et 2023**

Nom du SIA ou du projet de SIA	Direction ou service	Stade de développement	Description de la fonction du SIA
Chatbot Chorus Pro	AIFE	SIA déployé et exploité	<i>Chatbot</i> d'accompagnement des utilisateurs de Chorus-pro caractérisé par sa capacité d'enrichissement continu. Il gère en 2023 80% des questions (100000 sollicitations par mois).
Assistance au contrôle hiérarchisé de la dépense (IA CHD)	AIFE	SIA déployé et exploité	Outil d'amélioration du ciblage des contrôles hiérarchisés de la dépense. Il se distingue par une actualisation plus régulière des règles métier de hiérarchisation.
Callbot Chorus Pro	AIFE	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	<i>Callbot</i> sur Chorus-pro aidant les fournisseurs à saisir leurs factures, en remplacement d'un call center dont les temps d'attente étaient longs . Si le callbot ne sait pas répondre, un livechat est proposé.
IA RéponseConso	DGCCRF	SIA déployé et exploité	Système automatisé d'aide à la réponse aux consommateurs, utilisé par le centre de contact multicanal de la DGCCRF. La solution propose automatiquement les bons paragraphes pour constituer le courrier de réponse au consommateur, à partir d'une banque de paragraphes types et de l'analyse automatique du contenu texte de la demande et de ses pièces jointes (*) par une IA hybride : un algorithme « moteur de règles », combiné à un algorithme d'apprentissage « classifieur statistique ».
SESAM IA	DGCCRF	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Système automatisé d'analyse textuelle de documents ou de procès-verbaux pour restructurer les données qui y apparaissent et améliorer les enquêtes futures

L'INTELLIGENCE ARTIFICIELLE DANS LES POLITIQUES PUBLIQUES : L'EXEMPLE DU
MINISTÈRE DE L'ÉCONOMIE ET DES FINANCES

Nom du SIA ou du projet de SIA	Direction ou service	Stade de développement	Description de la fonction du SIA
Chatbot Code de la consommation	DGCCRF	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Le projet couvrirait deux fonctions: information des consommateurs et des professionnels (1), aide et appui à l'enquête pour la détection de fraude (2)
3D - Chatbot Accompagnement RH des agents	DGDDI	SIA déployé et exploité	Chatbot basé sur des règles métier pour accompagner les agents dans leurs actes de ressources humaines.
3D - Lutte contre la fraude	DGDDI	SIA en cours de développement	Outil visant à détecter des fraudes et orienter les contrôles dans différents domaines douaniers
3D - Détection non intrusive	DGDDI	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Expérimentation d'un système de détection de produits stupéfiants fondés sur des images à Rayons X de colis postaux et d'alerte des cibles jugées crédibles pour étayer une décision de contrôle.
Vielle Eco SISSE	DGE	SIA en cours de développement	Outil d'automatisation de la veille informative concernant l'intelligence économique et la détection de menaces de sécurité économique
Signaux faibles	DGE	SIA déployé et exploité	Outil de prédiction de défaillances d'entreprises au sein d'une zone géographique (SIA), associé à des renseignements sur l'ensemble des 426 000 entreprises et 1,8 million d'établissements en base (IHM) pour identifier et accompagner les entreprises fragilisées.
Ciblage de la fraude et valorisation des requêtes (CFVR)	DGFIP	SIA déployé et exploité	Aide au contrôle fiscal des entreprises et des particuliers en proposant aux services de contrôle une hiérarchisation des dossiers selon leur niveau de risque de fraude
Foncier Innovant	DGFIP	SIA déployé et exploité	Outil d'identification de bâtiments, de comparaison avec les autorisations d'urbanisme et les déclarations fiscales et de signalements
Econtact+	DGFIP	SIA déployé et exploité	Outil d'analyse-sémantique aidant les agents en centre de contacts à traiter les messages des usagers, en leur proposant une réponse type pour certaines questions.

Nom du SIA ou du projet de SIA	Direction ou service	Stade de développement	Description de la fonction du SIA
VCI – Valorisation des cessions immobilières	DGFIP	SIA en cours de développement	Outil de prédiction la valeur vénale (prix du marché) d'un bien immobilier à usage d'habitation (maison ou appartement), dans le but de la comparer au prix de vente déclaré.
AGATHE – Prédire les dysfonctionnements applicatifs	DGFIP	SIA en cours de développement	Outil d'analyse automatique (de masse) de la qualité de service des applications de la DGFIP envers les agents et à mieux anticiper les difficultés à venir concernant ces applications
AGATHE – Traitement Naturel du Langage sur les remontées utilisateurs	DGFIP	SIA en cours de développement	Outil automatique d'aide au classement des remontées agents issus de canaux de communication distincts
IA CHD des collectivités	DGFIP	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil d'amélioration du contrôle hiérarchisé de la dépense des collectivités.
Suivi des amendements PLF	DGFIP	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil de suivi des amendements aux PLF, permettant leur analyse sémantique, le regroupement de familles d'amendements, l'identification des doublons et produisant un résumé.
TNMJ (Transformation numérique des métiers du juridique)	DGFIP	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil de génération de réponses permettant de donner aux usagers la position de l'administration sur des sujets contentieux, en mobilisant des données plus nombreuses et en moins de temps que ce qui est possible humainement.
OCRisation	DGFIP	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil de lecture pour extraire des données numériques de qualité à partir de documents scannés (typographiques et à terme manuscrits, formulaires, lettres) en cas d'absence ou de limite des téléprocédures

Nom du SIA ou du projet de SIA	Direction ou service	Stade de développement	Description de la fonction du SIA
Econtact (volet GMBI)	DGFIP	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil de génération de texte aidant à répondre aux demandes des usagers, fondé sur les FAQ de la DGFIP aux usagers ainsi qu'un échantillon de questions/réponses réelles.
Génération de code informatique	DGFIP	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil d'écriture de code pour transformer du code SAS en Python.
Paye des agents	DGFIP	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil de définition des jeux d'essais représentatifs pour les tests de paye des agents de l'État compte-tenu de l'importante diversité des situations
Codification APE SIRENE	Insee	SIA déployé et exploité	Outil remplaçant un système expert en fin de vie (SICORE) qui proposait aux gestionnaires SIRENE un/des code(s) APE en fonction de la description de l'activité d'une entreprise
Codification PCS RP	Insee	SIA en cours de développement	Outil remplaçant un système expert en fin de vie (SICORE) qui proposait aux gestionnaires RP un/des code(s) PCS en fonction de la description de l'activité professionnelle des personnes enquêtées dans le recensement de la population
Passage du langage SAS au langage R	Insee	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil d'appui à la conversion de code permettant d'explicitier un code existant mal documenté et de transformer des parties de code SAS en langage "R"
Assistant de veille sociale	SG et Bercy Hub	SIA en cours de développement	Outil d'étude textuelle des publications des organisations syndicales du MEFSIN, traduite par nuage de mots et tableau de bord, filtrable par OS. Il alimente la note ministérielle sur le climat social. Méthode non supervisée "text rank" inspiré de "page rank" qui regarde des similarités et redondances dans des corpus de texte. L'algo classe les phrases importantes (stabilo magique).

L'INTELLIGENCE ARTIFICIELLE DANS LES POLITIQUES PUBLIQUES : L'EXEMPLE DU
MINISTÈRE DE L'ÉCONOMIE ET DES FINANCES

Nom du SIA ou du projet de SIA	Direction ou service	Stade de développement	Description de la fonction du SIA
Baromètre numérique agents (BNA)	SG et Bercy Hub	SIA déployé et exploité	Outil de mesure de sentiments analysant les 15 000 réponses des agents au questionnaire annuel du SG sur leur environnement numérique, périphérique, réseaux + notes et verbatim, espaces d discussion.
Labellisation et texte dans les vidéos	SG et Bercy Hub	SIA déployé et exploité	Outil de labellisation automatique des prononcés produits dans des vidéos, utilisé pour la recherche documentaire et pour la traduction en temps réel.
Archivage des mails des cabinets ministériels	SG et Bercy Hub	SIA en cours de développement	Outil de filtrage des messageries des conseillers de cabinet, pour distinguer les messages professionnels des messages personnels, fondés sur la labellisation d'un échantillon de données effectuée avec des archivistes.
Attentes des usagers sur la qualité des données	SG et Bercy Hub	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Outil de mesure de sentiments, fondé sur l'historique des discussions publiques sur data.gouv concernant les jeux de données de Bercy, utilisé pour entraîner pour entraîner le NLP, identifier les grandes thématiques et inférer sur les nouveaux enjeux de qualité des données.
Intégration de déclarations	Autre service	SIA déployé et exploité	Système automatique de correction et d'intégration de signalements.
Orientations de déclarations	Autre service	SIA déployé et exploité	Système d'orientation automatisé de signalements vers des services d'analyse appropriés.
Entités nommées	Autre service	SIA envisagés, ou faisant l'objet d'une étude de faisabilité	Système automatique d'exploitation permettant d'identifier des relations entre différents signalements.

Annexe n° 3. Présentation de la méthode d'évaluation de la maturité des systèmes d'IA en termes de confiance

La Cour a souhaité évaluer la maturité des systèmes d'IA de Bercy en termes de confiance. Pour ce faire, la Cour a établi un questionnaire lié à six axes déterminants pour évaluer le niveau de confiance pouvant être accordé à une IA.

I – Le questionnaire

Principes	Axes d'analyse
<i>Transparence</i>	Existe-t-il une charte ou politique interne encadrant la conception et le déploiement du système d'IA (SIA) ?
	L'objectif (finalité) du traitement reposant sur l'utilisation d'un système d'IA (SIA) est-il clairement défini et accessible* aux usagers* du système d'IA ?
	Un cahier des charges a-t-il été réalisé et est-il accessible* aux usagers du SIA* ?
	Le code du système d'IA est-il expliqué*, maîtrisé (droit de modification ou d'intervention sur le code) et publié (open source) ?
<i>Responsabilisation</i>	Y-a-t-il eu un audit interne (vérifiant la réalité des besoins et la conformité au cahier des charges et mesurant l'impact organisationnel) avant le déploiement du SIA ?
	Est-il possible d'expliquer* le résultat proposé par le SIA ?
<i>Facteur et contrôle humain</i>	Les personnes physiques en charge de la conception, du développement tout le long du cycle de vie du SIA* ont-elles été sensibilisées aux enjeux juridiques, techniques, éthiques et moraux de l'IA ?
	Les responsabilités des différentes personnes (physiques et morales) dans le processus du système d'IA sont-elles définies et contrôlées ?
	Les phases d'apprentissage et de production du système d'IA sont-elles distinctes ? Si oui, une seconde évaluation est-elle prévue concernant la phase de production ?
<i>Respect de la vie privée et gouvernance des données</i>	L'infrastructure hébergeant le système d'IA respecte-t-il SecNumCloud ?
	Les données d'apprentissage du système d'IA sont-elles mises à jour ?
<i>Diversité, équité et non-discrimination</i>	Quels contrôles existent-ils sur les biais de données, de modélisation et de fiabilité des explications ?
	Existe-t-il un comité d'éthique et scientifique ayant analysé le SIA ? Dans l'hypothèse positive, ces analyses sont-elles accessibles ?
<i>Bien être sociétal et environnemental</i>	Dans quelle mesure le système d'IA est aligné avec les objectifs d'un État exemplaire dans la transition écologique et ses critères de conception et de réalisation respectent les engagements de l'État en matière de services publics écoresponsables ?

L'INTELLIGENCE ARTIFICIELLE DANS LES POLITIQUES PUBLIQUES : L'EXEMPLE DU
 MINISTERE DE L'ECONOMIE ET DES FINANCES

	La CNIL recommande, lors de la mise en place d'un système d'IA, de s'interroger sur son objectif (sa finalité) et sur la proportionnalité des techniques choisies : sont-elles strictement nécessaires pour atteindre cet objectif ?
<i>Sécurité robustesse technique</i>	Les accès aux données d'entrée sont-ils restreints et sécurisés ?
	Quel chiffrement des données est employé ?
	Y-a-t-il une alerte si les données d'entrée sont modifiées?
	Un plan de continuation de l'activité (PCA)* a-t-il été défini ?
	Y-a-t-il eu des tests à l'inverse* qui ont été réalisés ?
<i>Ecosystème</i>	Quel accompagnement a été sollicité auprès du SG des MEF, de la DINUM, de la DITP, de la CNIL au cours du cycle de vie du SIA?

II – Grille de maturité

Principes	Niveau 0 Non pris en compte	Niveau Un Initial	Niveau deux Géré	Niveau trois Standard	Niveau Piloté quatre	Niveau cinq Optimisé
<i>Transparence</i>	Non pris en compte.	Une charte encadrant la conception et le déploiement du système d'IA est en cours de rédaction	Il existe une charte ou politique interne encadrant la conception et le déploiement du système d'IA L'objectif (finalité) du traitement reposant sur l'utilisation d'un système d'IA (SIA) est-il clairement défini	Un cahier des charges a été réalisé Le code du système d'IA est expliqué	L'objectif (finalité) du traitement reposant sur l'utilisation d'un système d'IA est accessible aux usagers Le cahier des charges est accessible aux usagers Le code du système d'IA est maîtrisé (droit de modification ou d'intervention sur le code)	Le code du système d'IA est publié en open source
<i>Responsabilisation</i>	Non pris en compte.			Il est possible d'expliquer le résultat proposé par un SIA	Avant chaque déploiement de systèmes d'IA un audit interne vérifiant la réalité des besoins et la conformité au cahier des charges et mesurant l'impact organisationnel est réalisé	

L'INTELLIGENCE ARTIFICIELLE DANS LES POLITIQUES PUBLIQUES : L'EXEMPLE DU
MINISTÈRE DE L'ÉCONOMIE ET DES FINANCES

<p><i>Facteur et contrôle humain</i></p>	<p>Non pris en compte.</p>	<p>Les phases d'apprentissage et de production du système d'IA sont distinctes</p>	<p>Les personnes physiques en charge de la conception, du développement tout le long du cycle de vie du système d'IA sont sensibilisées aux enjeux juridiques, techniques, éthiques et moraux de l'IA</p> <p>Les responsabilités des différentes personnes (physiques et morales) dans le processus du système d'IA sont définies</p>		<p>Les responsabilités des différentes personnes (physiques et morales) dans le processus du système d'IA sont contrôlées</p> <p>une seconde évaluation est prévue concernant la phase de production</p>	
<p><i>Respect de la vie privée</i></p>	<p>Non pris en compte.</p>		<p>L'infrastructure hébergeant le système d'IA respecte SecNumCloud</p>			<p>Les données d'apprentissage du système d'IA sont mises à jour</p>
<p><i>Diversité, équité et</i></p>	<p>Non pris en compte.</p>			<p>Il existe un comité d'éthique et scientifique ayant analysé le SIA</p>	<p>Des contrôles existent sur les biais de données, de modélisation et de fiabilité des explications</p>	<p>Les analyses du comité d'éthique sont accessibles</p>
<p><i>Bien être sociétal et environnemental</i></p>	<p>Non pris en compte.</p>		<p>L'objectif (sa finalité) et la proportionnalité des techniques choisies pour le système d'IA ont été évaluées et strictement nécessaires pour atteindre l'objectif</p>		<p>Le système d'IA est aligné avec les objectifs d'un État exemplaire dans la transition écologique et ses critères de conception et de réalisation respectent les engagements de l'État en matière de services publics écoresponsables</p>	

S sécurité et	Non pris en compte.	Les accès aux données d'entrée sont restreints et sécurisés	Les données sont chiffrées		Une alerte existe si les données d'entrée sont modifiées	

III- Analyse du niveau de confiance par type de SIA

Le questionnaire présenté dans la 1^{ère} partie a été transmis aux porteurs de systèmes d'IA des administrations de Bercy afin qu'ils apportent une auto-évaluation de leur IA. 21 systèmes d'IA ont fait l'objet d'une réponse durant l'instruction et ont pu être pris en compte. Les réponses apportées sont déclaratives et n'ont pas fait l'objet de vérification.

Lors de l'analyse des réponses, nous notons que deux questions semblent avoir pu être interprétées de différentes manières et amènent des réponses diverses. Il s'agit des questions liées à la sécurité et la robustesse : « Y-a-t-il une alerte si les données d'entrée sont modifiées ? » et « Y-a-t-il eu des tests à l'inverse qui ont été réalisés ? ».

Afin d'obtenir des résultats les plus égaux possibles, le parti a été pris de ne pas prendre en compte les réponses apportées dans ces deux questions et de ne pas inclure les résultats dans l'analyse.

Pour chaque systèmes d'IA, une note a été attribuée permettant d'évaluer la confiance suivant la grille de maturité ci-dessus.

A partir de cette évaluation du niveau de confiance, des graphiques en fonction des caractéristiques des systèmes d'IA ont été réalisés. Ils sont présentés dans le corps du rapport.

Annexe n° 4. Sept principes d'IA de confiance

Les travaux de recherche, les rapports administratifs, les résolutions et les chartes précitées⁸⁹ ont ainsi dégagé un corpus de principes caractérisant l'IA de confiance et luttant contre les risques et biais présentés ci-dessus.

– la transparence dans leur emploi ;

Les organisations mettant en œuvre des systèmes d'IA s'engagent à assurer la transparence et une divulgation responsable des informations liées aux systèmes d'IA. À cet effet, ils doivent fournir des informations pertinentes, adaptées au contexte et à l'état de l'art, afin de favoriser une compréhension générale des systèmes d'IA, d'informer les parties prenantes de leurs interactions avec les systèmes d'IA. La finalité et la légitimité des algorithmes doivent être explicitées et encadrées pour valider la pertinence de l'IA au regard des objectifs recherchés.

– la primauté humaine ;

Les algorithmes et solutions d'intelligence artificielle doivent être développés et utilisés dans le respect de l'autonomie humaine. Ils doivent pouvoir aider les individus à prendre des décisions et éclairer leurs choix, sans pour autant ni les y contraindre, ni éluder leur responsabilité ;

– L'explicabilité et la responsabilité :

Le résultat du système d'IA est explicable et expliqué. Les organisations mettant en œuvre des systèmes d'IA fournissent les informations, adaptées au contexte et à l'état de l'art, permettant aux personnes subissant les effets néfastes d'un système d'IA de contester les résultats sur la base d'explications claires et facilement compréhensibles sur les facteurs et sur la logique ayant servi à la formulation de prévisions, recommandations ou décisions.

Les acteurs de l'IA sont responsables du bon fonctionnement des systèmes d'IA et du respect des décisions prises, selon leurs rôles, le contexte et l'état de l'art.

– l'équité et la non-discrimination ;

Les systèmes d'intelligence artificielle peuvent reproduire, renforcer ou générer des biais discriminatoires. C'est pourquoi les organisations mettant en œuvre des systèmes d'IA s'engagent à produire des mesures, adaptées au contexte et à l'état de l'art, visant à garantir l'équité entre les individus, l'absence de discriminations et la fiabilité des solutions proposées. Les concepteurs du système d'IA doivent formaliser leur définition de l'équité, présenter les moyens mis en place pour lutter contre les biais discriminatoires involontaires et garantir en toutes circonstances l'accessibilité et l'universalité des systèmes vis-à-vis des usagers. Les acteurs de l'IA doivent, selon leurs rôles respectifs, le contexte et leur capacité à agir, appliquer

⁸⁹ Rapport de Cédric Villani : donner un sens à l'intelligence artificielle (IA), 2018 ; OCDE, recommandation du conseil sur l'intelligence artificielle, 2019 ; Unesco, recommandation sur l'éthique de l'IA, 2021 ; conseil d'État, intelligence artificielle et action publique : construire la confiance, servir la performance, 2022 P22. L'IA a conduit à la publication de rapports publics plus récents. Il s'agit par exemple du rapport du comité interministériel sur l'IA générative (2024) et celui du Sénat, l'IA et l'avenir du service public, rapport thématique, 1 impôts, prestations sociales et lutte contre la fraude, Didier Rambaud et Sylvie Vermeillet (2024).

de manière continue une approche systématique de la gestion du risque, à chaque phase du cycle de vie des systèmes d'IA, afin de gérer les risques liés aux biais.

- la sécurité et la robustesse techniques des infrastructures et des modèles, tout au long du cycle de vie des solutions ;

Il s'agit de s'assurer de la résilience des systèmes d'IA en cas d'attaques. Les systèmes d'IA devraient être robustes, sûrs et sécurisés tout au long de leur cycle de vie, de sorte que, dans des conditions d'utilisation normales ou prévisibles, ou en cas d'utilisation abusive ou de conditions défavorables, ils soient à même de fonctionner convenablement, et ne fassent pas peser un risque de sécurité démesuré. Les acteurs de l'IA doivent, selon leurs rôles respectifs, le contexte et leur capacité à agir, appliquer de manière continue une approche systématique de la gestion du risque, à chaque phase du cycle de vie des systèmes d'IA, afin de gérer les risques y afférents, notamment ceux liés à la sécurité numérique.

- une politique de gestion des données pour assurer le respect de la vie privée ;

Pour ce faire, les acteurs de l'IA doivent veiller à la traçabilité, notamment pour ce qui est des ensembles de données, afin de permettre l'analyse des résultats produits par lesdits systèmes d'IA et le traitement des demandes d'information, compte tenu du contexte et de l'état de l'art de la technologie. Les acteurs de l'IA doivent, selon leurs rôles respectifs, le contexte et leur capacité à agir, appliquer de manière continue une approche systématique de la gestion du risque, à chaque phase du cycle de vie des systèmes d'IA, afin de gérer les risques y afférents, notamment ceux liés au respect de la vie privée.

- L'ambition de contribuer au bien-être sociétal et environnemental parachève les principes dégagés pour instaurer la confiance des parties prenantes de l'IA dans les systèmes d'IA.