

Cour des comptes



ÉVALUATION DES POLITIQUES PUBLIQUES

LA STRATÉGIE NATIONALE DE RECHERCHE EN INTELLIGENCE ARTIFICIELLE

Une stratégie à structurer
et à pérenniser

Rapport public thématique
Avril 2023

Sommaire

Synthèse	9
Récapitulatif des recommandations	15
Glossaire	17
Introduction.....	21
Chapitre I Une priorité donnée à la recherche qui n’a pas encore produit les résultats attendus	25
I - Un rôle prépondérant accordé à la recherche en intelligence artificielle donnant lieu à un plan spécifique	26
A - La France, un des premiers pays à s’être doté d’une stratégie spécifique en intelligence artificielle	26
B - Des financements du plan national portant principalement sur la recherche	29
C - L’IA, une priorité qui se diffuse au-delà du strict cadre de la stratégie.....	32
D - Un développement de l’IA éclaté entre plusieurs politiques publiques	34
II - Une efficacité difficile à apprécier en raison d’un suivi insuffisant des financements associés.....	37
A - Un dispositif budgétaire reposant surtout sur des outils financiers déjà existants et ne permettant pas d’appréhender la réalité de la dépense.....	37
B - Des outils de financements de courte durée complétés par des financements génériques importants, français et européens	39
C - Un suivi non exhaustif de l’exécution financière de la stratégie.....	42
III - Un positionnement scientifique de la France en IA en retrait depuis 2018.....	46
A - Une production scientifique globalement constante sur la période.....	46
B - Une amélioration depuis 2018 de la valorisation de la recherche en IA	49
Chapitre II Un écosystème à mieux structurer et à pérenniser	55
II - Une stratégie de pôles d’excellence pertinente qui doit gagner en maturité.....	56
A - Le renforcement des bassins actifs en IA par la labellisation d’instituts 3IA et de chaires individuelles.....	57
B - Une performance des 3IA et des autres centres d’excellence comparable en matière de production scientifique.....	60
C - Une visibilité des instituts sur la scène internationale encore à construire	62
D - Des modes de financement des centres d’excellence reposant sur une logique de trop court terme.....	63

III - Un investissement en capital humain pour la recherche en IA ciblé sur les chercheurs juniors	65
IV - Un accès renforcé aux infrastructures souveraines de calcul.....	70
A - Une appropriation rapide par la communauté scientifique du supercalculateur Jean Zay.....	70
B - Les effets du supercalculateur sur l'attractivité et la crédibilité de la France en matière d'IA	74
V - Un écosystème national encore structuré autour des acteurs académiques et des administrations nationales	76
A - Un rôle encore prépondérant des organismes de recherche, sites universitaires et administrations dans la structuration de l'écosystème	77
B - Les instituts interdisciplinaires allemands, plus récents et plus efficaces en matière de structuration de l'écosystème d'IA	81
C - Des « communs » mis en place pour renforcer la communauté	83
VI - Une gouvernance affaiblie avec le temps et fragmentée à laquelle il faut mieux associer les parties prenantes	85
Chapitre III L'accélération de la stratégie, avec une approche européenne, frugale et de confiance à approfondir.....	91
I - Des objectifs centrés sur la formation et l'attractivité des talents jusqu'à présent moins prises en considération	95
A - Une croissance des talents formés en IA en progression mais qui demeure en deçà des enjeux.....	95
B - Les ambitions contrariées d'excellence de la recherche et de la massification	104
C - Une attractivité envers les chercheurs et étudiants étrangers encore fragile	108
II - Une projection internationale et européenne de la stratégie à poursuivre	113
A - Des coopérations internationales à amplifier	113
B - Une nécessité de capitaliser davantage sur les financements européens	115
III - Un positionnement stratégique autour des enjeux sociétaux de confiance et de frugalité	116
A - Une déclinaison opérationnelle de l'IA de confiance par plusieurs mesures dans la seconde phase de la stratégie	119
B - L'IA frugale, principe fondateur qui doit irriguer de manière concrète l'ensemble des recherches financées par la seconde phase	120
Conclusion générale	125
Liste des abréviations	127
Annexes	131

Procédures et méthodes

Les rapports de la Cour des comptes sont réalisés par l'une des six chambres¹ thématiques que comprend la Cour ou par une formation associant plusieurs chambres et/ou plusieurs chambres régionales ou territoriales des comptes.

Trois principes fondamentaux gouvernent l'organisation et l'activité de la Cour ainsi que des chambres régionales et territoriales des comptes, donc aussi bien l'exécution de leurs contrôles et enquêtes que l'élaboration des rapports publics : l'indépendance, la contradiction et la collégialité.

L'indépendance institutionnelle des juridictions financières et l'indépendance statutaire de leurs membres garantissent que les contrôles effectués et les conclusions tirées le sont en toute liberté d'appréciation.

La contradiction implique que toutes les constatations et appréciations faites lors d'un contrôle ou d'une enquête, de même que toutes les observations et recommandations formulées ensuite, sont systématiquement soumises aux responsables des administrations ou organismes concernés ; elles ne peuvent être rendues définitives qu'après prise en compte des réponses reçues et, s'il y a lieu, après audition des responsables concernés.

La collégialité intervient pour conclure les principales étapes des procédures de contrôle et de publication. Tout contrôle ou enquête est confié à un ou plusieurs rapporteurs. Le rapport d'instruction, comme les projets ultérieurs d'observations et de recommandations, provisoires et définitives, sont examinés et délibérés de façon collégiale, par une formation comprenant au moins trois magistrats. L'un des magistrats assure le rôle de contre-rapporteur et veille à la qualité des contrôles.

Sauf pour les rapports réalisés à la demande du Parlement ou du Gouvernement, la publication d'un rapport est nécessairement précédée par la communication du projet de texte que la Cour se propose de publier aux ministres et aux responsables des organismes concernés, ainsi qu'aux autres personnes morales ou physiques directement intéressées. Leurs réponses sont présentées en annexe du rapport publié par la Cour.

¹ La Cour comprend aussi une chambre contentieuse, dont les arrêts sont rendus publics.

Le présent rapport d'évaluation est issu d'une enquête conduite sur le fondement de l'article L. 143-6 du code des juridictions financières qui permet à la Cour des comptes de mener des enquêtes thématiques et de l'article L. 111-13 du même code, selon lequel la Cour des comptes contribue à l'évaluation des politiques publiques.

Dans ses évaluations, la Cour s'attache notamment à apprécier les résultats de la politique publique examinée au regard à la fois des objectifs poursuivis (pertinence, cohérence et efficacité) et des moyens mis en œuvre (efficience). Ce rapport a été préparé par la première section « Enseignement supérieur et recherche » de la Troisième chambre.

La Cour s'est dotée d'un comité d'accompagnement composé de 14 membres (cf. annexe n° 1) issus de la diversité des acteurs du domaine de l'enseignement supérieur, de la recherche et de l'innovation en intelligence artificielle : université, organisme de recherche, association de doctorants, administrations, milieu socio-économique et trois chercheurs (un sociologue, un économiste et un mathématicien). Il s'est réuni quatre fois au cours de l'instruction : au lancement, lors des résultats préliminaires, lors de l'élaboration des recommandations et à la fin du travail d'instruction.

Pour mener cette évaluation, des méthodes qualitatives (entretiens semi-directifs, panels, ateliers participatifs, consultation) et quantitatives (scientométrie, économétrie, analyses sémantiques, analyses de réseaux) ont été mises en œuvre et irriguent l'analyse tout au long du rapport. Les méthodes les plus singulières sont détaillées dans les annexes. L'évaluation mobilise des données sur la période 2014-2021, en distinguant les années 2014 à 2017 et 2018 à 2021, soit quatre années avant et après la mise en œuvre de la stratégie nationale pour l'intelligence artificielle (SNIA). Des données plus anciennes ont parfois été prises en compte pour disposer d'une profondeur historique suffisante pour des analyses spécifiques. Les données relatives à 2022 ont été intégrées dans la mesure du possible.

Dans l'optique de retracer les dynamiques de recherche et de formation en IA sur la base d'un vocabulaire homogène et aussi consensuel que possible, la Cour a identifié et positionné, avec l'appui des membres du comité d'accompagnement, les mots clés de l'IA « de cœur », et a développé une méthode d'analyse sémantique (cf. annexe n° 5). Cette approche sémantique est fondée sur la recherche automatisée d'une centaine de mots clés caractéristiques de l'IA, en français et en anglais, au sein de grands ensembles de données relatives à la production scientifique, aux projets de recherche, aux thèses ou encore aux descriptifs des formations tant initiale que continue. La méthode a permis d'identifier les contenus relatifs à l'IA suivant une méthode harmonisée, explicable et reproductible. Le rapport propose, à date, une liste hiérarchisée de mots-clés en IA susceptible d'être utilisée dans le cadre de divers requêtes pour détecter automatiquement des contenus relatifs à ce domaine.

Pour le positionnement international, la Cour a utilisé les données issues de l'observatoire des politiques publiques sur l'IA² mené par l'Organisme de coopération et de développement économiques (OCDE). Pour l'analyse de la production scientifique, la Cour a utilisé les bases de données bibliométriques *InCites (Web of Science)* et l'archive ouverte pluridisciplinaire HAL. Pour les financements, la Cour a utilisé la base de données « projet » de l'Agence nationale de la recherche et E-Corda, la base de données des projets financés par la Commission européenne. Pour les thèses, la base ouverte thèses.fr et la base de l'Association nationale de recherche et technologie (ANRT) sur les conventions industrielles de formation par la recherche ont été mobilisées. Pour l'analyse de réseaux web, le logiciel Hyphe a été mis à disposition par SciencesPo Média Lab afin d'être en capacité d'analyser les liens entre sites web des différents acteurs de l'intelligence artificielle.

Une consultation de la communauté de recherche a également été menée par la Cour pour évaluer la perception de la première phase de la stratégie par la communauté, les conditions de travail des personnels d'enseignement et de recherche, l'attractivité de la France ou encore les conditions de valorisation des travaux scientifiques et éclairer la mise en œuvre de la seconde phase, notamment en matière de maturité de l'écosystème.

Cette consultation a été adressée aux chercheurs, enseignants-chercheurs, ingénieurs de recherche, doctorants, post-doctorants, et à toute autre personne identifiée comme travaillant dans le domaine de l'IA par le CNRS, Inria, le CEA et les universités (76 % d'entre elles ont répondu à la sollicitation de la Cour). Entre le 22 juin et le 29 août 2022, 772 réponses ont été reçues complètes et exploitables sur la base d'un envoi à 1 978 destinataires, soit un taux de réponses de 39 %. Cette consultation n'a pas valeur de sondage étant donné que la cohorte a été identifiée de manière hétérogène par les institutions et qu'il n'est pas possible de corriger des biais de réponses fondées sur le volontariat. Néanmoins, au vu du taux de retour, la consultation a été jugée suffisamment robuste pour être mentionnée dans le rapport par la Cour.

Un atelier participatif avec une trentaine de parties prenantes (administrations, organismes de recherche, universités, grandes écoles, *think-tanks*, monde économique) a eu lieu le 19 juillet 2022 pour partager les constats de la Cour, identifier des biais de perception ainsi que des manques dans le diagnostic ; opérationnaliser les projets de recommandations pour mesurer leur pertinence, leur efficacité ; et favoriser une meilleure appropriation par la communauté d'acteurs. Les résultats de cet atelier sont résumés en annexe n° 9.

² Observatoire OCDE des Politiques Publiques sur l'Intelligence Artificielle - OECD.AI

Les remarques formulées sur le rapport provisoire dans la phase contradictoire ont été prises en considération par la Cour qui a également auditionné le Président de France universités et à sa demande, le Président-directeur général d'Inria.

Le projet de rapport a été préparé, puis délibéré le 24 novembre 2022, par la troisième chambre, présidée par M. Meddah, président de chambre, et composée de M. Rousselot, président de section, MM. Bouvard, Feller, Guaino, Metzger, Miller, Montarnal, et Mme Delétang conseillers-maitres ainsi que, en tant que rapporteurs, Mmes Séguin, conseillère référendaire, et Roger-Vasselin, auditrice et M. Mousson, conseiller-maître, MM. Moslonka-Lefebvre et Roudaut, conseillers référendaires en service extraordinaire, M. Luscié, vérificateur et M. Netter, conseiller-expert, et, en tant que contre-rapporteur, M. Guibert, conseiller-maître.

Il a été examiné et approuvé le 5 janvier 2023 par le comité du rapport public et des programmes de la Cour des comptes, composé de M. Moscovici, Premier président, Mme Camby, rapporteure générale, M. Andréani, Mme Podeur, M. Charpy, Mme Démier, M. Bertucci, Mme Hamayon et M. Meddah, présidentes et présidents de chambre de la Cour, M. Michaut, M. Lejeune, M. Advielle, Mme Gaspari et Mme Renet, présidentes et présidents de chambre régionale des comptes, M. Gautier, procureur général, entendu en ses avis.

Les rapports publics de la Cour des comptes sont accessibles en ligne sur le site internet de la Cour et des chambres régionales et territoriales des comptes : www.ccomptes.fr.

Synthèse

L'intelligence artificielle (IA) est un concept ancien, qui apparaît dès les années 1950 dans les travaux notamment de l'informaticien britannique Alan Turing. S'il existe un débat sémantique important au sein de la communauté scientifique sur ce que recouvre l'IA et la délimitation de son périmètre, on peut la définir par son objet, qui est de reproduire l'intelligence humaine par l'utilisation de l'informatique et des mathématiques. L'IA s'est surtout développée à partir des années 1980, avec l'émergence de l'apprentissage automatique des algorithmes (*machine learning*). Dans les années 2000, la croissance des capacités de calcul et de l'accès aux données a favorisé le développement des techniques d'apprentissage profond (*deep learning*).

Les applications de l'IA sont aujourd'hui multiples, sources d'innovations et de gains de productivité dans de nombreux secteurs. Il en résulte une croissance constante des investissements économiques observée depuis les années 2010. Selon l'OCDE, les start-up spécialisées dans l'IA ont ainsi attiré près de 12 % du capital-investissement mondial au cours du premier semestre de 2018, par rapport à 3 % en 2011. Les publications de recherche ont connu une tendance similaire, avec plus de 1,2 million de publications en 2019 contre moins de 40 000 en 2010. À côté de ces opportunités, son essor est porteur d'enjeux, notamment éthiques, en particulier en matière de protection des droits des citoyens.

Par conséquent, l'IA est devenue une priorité croissante des autorités publiques. L'adoption, depuis 2017, par de nombreux États de plans nationaux pour encourager son développement en témoigne, et constitue une réponse à la forte compétition qui existe à l'échelle internationale pour accroître le niveau technologique des pays et attirer les meilleurs talents. En France, une stratégie nationale pour l'intelligence artificielle (SNIA) a été lancée en mars 2018, avec pour objectif de positionner la France comme l'un des acteurs majeurs en IA sur l'échiquier mondial. Initialement dotée de 1 527 M€ de financements publics pour la période 2018-2022, elle s'est déployée autour de cinq volets : 1) la recherche, 2) l'enseignement supérieur, 3) la transformation publique, 4) la diffusion dans l'économie, 5) la défense et la sécurité. En novembre 2021, une nouvelle phase de la SNIA dite d'« accélération » a été annoncée pour la période 2022-2025, avec comme

objectif de renforcer la compétitivité et l'attractivité de la France dans ce domaine. Cette nouvelle étape s'inscrit dans le prolongement des ambitions de la première phase de la stratégie, et les financements publics qui lui sont alloués correspondent *a priori* à un effort similaire à celui engagé pour la période 2018-2022. Elle a, par ailleurs, été élaborée en accord avec les priorités à l'échelle européenne.

Le présent rapport constitue une évaluation *in itinere* de la SNIA. Il porte sur les volets « recherche » et « enseignement supérieur », soit les principaux volets en matière de financement, respectivement d'un montant de 1 527 M€ et de 1 545 M€ dans la première et dans la seconde phase. Sur la période 2018-2022, 445 M€, soit près de 30 % des financements alloués à la stratégie, étaient en effet orientés vers la recherche, contre 134 M€, soit 8,7 %, dans la seconde phase. *A contrario*, les financements prévus pour la période 2022-2025 en faveur de la formation connaissent une nette hausse (50,2 % des crédits alloués, contre 8,4 % dans la phase précédente).

Les questions évaluatives ont été définies en concertation avec les parties prenantes de la SNIA et les autorités publiques chargées de sa mise en œuvre. Elles se sont réparties en quatre grandes interrogations :

- la stratégie nationale de recherche a-t-elle permis de renforcer la position de la France aux niveaux mondial et européen ? [cohérence, efficacité et efficience] ;
- la stratégie nationale de recherche a-t-elle accéléré la structuration de l'écosystème français en IA ? [pertinence et efficience] ;
- la stratégie nationale de recherche des pôles d'excellence est-elle efficace et efficiente ? [efficacité et efficience] ;
- la stratégie nationale de recherche a-t-elle amélioré la prise en compte des enjeux éthiques (IA frugale et de confiance) ? [pertinence, cohérence et efficacité].

Pour y répondre, un travail inédit d'analyse sémantique et d'exploitation de nombreuses bases de données reposant sur des méthodes statistiques et économétriques a été réalisé afin de quantifier et d'apprécier les résultats de la stratégie. Ce volet quantitatif a été complété par de nombreux entretiens semi directifs et focus groupes, ainsi que par une consultation auprès de chercheurs en IA et un atelier participatif avec des experts du domaine.

Une priorité donnée à la recherche qui a permis d'éviter un décrochage

En 2018, la France a été l'un des premiers pays à l'échelle mondiale à s'être doté d'un plan formalisé pour l'IA. Depuis, des stratégies nationales ou mesures spécifiques ont été élaborées par de nombreux États.

Dans un premier temps, la stratégie française a donné la priorité à la recherche en IA. Outre les 30 % des financements qui lui étaient alloués pour la période 2018-2022, la recherche a ainsi fait l'objet d'un plan spécifique, intitulé « stratégie nationale de recherche en intelligence artificielle » (SNRIA), et dont la coordination a été confiée à l'opérateur Institut national de recherche en informatique et en automatique (Inria). Le parangonnage international réalisé à partir des données de l'OCDE, et l'étude plus spécifique menée par la Cour sur les stratégies ou politiques publiques en matière d'IA de 10 pays³, montrent que l'identification de la recherche comme priorité stratégique est le choix le plus fréquent retenu par les États.

Depuis son lancement, la majorité des mesures prévues dans le cadre de la SNRIA ont été mises en œuvre. En formalisant une stratégie, les pouvoirs publics ont donné un signal politique fort sur l'importance de l'IA pour la recherche française. De fait, au-delà même des actions prévues par la stratégie, cette thématique est désormais prégnante dans toutes les réflexions au sein de plusieurs organismes de recherche. L'évaluation, et notamment les analyses économétriques sur les données mondiales, conforte le choix d'adopter un plan stratégique.

L'efficacité de la stratégie pour renforcer le positionnement de la France en IA, conformément à l'objectif qui lui était initialement fixé, n'est cependant pas avérée. Sur la période analysée, en nombre de publications en IA et sur un total de 47 pays comparés, la France conserve difficilement une place au 10^e rang à l'échelle mondiale et se maintient au 2^e rang au niveau européen. En raison du temps long de la recherche, les effets concrets de la stratégie sur la production scientifique ne sont néanmoins pas encore appréhendables de manière fiable.

S'agissant par ailleurs des financements mobilisés, ils doivent être davantage suivis afin de mesurer les effets des efforts financiers de cette stratégie en IA sur le rang scientifique de la France et sur la structuration de son organisation. Les moyens alloués à la stratégie ne retracent pas l'exhaustivité de l'investissement public en faveur de l'IA. Sur le volet recherche, 554,6 M€ ont *in fine* été engagés sur la période 2018-2022, sans que l'exécution réelle des crédits soit suivie de manière complète et synthétique.

³ États-Unis, Canada, Allemagne, Finlande, Italie, Pays-Bas, Royaume-Uni, Suisse, Israël, Japon.

Dans une perspective d'attractivité des talents, certains outils de financement gagneraient à être pérennisés. La grande majorité (+ de 80 %) des crédits a en effet été distribuée *via* des instruments financiers de courte durée, sur une logique d'appels à projets. Or, l'absence de visibilité sur la pérennité de ces guichets de financement est susceptible de créer des effets de rupture dans la formation des jeunes chercheurs (programme doctoral) ou la poursuite des programmes de recherche (chaire académique).

Un écosystème d'excellence dont la pérennité doit être garantie

L'orientation principale de la stratégie consiste en la constitution de pôles d'excellence en IA⁴, *via* la labellisation d'instituts interdisciplinaires en IA (3IA), la mise en place de chaires individuelles, puis l'identification de centres d'excellence en dehors des instituts 3IA. Il en ressort un renforcement de bassins géographiques déjà actifs en intelligence artificielle, une structuration d'un écosystème et un accroissement de la production scientifique des sites sans qu'il soit possible de démontrer l'impact de la stratégie dans cette dernière évolution.

Les synergies entre centres d'excellence doivent être renforcées, en s'inscrivant notamment dans une démarche plus systématique de valorisation réciproque de leurs travaux. Cela contribuerait à améliorer leur visibilité au niveau tant national qu'international, ainsi que l'image de marque de la France pour attirer des talents étrangers.

En parallèle, une clarification des missions des centres d'excellence 3IA et hors 3IA apparaît nécessaire. Les centres d'excellence hors 3IA ont en effet été identifiés après la labellisation des instituts 3IA, et, contrairement à ces derniers, en dehors de tout appel à projets compétitif impliquant un jury indépendant. Les attendus de la puissance publique à leur égard sont dès lors moins explicites, et leur modèle de développement moins contraint par les conditions de financement. Ils sont pourtant impliqués dans la seconde phase de la stratégie au même titre que les instituts 3IA. Cette réflexion doit s'accompagner d'une révision de la temporalité des financements alloués aux instituts labellisés (actuellement de quatre ans), qui repose sur une logique de trop court terme pour permettre des effets de leviers.

⁴ Les « pôles d'excellence » regroupent trois types d'entités : les quatre instituts interdisciplinaires en IA (instituts 3IA) identifiés au cours de la première phase *via* un AAP spécifique de l'ANR ; les 43 chaires individuelles portées par un chercheur identifié au travers d'un autre AAP spécifique de l'ANR ; les trois autres centres d'excellence, dits « hors 3IA », identifiés par le coordonnateur de la stratégie en 2021, sans faire l'objet d'un AAP et dont les membres peuvent détenir une chaire individuelle.

Le manque de visibilité dans le temps des financements associés est aussi identifié pour la formation des jeunes talents au travers des programmes doctoraux et des contrats Convention industrielle de formation par la recherche (CIFRE). Si la stratégie a donné un signal fort pour la favoriser, il importe désormais de pérenniser les financements nécessaires à cette dynamique.

L'évaluation montre que la SNRIA a permis de structurer les acteurs de la recherche en IA, alors que l'IA n'était pas identifiée comme discipline en tant que telle. Cette structuration doit cependant gagner encore en maturité : une comparaison des acteurs français et allemands au travers d'une analyse de réseaux montre que cette structuration s'organise dans notre pays encore majoritairement autour des principaux organismes de recherche, tandis que les sites universitaires et les centres pluridisciplinaires allemands sont plus efficaces dans la structuration de leur écosystème national.

Pour améliorer cette articulation entre acteurs, procéder à une adaptation de la gouvernance et du pilotage de la stratégie est nécessaire et urgent. La maquette de performance initialement élaborée n'a *in fine* pas été utilisée. Il n'y a pas de rapport annuel permettant de suivre la marche de la stratégie. Son pilotage, assuré par la coordination nationale, n'est pas en mesure de suivre les étapes de réalisation, de s'assurer des financements et des dépenses, ni de redresser les éventuelles difficultés. Il s'est affaibli avec le temps et il court le risque d'être inopérant. Aucune gouvernance spécifique n'a été mise en place pour le volet recherche, Inria étant alors le seul opérateur représenté au sein du comité de pilotage global de la stratégie. L'adaptation de la gouvernance pourrait être l'occasion de construire une vision commune et globale de l'action publique en IA, alors que les mesures décidées en ce domaine restent dispersées au sein de plusieurs politiques publiques.

La stratégie d'accélération, opportunité d'une approche frugale, de confiance et européenne renforcée

Si la stratégie nationale de recherche n'a pas encore permis de renforcer la position de la France au niveau mondial, le premier volet de la stratégie a permis d'éviter un décrochage scientifique depuis 2018. Le second volet est à présent déterminant pour enfin améliorer le positionnement de la France en IA dans la compétition mondiale. Cette « stratégie d'accélération en IA » est recentrée sur l'objectif de formation des talents en IA, priorité jusqu'ici peu prise en compte. Les financements totaux consacrés à cette priorité s'élèvent à 776 M€, dont 500 M€ pour un volet « excellence » et 276 M€ pour un volet « massification ». Le lancement de l'appel à manifestation d'intérêt « Compétences et Métiers d'Avenir » (AMI CMA) en 2022 constitue l'une des déclinaisons opérationnelles de cet effort.

Le nombre actuellement limité de formateurs publics de haut niveau pourrait contrarier les ambitions affichées, d'autant plus qu'il existe une tension entre investissement dans l'enseignement et excellence de la recherche. Le nombre de formateurs spécialisés dans l'enseignement supérieur public est pour le moment sous-dimensionné par rapport aux besoins de formation en IA, tant en formation initiale qu'en formation continue.

L'approche française gagnerait à s'intégrer encore davantage dans l'approche européenne. Divers programmes européens de soutien à la recherche ont pour objet de favoriser le développement de l'IA, dont « Horizon Europe » (près de 100 Md€ au total sur la période 2021-2027) et « pour une Europe numérique » (7,5 Md€ au total sur la même période). En 2018, les priorités de la stratégie française ont été élaborées en tenant compte du plan européen pour l'IA initié en 2018 et mis à jour en 2021. La phase d'accélération offre désormais l'opportunité de capitaliser davantage sur les efforts menés à l'échelle européenne.

La confiance⁵ et la frugalité⁶ sont ainsi deux des quatre axes structurants du programme et équipements prioritaires de recherche (PEPR) de 73 M€ adossé à la stratégie d'accélération. Il existe néanmoins un enjeu d'amélioration de la maturité de la communauté scientifique sur ces questions. La consultation des chercheurs en IA opérée par la Cour montre que ces thématiques sont actuellement peu prises en compte dans les travaux de recherche.

Ce sujet se pose avec une acuité particulière concernant la notion de frugalité de l'IA, avec une tension potentielle entre l'économie des ressources et la performance. La frugalité devrait être mieux intégrée aux appels à projets, *via* par exemple l'élaboration d'une charte ou d'un guide de bonnes pratiques.

Sur le volet recherche, les priorités sont recentrées dans cette seconde phase sur l'attractivité des talents et la prise en compte d'enjeux sociétaux, tels que la confiance dans l'IA et la frugalité dans son utilisation. Cette dernière évolution témoigne d'une réorientation du volet recherche, davantage tourné vers la recherche appliquée pour tenir compte de l'essor de l'IA industrielle.

⁵ Une intelligence artificielle de confiance se caractérise par son interprétabilité, son explicabilité, sa transparence et une identité « responsable ».

⁶ Une intelligence artificielle frugale est durable et respectueuse de l'environnement en cherchant à minimiser sa consommation d'énergie et de ressources.

Récapitulatif des recommandations

1. Traduire la politique publique sur l'intelligence artificielle dans un document budgétaire de synthèse qui permettra de la suivre et d'en mesurer les effets (*MEFSIN*).
2. Préciser les missions respectives des centres d'excellence 3IA et hors 3IA, et clarifier en conséquence les financements pluriannuels qui leur sont alloués (*MESR, SGPI*).
3. Établir de manière partagée les objectifs et les indicateurs prioritaires de la politique publique en IA, en lien avec la stratégie européenne (*MEFSIN, SGPI*).
4. Créer un comité scientifique et de pilotage auprès d'Inria, co-présidé par France Universités, pour suivre de manière concertée la mise en œuvre de la stratégie et définir les futures orientations stratégiques (*MESR, Inria*).
5. Réaliser une cartographie harmonisée et actualisée des formations en IA à valoriser au travers d'un label commun pour les rendre visibles et accompagner leur massification (*MESR*).
6. Prévoir les besoins en enseignants du secondaire, en enseignants-chercheurs et en chercheurs formés à l'usage de l'IA et établir des plans de formation en adéquation (*MESR*).
7. Élaborer une charte et un catalogue de bonnes pratiques visant à définir et suivre l'impact environnemental de la recherche en IA, et à favoriser le développement d'une IA responsable (*SGPI, MESR*).

Glossaire

- *Cloud* : le *cloud* n'est pas une entité physique mais un vaste réseau de serveurs distants, reliés entre eux, et destinés à fonctionner comme un écosystème unique. Ces serveurs sont conçus pour stocker et gérer des données accessibles à distance, exécuter des applications, ou fournir du contenu ou des services (vidéos diffusées en continu, courrier web, logiciels bureautiques de productivité et autres réseaux sociaux).
- *Deep learning* : méthode d'apprentissage profond de l'intelligence artificielle issue du *machine learning*. Cette méthode d'apprentissage se fonde sur des réseaux de neurones artificiels et peut traiter des données complexes dites « en très grande dimension » (son, texte, image, etc.).
- GPU : unité de traitement graphique (ou *Graphics Processing Unit* en anglais) ou puce informatique disposée sur la carte graphique de l'ordinateur. Ce type d'unité de traitement est plus adapté que les processeurs « classiques » (CPU pour *Central Processing Unit*) pour mettre en œuvre les calculs associés aux algorithmes de *deep learning*.
- Centres d'excellence : terme générique désignant les instituts dits 3IA (Paris, Grenoble, Nice et Toulouse) et les instituts non labellisés 3IA dans le cadre de la stratégie nationale (DATAIA, SCAI et Hi! PARIS).
- Intelligence artificielle (IA ou AI en anglais) : l'intelligence artificielle regroupe l'ensemble des outils, techniques et technologiques, fondés sur les algorithmes permettant d'accomplir une tâche, de formuler une proposition, de prendre des décisions, dans un système multiparamétrique et dans une dynamique d'apprentissage.
- Cœur de l'IA : activités de recherche sur les thèmes centraux de l'IA tels que l'algorithmique pour l'IA (apprentissage machine, évolution artificielle, optimisation spécifique pour l'IA, etc.) ou le raisonnement symbolique.
- Intégration de l'IA : activités de recherche sur des thèmes intégrateurs (vision artificielle, robotique, traitement des langues naturelles, systèmes multi-agents, interfaces homme machine, science des données, architectures et composants matériels pour l'IA, etc.).

- Applications de l'IA : activités de recherche interdisciplinaire avec des spécialistes des domaines d'applications de l'IA, en particulier ceux mentionnés par la stratégie nationale (sécurité et défense ; transport et mobilité ; santé ; environnement).
- Intelligence artificielle forte : intelligence artificielle qui permettrait à une machine de répondre au test de Turing⁷.
- Intelligence artificielle d'usage : ou intelligence artificielle dite « faible » qui permet aux algorithmes de trouver une application ou un usage pratique.
- Intelligence artificielle symbolique : méthode qui vise à formaliser une intelligence artificielle selon des méthodes d'apprentissage analogues à l'esprit humain. L'IA symbolique s'attache à représenter la connaissance humaine à travers des symboles explicites ainsi qu'à travers des règles qui rendent possible la manipulation de ces symboles.
- Intelligence artificielle décentralisée et embarquée : si le développement de l'intelligence artificielle reposait jusqu'à présent sur des briques logicielles installées sur des réseaux et des serveurs informatiques centralisés (*cloud*), la multiplication des objets « connectés » impose une décentralisation. La nouvelle révolution de l'intelligence artificielle passe dorénavant par le développement de composants électroniques dédiés qui « décentralisent » et « embarquent » l'intelligence artificielle et le traitement algorithmique au sein de ces composants.
- Intelligence artificielle de confiance : intelligence artificielle qui se caractérise par son interprétabilité, son explicabilité, sa transparence et une identité « responsable ». Au-delà des dimensions éthiques, juridiques et environnementales, l'intelligence artificielle de confiance doit également garantir une robustesse et une fiabilité technologiques de son fonctionnement.
- Intelligence artificielle frugale ou IA sobre : intelligence artificielle durable et respectueuse de l'environnement qui cherche à minimiser sa consommation d'énergie et de ressources, en particulier les données qui peuvent être rares ou coûteuses à obtenir dans certains domaines.

⁷ En octobre 1950, Alan Turing publie un article intitulé « l'ordinateur et l'intelligence » qui jette les bases de l'intelligence artificielle en présentant un jeu de l'imitation dans lequel il cherche à montrer comment faire passer la machine pour un être humain.

- *Machine learning* : méthode d'apprentissage automatique par les machines sans qu'elles aient été préalablement programmées à cet effet. Le *machine learning* est rendu possible par l'utilisation de données structurées (valeurs numériques dans des tableaux, etc.).
- Programmes et équipements prioritaires de recherche (PEPR) : est une action du programme 424 – *Financement des investissements stratégiques* de la mission Investir dans la France de 2030. Ces PEPR visent à construire ou consolider un *leadership* français dans des domaines scientifiques liés à une transformation technologique, économique, sociétale, sanitaire ou environnementale et qui sont considérés comme prioritaires au niveau national ou européen. L'État désignent les pilotes associés, et prévoit le financement alloué. Le ou les pilotes scientifiques décrivent concrètement leur programme dans un document de cadrage.
- Supercalculateur : machine conçue pour obtenir les plus hautes performances possibles avec les techniques connues au moment de sa conception, en particulier s'agissant des capacités de calcul. Le supercalculateur dénommé Jean Zay en hommage au ministre de l'éducation, de juin 1936 à septembre 1939, est un outil phare de la mise en œuvre de la stratégie nationale en IA.

Introduction

L'intelligence artificielle (IA), qui ambitionne de reproduire l'intelligence humaine par l'utilisation de l'informatique et des mathématiques, a émergé dès les années 1950, sous l'influence notamment des travaux d'Alan Turing⁸. Elle s'est ensuite développée par cycles d'innovation successifs, au gré de l'augmentation des capacités de calcul informatique et des percées scientifiques dans ce domaine. L'apprentissage automatique (*Machine learning*) apparaît dans les années 1980, avec l'essor des algorithmes « apprenants » dont la majorité est encore utilisée aujourd'hui. À partir des années 2000, la croissance de la puissance de calcul tout comme l'accès aux données permet le développement des techniques d'apprentissage profond (*deep learning*) et favorise des avancées importantes, par exemple dans la reconnaissance de textes et d'images, la traduction, ou encore l'analyse documentaire, la santé ou les transports.

Sur le plan sémantique, il existe un débat important au sein de la communauté sur la définition de l'IA et son périmètre⁹. En l'absence de consensus, les institutions ont adapté des définitions opérationnelles de l'IA. D'après la définition publiée au Journal officiel de la République française du 9 décembre 2018¹⁰, l'IA est « *un champ interdisciplinaire théorique et pratique qui a pour objet la compréhension des mécanismes de la cognition et de la réflexion et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à l'activité humaine* ». L'OCDE définit un système d'intelligence artificielle comme un système fonctionnant grâce à une machine, et capable d'influencer son environnement en produisant des résultats tels que des prédictions, des classifications ou des regroupements automatiques, des recommandations ou des décisions, pour répondre à un ensemble donné d'objectifs.

⁸ Dans l'article intitulé « L'ordinateur et l'intelligence » publié en 1950, l'auteur présente le « test de Turing », qui permet de vérifier la capacité d'une machine à se faire passer pour un humain, et pose les premiers concepts associés à la notion d'intelligence artificielle.

⁹ Identifier les contenus qui relèvent de l'IA suscite de vifs débats entre les experts du domaine, notamment entre les partisans des écoles « données / apprentissage » (dont la communauté des « connexionnistes » qui s'intéressent plus particulièrement aux réseaux de neurones profonds) de l'IA, d'une part, et l'école « symbolique » de l'IA (école historique qui a précédé l'IA par apprentissage), d'autre part, avec les partisans de l'IA « hybride » qui cherchent à combiner ces deux grandes approches du domaine.

¹⁰ JORF n° 0285 du 9 décembre 2018, texte n° 58.

L'évaluation de la politique publique française en matière d'IA conduite par la Cour n'a pas vocation à se prononcer sur ce débat scientifique et sémantique. Pour le présent rapport, elle a pris en considération l'ensemble des actions et acteurs s'identifiant à l'écosystème IA entendu au sens le plus large. Il regroupe l'IA dite « de cœur », qui correspond au développement des fondements de l'IA, l'IA dite « intégrative », qui regroupe les recherches sur des thèmes intégrateurs du type vision ou traitement naturel du langage, et l'IA « applicative », qui utilise l'IA au service de problématiques scientifiques telles que définies dans les appels à projets ou manifestation d'intérêts de l'Agence nationale de la recherche (ANR) dans le cadre de la stratégie nationale de recherche en intelligence artificielle.

Au cours de la dernière décennie, les investissements économiques dans l'IA ont progressé à un rythme soutenu. Les publications de recherche ont suivi la même tendance avec plus 1,2 million de publications en 2019 contre moins de 40 000 en 2010. L'année 2016 marque un très net décollage des publications mondiales en IA¹¹. Associée volontiers à la « 4^e révolution industrielle » ou « industrie 4.0 », l'intelligence artificielle est désormais une technologie d'usage général¹². Elle peut irriguer l'ensemble des domaines socio-économiques et environnementaux en améliorant la productivité et en favorisant l'innovation. L'IA est susceptible d'induire des mutations profondes sur le marché du travail, avec la transformation importante de secteurs économiques¹³. Son développement est porteur de nombreux enjeux en matière de protection des données personnelles¹⁴ et de préservation des droits fondamentaux des citoyens¹⁵.

L'intelligence artificielle est en conséquence devenue une priorité croissante des autorités publiques. L'adoption depuis 2017 par de nombreux États de plans nationaux pour encourager son développement en témoigne. La stratégie française a été formellement lancée le 28 mars 2018, à la suite, notamment, des travaux initiés par France Stratégie en 2017 et du rapport *Donner un sens à l'intelligence artificielle* de M. Cédric Villani de 2018, et en articulation avec les priorités définies au niveau européen¹⁶. Établie dans un premier temps pour une première phase de cinq ans, jusqu'en 2022, elle

¹¹ AI Index Report- source Scopus, Université de Stanford, 2021.

¹² Terme issue de la terminologie adoptée par les économistes de la croissance, T. Bresnahan & M. Trajtenberg, « *General purpose technologies 'engines of growth' ?* » *Journal of econometrics*, 1995.

¹³ France stratégie, *Intelligence artificielle et travail*, 2018.

¹⁴ CNIL, *Intelligence artificielle*.

¹⁵ Commission nationale consultative pour les droits de l'homme, *Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux (A - 2022 - 6)*, 7 avril 2022.

¹⁶ La Commission européenne a publié en avril 2018 une stratégie européenne pour l'IA et présenté un plan coordonné dans le domaine de l'IA en décembre de la même année.

affichait comme objectif de positionner la France comme l'un des acteurs majeurs en intelligence artificielle sur l'échiquier mondial, et poursuivait trois priorités stratégiques identifiées : développer une attractivité de talents (i), diffuser l'intelligence artificielle et les sciences des données dans l'économie et dans l'administration (ii) et promouvoir un modèle éthique équilibré entre innovation et protection des droits fondamentaux (iii). Bien que particulièrement orientée sur la recherche, cette première étape s'est déployée autour de cinq volets : 1) la recherche, 2) l'enseignement supérieur, 3) la transformation publique, 4) la diffusion dans l'économie ainsi que 5) la défense et la sécurité. Elle mettait l'accent sur quatre secteurs : la défense, la santé, les transports et l'environnement.

La présente évaluation se concentre sur les deux premiers volets¹⁷, la recherche et l'enseignement supérieur, définis par la réunion interministérielle (RIM) du 27 mars 2018¹⁸ et par le document de programmation du volet recherche établi en novembre 2018. Le présent rapport de la Cour débouche sur des orientations à retenir, le cas échéant, dans la mise en œuvre de la stratégie d'accélération annoncée à la fin de l'année 2021, dans le cadre du 4^e programme d'investissements d'avenir (PIA) et de France 2030. Définie pour une période allant jusqu'en 2025, cette nouvelle étape affiche comme objectif la transformation du potentiel scientifique en succès économiques et s'inscrit également dans le programme européen de recherche Horizon Europe pour des financements obtenus jusqu'en 2027.

¹⁷ Le Conseil d'État a présenté le 30 août 2022 une étude intitulée *Intelligence artificielle et action publique : construire la confiance, servir la performance*. Il y plaide pour une politique de déploiement de l'intelligence artificielle, résolument volontariste au service de l'intérêt général et de la performance publique. La France doit mettre en place un cadre réglementaire notamment au niveau européen à partir de lignes directrices pragmatiques. L'intelligence publique de confiance doit ainsi reposer sur sept principes : primauté humaine, performance, équité et non-discrimination, transparence, sûreté, soutenabilité environnementale et autonomie stratégique. Quelques pistes sont proposées : former les dirigeants, recruter des experts, assouplir le cadre juridique du partage des données entre administrations, renforcer Etalab et le coordinateur national et transformer profondément la CNIL.

¹⁸ Les « Grands défis », visant à répondre à des enjeux sociétaux mais nécessitant la levée de barrières technologiques que le Gouvernement a choisi de soutenir, n'ont pas été évalués. Deux relèvent cependant explicitement de l'intelligence artificielle : le Grand défi « Santé : améliorer les diagnostics médicaux à l'aide de l'intelligence artificielle » doté de 30 M€ d'investissements publics et le Grand défi « Sécuriser, certifier et fiabiliser les systèmes fondés sur l'intelligence artificielle » avec le programme Confiance.ai, associant partenaires académiques et industriels français qui a permis le développement d'une plateforme d'outils logiciels pour la conception, la qualification et le déploiement d'une IA de confiance au service de l'Industrie. Ce programme est doté de 30 M€ d'investissements publics et 15 M€ d'investissements privés.

Si une grande partie de la recherche et développement en IA est portée par le secteur privé, tant par des grands groupes que par des start-up, petites et moyennes entreprises (PME) ou entreprises de taille intermédiaire (ETI), il n'a pas été possible dans les délais de l'évaluation de déterminer la part de cet investissement dans l'écosystème et ses effets de levier. De même, la défense¹⁹ ayant fait l'objet d'une feuille de route disposant de financements spécifiques, la recherche duale (civile et militaire) n'est que partiellement prise en compte dans ces travaux. Ces éléments ne sont pas mesurés dans le présent rapport, en dehors des effets des partenariats entre acteurs privés et académiques ou de la valorisation de la recherche *via* notamment les centres d'excellence, les instituts de recherche technologique (IRT) ou encore la création dans ce domaine de start-up « *deep tech* ».

Dans son enquête, la Cour a cherché à répondre à quatre grandes questions évaluatives :

- La stratégie nationale de recherche a-t-elle permis de renforcer la position de la France aux niveaux mondial et européen ? (chapitre I) ;
- La stratégie nationale de recherche a-t-elle accéléré la structuration de l'écosystème français en IA ? (chapitre II) ;
- La stratégie nationale de recherche des pôles d'excellence est-elle efficace et efficiente ? (chapitre II) ;
- La stratégie nationale de recherche a-t-elle amélioré la prise en compte des enjeux éthiques (IA frugale et de confiance) ? (chapitre III).

¹⁹ [Feuille de route du ministère des Armées.pdf \(vie-publique.fr\)](#).

Chapitre I

Une priorité donnée à la recherche qui n'a pas encore produit les résultats attendus

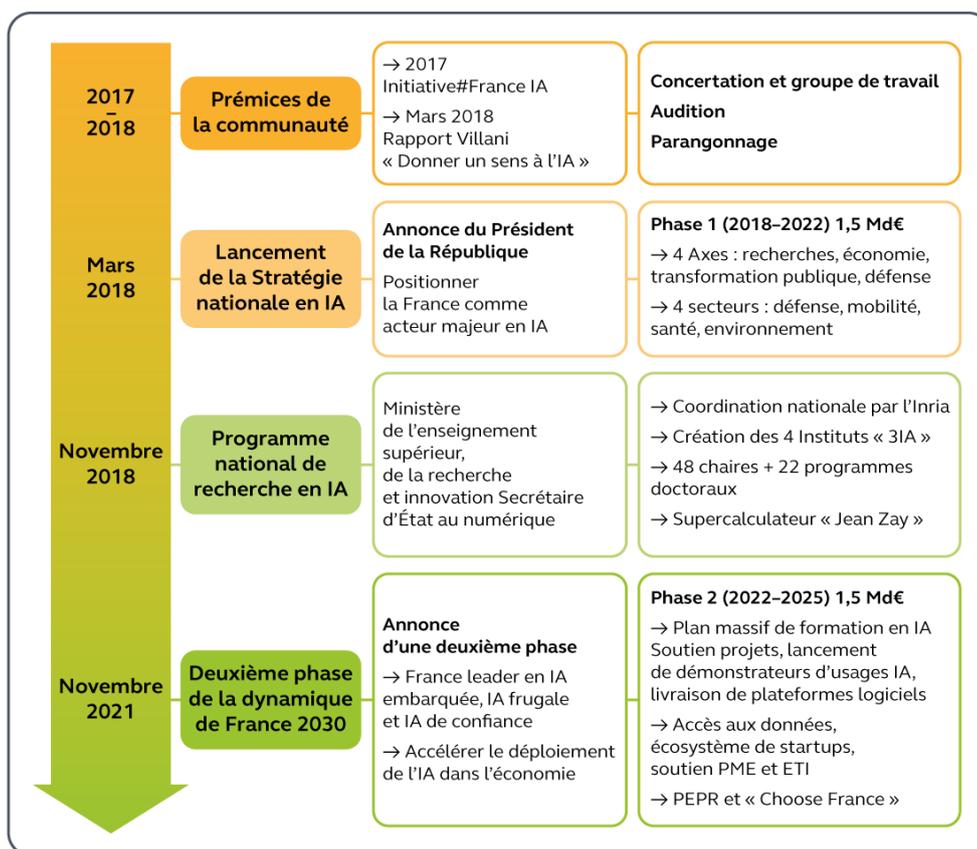
L'intelligence artificielle est à la fois un enjeu de souveraineté nationale et européenne et une technologie déterminante pour la compétitivité économique et la transformation de nombreux secteurs. Il existe une compétition entre les pays autour de cette technologie de l'amont à travers la recherche à l'aval concernant les applications. Chaque pays cherche donc à se positionner comme pionnier et moteur de cette technologie.

Ce chapitre vise à répondre à la question évaluative : « La stratégie nationale de recherche a-t-elle permis de renforcer la position de la France aux niveaux mondial et européen ? ».

I - Un rôle prépondérant accordé à la recherche en intelligence artificielle donnant lieu à un plan spécifique

A - La France, un des premiers pays à s'être doté d'une stratégie spécifique en intelligence artificielle

Schéma n° 1 : les étapes d'élaboration de la stratégie nationale en IA



Source : Cour des comptes

En France, la volonté de bâtir une stratégie globale pour l'intelligence artificielle s'est inscrite dans une démarche plus large de mobilisation de toute la communauté française en IA (schéma n° 1). L'initiative #France IA a été prise dès 2017 par les secrétaires d'État chargés de l'industrie, du numérique et de l'innovation, et de l'enseignement supérieur et de la recherche. Elle a donné lieu à une large consultation des parties prenantes, avec l'objectif de structurer un plan national permettant d'encourager l'innovation permise par l'intelligence artificielle (i), d'anticiper les impacts en matière économique et sociale (ii), et de préparer les évolutions futures en matière de formation et de recherche (iii). En 2017 également, les rapports de l'office parlementaire des choix scientifiques et techniques (OPECST)²⁰ et de la Commission nationale de l'informatique et des libertés (CNIL)²¹ ont aussi contribué à nourrir la réflexion de la communauté.

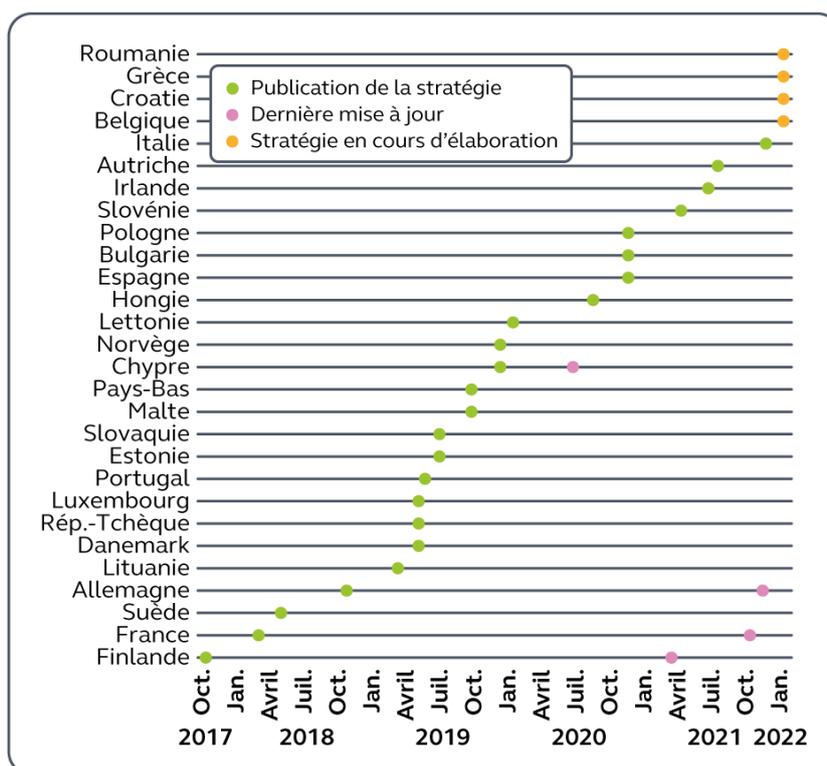
Cette dynamique originelle a été ensuite amplifiée, avec le lancement de la mission confiée par le Premier ministre au député Cédric Villani sur la mise en œuvre d'une stratégie française européenne en intelligence artificielle. Le rapport *Donner un sens à l'intelligence artificielle, pour une stratégie nationale et européenne*²² (2018) a été élaboré à partir de l'audition de 420 experts, d'une consultation publique à travers le Parlement et les citoyens pour plus de 2 000 contributions et d'une étude comparative auprès de 15 pays. Il constitue à ce jour le principal socle réflexif de la stratégie française exposée par le président de la République à l'occasion d'un discours de lancement tenu le 29 mars 2018 (#AI for humanity).

²⁰ M. Claude de Ganay et Mme Dominique Guillot, Pour une intelligence artificielle maîtrisée, utile et démystifiée, OPECST, Assemblée nationale et Sénat, 29 mars 2017.

²¹ CNIL, *Comment permettre à l'homme de garder la main ?* 2017.

²² Auteurs : Cédric Villani, Marc Schoenauer, Yann Bonnet et Charly Berthet. La mission se composait également d'Anne-Charlotte Cornut, François Levin et Bertrand Rondepierre.

Graphique n° 1 : panorama des stratégies nationales en IA des États-membres de l'Union européenne et de la Norvège



Source : Cour des comptes d'après les données du rapport du centre commun de recherche de l'Union européenne et de l'OCDE, *AI Watch. National strategies on Artificial Intelligence : A European perspective* (2022)

L'élaboration de la stratégie française est ainsi intervenue alors que le développement de l'IA connaissait une accélération d'ampleur mondiale²³, dans un contexte plus large de compétition très sélective entre les pays pour accroître leur niveau technologique et attirer les meilleurs talents. Le parangonnage montre que la mise en place d'un plan national stratégique structuré en faveur de l'intelligence artificielle n'est pas une originalité française : depuis 2016, de nombreux États se sont dotés de stratégies spécifiques consacrées à l'IA (cf. graphique n° 1 présentant la progression des stratégies des pays membres de

²³ Selon le rapport #France IA, plus de 1 600 start-up spécialisées en intelligence artificielle étaient recensées dans le monde par la plateforme d'intelligence économique CBInsights en 2016. Entre 2012 et 2017, les investissements dans les start-up spécialisées en IA n'ont cessé de croître, passant de 415 M\$ à 5 Md\$.

l'Union européenne et de la Norvège). Lorsqu'il n'existe pas de plan spécifique, les enjeux liés à l'IA sont le plus souvent pris en compte dans le cadre de politiques publiques ou de programmes plus généraux²⁴.

La Cour a examiné plus spécifiquement les stratégies ou politiques publiques en matière d'IA menées au sein de 10 pays. Les pays ont été choisis en fonction de la disponibilité et de la fiabilité des sources, ainsi que d'un socle minimum de convergence sur les règles en matière d'usage de l'IA. Au total, l'étude a porté sur deux pays du continent Nord-américain (États-Unis, Canada), six pays européens, dont quatre de l'Union européenne (Allemagne, Finlande, Italie, Pays-Bas), deux proches de l'Union (Royaume-Uni, Suisse) et deux du reste du monde (Israël, Japon).

Il ressort de cette comparaison qu'il existe des ambitions communes poursuivies par les autorités publiques en matière d'IA, parmi lesquelles : le renforcement du secteur de la recherche et de son attractivité (*via* notamment la création de centres pluridisciplinaires et le développement de chaires), le transfert des résultats de la recherche vers le monde économique, le développement des capacités de calcul et l'élaboration de règles d'usage et de normes en matière d'IA. Ces priorités se retrouvent dans les trois axes stratégiques définis au lancement de la stratégie française : le développement d'une attractivité de talents (i), la diffusion de l'intelligence artificielle et de la science des données dans l'économie et dans l'administration (ii) et la promotion d'un modèle éthique équilibré entre innovation et protection des droits fondamentaux (iii). Les actions de mise en œuvre ont été élaborées dans la perspective de consolider les atouts existants de la France alors identifiés, à savoir : son école de mathématiques, l'excellence de ses laboratoires de recherche²⁵, des grands groupes performants dans le secteur, un écosystème dynamique²⁶ et l'accès à des bases de données de qualité.

B - Des financements du plan national portant principalement sur la recherche

Parmi les cinq volets de la première phase de la stratégie (recherche, enseignement supérieur, transformation de l'action publique, diffusion de l'IA dans l'économie, défense et sécurité), les efforts se sont principalement concentrés sur la recherche, comme en témoigne l'importance des financements étatiques lui étant alloués (près du tiers) au sein du budget prévisionnel de la stratégie (cf. tableau n° 1).

²⁴ C'est le cas notamment de la Suisse : si aucune stratégie spécifique à l'intelligence artificielle n'a été mise en place, cet enjeu est depuis 2018 l'une des thématiques centrales de la stratégie « Suisse numérique » 2021-2024.

²⁵ Selon l'étude du cabinet Roland Berger, *Joining the dots – A map of Europe's AI ecosystem* publiée en 2018, la France se distingue par son nombre important de laboratoires de recherche, s'élevant à 82, soit 21 % du total européen.

²⁶ Selon la cartographie des start-up en IA en France élaborée par Bpifrance en juillet 2019, l'écosystème des start-up françaises en IA représentent en avril 2019 plus de 650 start-up et 2 Md€ investis.

Tableau n° 1 : ventilation du budget étatique initialement prévu pour la stratégie en IA pour la période 2018-2022

<i>Axes de la stratégie nationale en IA</i>	Financements étatiques prévisionnels (en M€)
<i>Recherche</i>	445
<i>Enseignement supérieur</i>	128
<i>Transformation de l'action publique</i>	154
<i>Économie</i>	390
<i>Défense et sécurité</i>	410
Total	1 527

Source : Traitement Cour des comptes d'après données du coordonnateur national de la stratégie IA

Comme le souligne l'OCDE à partir des données de son observatoire des politiques relatives à l'IA de plus de 60 pays et territoires, la priorité donnée dans un premier temps à la recherche en intelligence artificielle est une constante des stratégies nationales²⁷. En la matière, toutes poursuivent un triple objectif d'accroissement du nombre de chercheurs et de diplômés qualifiés en IA (i), de renforcement des capacités nationales de recherche en IA (ii), et d'application des résultats de la recherche en IA dans les secteurs public et privé (iii).

En France, cette priorité s'est matérialisée par la mise en place au sein de la stratégie globale en IA d'un plan spécifique pour la recherche. Présentée en novembre 2018, cette « stratégie nationale de recherche en intelligence artificielle » (SNRIA) – également dénommée « programme national de recherche en intelligence artificielle » (PNRIA) – s'est déployée à partir de 2019 sous le pilotage d'Inria²⁸. L'opérateur s'en est vu confier la coordination par une lettre de mission datée de juin 2019, soit six mois après le lancement de la SNRIA, et a mis en place une structure spécifique pilotée par un directeur de recherche (cf. annexe n° 2). L'élaboration et la mise en œuvre de ce volet recherche se sont appuyées sur le livre blanc élaboré par l'opérateur en 2016²⁹ et mis à jour en 2021³⁰, ainsi que le rapport de l'alliance Allistène consacrée aux sciences et technologies pour le numérique sur les infrastructures de recherche pour l'IA³¹.

²⁷ OCDE.IA 2021, données des pays sur leur politique nationale en IA-instruments de la politique nationale en IA-instruments de financements- fonds alloués à la recherche publique.

²⁸ Institut national de recherche en sciences et technologies du numérique.

²⁹ Bertrand Braunschweig, Livre blanc *Intelligence Artificielle : Les défis actuels et l'action d'Inria*, Inria, 2016.

³⁰ Bertrand Braunschweig, Livre blanc *Artificial Intelligence : Current challenges and Inria's engagement*, Inria, deuxième édition, 2021.

³¹ Jamal Atif et Bertrand Braunschweig, Rapport *Infrastructure pour une recherche pour l'Intelligence artificielle*, Allistène – Alliance des sciences et des technologies pour le numérique, 2018.

Afin de renforcer les capacités de recherche en IA, six axes structurants déclinés en plusieurs actions de mise en œuvre ont été identifiés dans le programme national de recherche. La pertinence, l'efficacité et l'efficience de ces axes pour atteindre les objectifs fixés sont évalués au chapitre 2. Le tableau n° 2 présente un premier état des lieux de leur réalisation.

Tableau n° 2 : axes structurants de la stratégie nationale de recherche en intelligence artificielle pour la période 2018-2022

<i>Axes structurants</i>	<i>Principales actions mises en œuvre</i>
<i>1) Mettre en place un réseau d'institut d'excellence en IA.</i>	- Lancement, développement de quatre instituts pluridisciplinaires en intelligence artificielle (réseau des « 3IA ») abritant 149 chaires.
<i>2) Lancer un programme d'attractivité et de soutien aux talents</i>	- Création de 43 chaires hors des instituts « 3IA » ; - 22 programmes doctoraux pour 374 contrats doctoraux supplémentaires, et 50 CIFRE en 2019 ; - Attirer les talents avec une expérimentation du CNRS sur le programme Choose France – Rising talents.
<i>3) Dynamiser la recherche en IA à l'ANR</i>	- Développement de projets de recherche, principalement collaboratifs, notamment au travers de l'appel à projets générique annuel ; - 2018 : 44 projets, dont 25 du comité d'évaluation scientifique (CES)23 dédiée à l'IA et 19 dans d'autres CES ; - 2019 : 43 projets, dont 30 du CES23 et 13 dans d'autres CES ; - 2020 : 44 projets dont 28 du CES23 et 16 dans d'autres CES ; - 2021 : 51 projets dont 40 du CES23 et 11 dans d'autres CES. - Challenge IA : Biodiversité avec deux projets sélectionnés.
<i>4) Renforcer les moyens de calcul</i>	- Faciliter l'accès aux données et à la puissance de calcul ; - Mise en place d'un réseau d'ingénieur (CNRS, CEA et Inria) pour faciliter l'utilisation des moyens de calcul relevant des centres dits « tiers » 0, 1 ou 2 – respectivement à usage européen, national ou régional – visant à favoriser les développements et l'utilisation de techniques de l'IA ; - Mise en place du supercalculateur convergé calcul intensif et IA appelé « Jean-Zay » : capacité portée à 36,8 pétaflops en 2022.
<i>5) Renforcer la recherche partenariale</i>	- Programme Labcom de l'ANR : sélection de quatre Labcoms en 2020 et de deux en 2021 ; - Fléchage de financements lors des appels à projet annuels ; - Mise en place de l'offre de service Engage IA par l'association French Institutes of technology.
<i>6) Renforcer les coopérations bilatérales, européennes et internationales</i>	- Renforcement de la collaboration franco-allemande : - Lancement de deux appels à projets bilatéraux en 2021 et en 2022 ; - Lancement du Partenariat Mondiale pour l'IA (25 pays membres) pour promouvoir un usage responsable de l'IA ; - Soutien au déploiement d'une stratégie ambitieuse de l'Union européenne pour l'intelligence artificielle.

Source : Cour des comptes d'après dossier de presse et éléments transmis par la coordination nationale et l'ANR.

C - L'IA, une priorité qui se diffuse au-delà du strict cadre de la stratégie

En formalisant une stratégie, la France a donné un signal politique sur l'importance de l'IA pour la recherche française. La thématique est devenue centrale dans plusieurs organismes de recherche, indépendamment et au-delà du strict périmètre de la stratégie.

L'IA est ainsi un axe important des contrats d'objectifs et de performance (COP) 2019-2023 d'Inria, du Centre national de la recherche scientifique (CNRS)³² et du Commissariat à l'énergie atomique et aux énergies alternatives (CEA).

Le nombre d'équipes projet en IA d'Inria est passé de 30 à 58 entre 2016 et 2021 (sur respectivement 223 et 210 équipe-projets), et un directeur délégué a été nommé.

Au CNRS, la recherche en IA est plus particulièrement pilotée par l'Institut des sciences de l'information et de leurs interactions (INS2I) avec l'appui de l'Institut national des sciences mathématiques et de leurs interactions (INSMI). La mission pour les initiatives transverses et interdisciplinaires (MITI) du CNRS lance par ailleurs chaque année de nombreux appels à projets (AAP), dont un certain nombre ont pu financer des projets en IA à l'instar de l'AAP spécifique « Enjeux scientifiques et sociaux de l'IA » (2020-2021), accompagné de huit thèses (du dispositif PRIME – projets de recherche interdisciplinaires multi équipes). Après une période de préfiguration de deux ans, le CNRS a également lancé en 2018 un groupement de recherche (GDR) « Aspects Formels et Algorithmiques de l'Intelligence Artificielle³³ ».

³² Dans son COP 2019-2023, le CNRS a identifié six grands défis sociétaux pour lesquels il a l'ambition d'apporter, en lien avec ses partenaires, une contribution substantielle dans les prochaines années *via* une mobilisation coordonnée des dix instituts : 1. Changement climatique 2. Inégalités éducatives 3. Intelligence artificielle 4. Santé et environnement 5. Territoires du futur 6. Transition énergétique.

³³ D'autres GDR contribuent à cette thématique. Le GDR ISIS s'intéresse aux méthodes de traitement de l'information, du signal, de l'image et de la vision. Le GDR MaDIC promeut des recherches interdisciplinaires positionnées dans un continuum « des données aux connaissances et à la prise de décision ». Le GDR TAL aborde les thématiques de modélisation informatique de la langue sous toutes ses formes (écrite, orale, signée), ses manifestations et ses applications.

Depuis 2018, le CEA a inscrit la majorité de ses activités en intelligence artificielle dans trois « mailles » programmatiques : microélectronique³⁴, systèmes numériques³⁵ et cybersécurité. Le caractère générique de l'utilisation de l'IA dans de nombreux domaines scientifiques et techniques explique que d'autres actions s'appuyant sur l'IA sont présentes dans les trois autres grandes missions du CEA « civil » : la transition énergétique qui comprend une dimension de simulation numérique importante, les technologies pour la médecine du futur, avec plusieurs actions autour de la santé numérique, et enfin le socle de recherche fondamentale (traitement de données de recherche issues de grands instruments par des techniques d'IA).

Tableau n° 3 : projets en IA déposés et retenus par l'ANR depuis 2014 dans le cadre de l'appel annuel à projets générique

Années	Déposés			Retenus			% de projets IA retenus
	IA	Total	%	IA	Total	%	
2014	273	3 035	9,00	79	815	9,69	28,93
2015	374	4 055	9,22	78	807	9,67	20,86
2016	344	3 371	10,20	102	918	11,11	29,65
2017	460	4 638	9,92	130	1 166	11,15	28,26
2018	405	3 411	11,87	136	1 120	12,14	33,58
2019	728	4 304	16,91	200	1 252	15,97	27,47
2020	544	3 284	16,57	232	1 251	18,55	42,65
2021	1 194	5 159	23,14	311	1 850	16,81	26,05

Source : Traitement Cour des comptes à partir de la base projet de l'ANR

Note de lecture : L'ensemble des abstracts des projets déposés auprès de l'ANR ont été soumis à l'analyse sémantique de la Cour (annexe n° 5). Les colonnes « Total » donnent le nombre de projets déposés ou retenus par l'ANR chaque année. Les colonnes IA donnent le nombre de projets déposés ou retenus qui font apparaître des termes relatifs à l'IA selon l'analyse sémantique de la Cour. La colonne % donne le ratio de projet en IA parmi le total. Ce filtre recouvre l'IA de cœur et les applicatifs et permet d'englober l'ensemble des usages de l'IA.

La diffusion de l'IA dans la recherche se mesure également par l'étude des appels à projets (AAP) de l'Agence nationale de recherche (ANR). L'agence constitue l'opérateur principal de mise en œuvre de la stratégie, en pilotant des AAP spécifiques, mais également en indiquant l'intelligence

³⁴ La maille Microélectronique intègre une inflexion programmatique portant sur le développement d'un processeur neuromorphique dédié à l'IA embarquée.

³⁵ La maille Systèmes Numériques intègre un axe dédié aux technologies pour l'IA. Dans ce cadre, le CEA développe des compétences aussi bien en IA symbolique qu'en IA connexionniste (réseaux de neurones et apprentissage comme pour le traitement des images vidéo ou l'analyse sémantique). Le CEA est en mesure de développer des solutions logicielles et de les porter sur des architectures de calcul contraintes pour les intégrer dans des systèmes embarqués. Une inflexion programmatique porte particulièrement sur le développement d'IA de confiance.

artificielle comme priorité stratégique dans les lignes directrices de l'AAP annuel générique. D'après l'analyse sémantique de la Cour (cf. tableau n° 3), la part des dossiers déposés utilisant des techniques d'intelligence artificielle dans l'appel à projets génériques (AAPG) est passée de moins de 10 % en 2014 à près de 17 % à partir de 2019³⁶ et celle des projets retenus de moins de 10 % en 2014 à plus de 16 % à partir de 2019. À partir de cette même année, le comité d'évaluation scientifique 23 sur la donnée a été renommé « Intelligence artificielle ». Une partie de cette croissance uniquement est soutenue par l'abondement par la SNRIA des crédits de recherche en IA.

Au niveau européen, cet investissement croissant est aussi identifiable au sein du programme *Horizon 2020* de financement de la recherche qui seront analysés en section I.B, et bénéficie à l'écosystème français. En appliquant la même analyse sémantique (cf. tableau n° 4), il est possible d'identifier une croissance du nombre de projets mobilisant de l'IA : alors que le nombre total de projets est presque stable sur la période, la part de projets en IA passe de 7,18% en 2014 à 21,02 % en 2020. Cette croissance est également visible pour les projets retenus. En revanche, la part des financements obtenue par la France demeure stable sur l'ensemble de la période, entre 7 % et 9,57 %, sans pouvoir identifier un effet d'impulsion de la mise en œuvre de la stratégie française.

Tableau n° 4 : nombre de projets en IA déposés et retenus depuis 2014 auprès d'Horizon 2020 et part française au niveau européen

Années	Déposés				Retenus				% de projets IA retenus
	IA	Total	%	% FR	IA	Total	%	%FR	
2014	2 761	38 439	7,18	8,48	301	4 770	6,31	7,64	10,90
2015	3 625	44 493	8,15	7,61	349	4 883	7,15	8,31	9,63
2016	3 787	40 753	9,29	7,92	439	5 225	8,40	9,57	11,59
2017	5 035	43 268	11,64	7,27	515	4 915	10,48	9,13	10,23
2018	6 194	42 580	14,55	7,62	702	5 578	12,59	8,97	11,34
2019	8 444	46 059	18,33	7,45	908	5 875	15,46	9,36	10,75
2020	8 893	42 300	21,02	7,55	787	4 551	17,29	7,12	8,85

Source : Traitement Cour des comptes à partir d'E-corda

Note de lecture : Les données 2021 ne sont pas suffisamment consolidées pour être incluses dans l'analyse. L'ensemble des abstracts des projets déposés auprès de la Commission européenne ont été soumis à l'analyse sémantique de la Cour (annexe n° 5). Les colonnes « Total » donnent le nombre de projets déposés ou retenus par la Commission européenne chaque année. Les colonnes IA donnent le nombre de projets déposés ou retenus qui font apparaître des termes relatifs à l'IA selon l'analyse sémantique de la Cour. La colonne % donne le ratio de projet en IA parmi le total. Ce filtre recouvre l'IA de cœur et les applicatifs et permet d'englober l'ensemble des usages de l'IA. La colonne « %FR » donne la part des financements obtenus par la France dans ces appels à projets en tant que responsable (et non partenaire).

³⁶ Il est à noter que l'ANR diffuse le plan d'action et l'AAPG de l'année N à l'été de l'année N-1. La stratégie en IA ayant été lancée en 2018, la prise en compte de cette priorité par l'ANR ne peut avoir lieu qu'à partir de l'AAPG 2019.

D - Un développement de l'IA éclaté entre plusieurs politiques publiques

En tant que technologie de nature générale et transverse, l'IA est mobilisée sous forme d'applicatifs dans de nombreux secteurs comme la Santé, les mobilités, la transition écologique ou l'éducation. L'IA ne peut se déployer sans le développement des capacités de calcul, de stockage, et des briques logicielles adaptées sur des réseaux centralisés, sur le *cloud*³⁷ ou de manière décentralisée au sein de composants électroniques. L'IA se situe donc à la croisée de nombreuses stratégies et politiques publiques actuelles (cf. schéma n° 2).

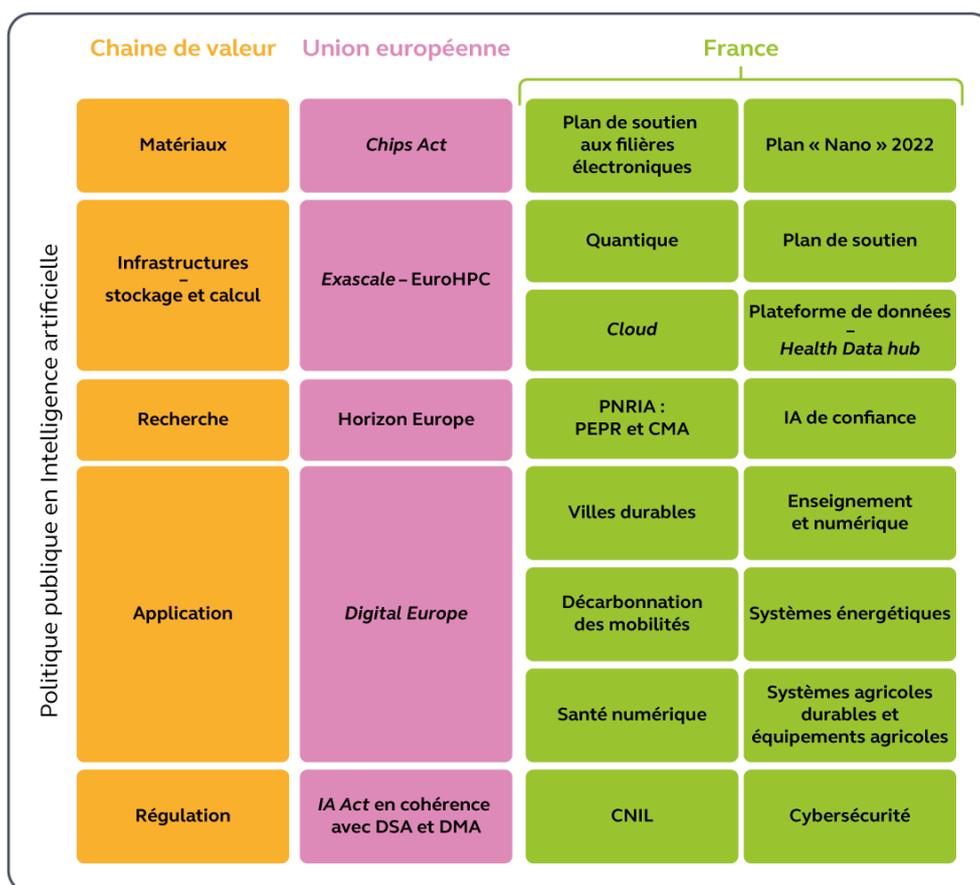
En matière d'infrastructures, les stratégies – notamment d'accélération en matière de « cloud », « 5G et futures technologies de réseaux de télécommunications », « cybersécurité », et « technologies quantiques³⁸ » – ont été lancées dans la même période, notamment sous l'impulsion du calendrier européen commun, et sont imbriquées avec celle de l'IA. Elles sont renforcées par le plan de soutien annoncé le 12 juillet 2022 aux filières électroniques (composants, semi-conducteurs, et systèmes électroniques). Ces stratégies mobilisent des écosystèmes connexes ayant pour objectif de doter la France de capacités souveraines et résilientes en matière numérique (notamment infrastructures, logiciels, composants, hébergement et protection des données individuelles et personnelles, compétences humaines). Il s'agit « placer » l'IA de manière avantageuse dans la compétition internationale pour certains secteurs innovants tels que les technologies quantiques ou l'*exascale*.

L'IA contribue à des politiques publiques d'accélération des transitions nécessaires dans des secteurs comme « l'enseignement et le numérique », « le numérique de santé », « la ville durable et les bâtiments innovants », « la digitalisation et la décarbonation des mobilités », « les technologies avancées pour les systèmes énergétiques », et les « systèmes agricoles durables et équipements agricoles contribuant à la transition écologique ».

³⁷ *Cloud* : vaste réseau de serveurs distants éparpillés tout autour de la planète, reliés entre eux, et destinés à fonctionner comme un écosystème unique. Ces serveurs sont conçus pour stocker et gérer des données, exécuter des applications, ou fournir du contenu ou des services (vidéos diffusées en continu, courrier web, logiciels bureautiques de productivité et autres réseaux sociaux). Au lieu d'accéder à des fichiers et données stockés sur un ordinateur local ou personnel, l'accès à ces ressources en ligne est possible à partir de n'importe quel appareil compatible internet : les informations sont disponibles en tout lieu et en tout temps. Cf. Glossaire également.

³⁸ Née au début du 20^e siècle, la physique quantique s'intéresse au comportement des particules élémentaires : atomes, neutrons, photons, quarks. Dès l'immédiat après-guerre, elle connaît des avancées majeures avec l'invention du laser et du transistor qui vont déboucher sur des applications encore omniprésentes dans notre vie quotidienne : faisceaux laser du lecteur BluRay, lecteurs de code-barres, lasers utilisés en médecine, horloges atomiques au cœur des systèmes GPS de la voiture, microélectronique, entre autres. L'utilisation des propriétés quantiques de la matière permet d'entrevoir des facteurs d'amélioration considérable par rapport aux instruments de mesure classique. Les perspectives sont innombrables sur le plan industriel, de l'automobile à la défense en passant par la santé et l'environnement.

Schéma n° 2 : cartographie des politiques publiques impactant le développement de l'intelligence artificielle



Source : Cour des comptes

Note de lecture : Cette cartographie regroupe les stratégies d'accélération et grands plans qui sont directement et de manière explicite liés à l'intelligence artificielle. L'encadrement en rouge identifie les dispositifs spécifiquement reliés à la SNIA.

Néanmoins, l'absence de gouvernance ou de coordination formelle pour l'ensemble des usages de l'IA et des infrastructures critiques associées peut occasionner des divergences de priorités, des difficultés de diffusion des innovations, ainsi qu'une réduction de l'efficacité des investissements qui peuvent être réalisés de manière dispersée. L'élaboration d'une vision commune et globale de l'action publique en IA serait pertinente pour assurer la coordination de l'ensemble des outils ou dispositifs mobilisables.

II - Une efficacité difficile à apprécier en raison d'un suivi insuffisant des financements associés

A - Un dispositif budgétaire reposant surtout sur des outils financiers déjà existants et ne permettant pas d'appréhender la réalité de la dépense

Tableau n° 5 : autorisations d'engagement initialement prévues pour la stratégie nationale en intelligence artificielle (en M€)

	2018	2019	2020	2021	2022	2018-2022
Crédits budgétaires nouveaux	0	17	38	38	38	131
<i>Dont SNRIA</i>	0	17	38	38	38	131
<i>Dont volet enseignement supérieur</i>	0	0	0	0	0	0
Redéploiements budgétaires	36	162	194	194	194	780
<i>Dont SNRIA</i>	26	39	54	54	54	227
<i>Dont volet enseignement supérieur</i>	0	5	20	20	20	65
Crédits PIA, FII, GPI, FTAP	58	123	145	145	145	616
<i>Dont SNRIA</i>	7	17	21	21	21	87
<i>Dont volet enseignement supérieur</i>	0	6	19	19	19	63
Total	94	302	377	377	377	1 527
<i>Dont SNRIA</i>	33	73	113	113	113	445
<i>Dont volet enseignement supérieur</i>	0	11	39	39	39	128

Source : Cour des comptes d'après les données du coordonnateur national pour l'intelligence artificielle et d'Inria en tant que coordonnateur du volet recherche

Note de lecture : Ces montants ont été actés lors de la RIM du 27 mars 2018. Seuls les trois premiers exercices ont cependant fait l'objet d'une prévision budgétaire détaillée, le budget prévisionnel pour 2020 ayant vocation à être reconduit en 2021 et en 2022.

Le financement de la stratégie (cf. tableau n° 5) repose sur des instruments financiers existants (le troisième programme d'investissements d'avenir PIA 3, le fonds pour l'innovation et l'industrie FII, le grand plan d'investissement GPI, le fonds de transformation pour l'action publique FTAP), des réorientations budgétaires et des financements tiers (collectivités, entreprises, Union européenne). Sur les 1,5 Md€ de financements étatiques initialement prévus, seuls 9 % sont des crédits budgétaires nouveaux, ce qui indique un fléchage important de financement publics sur des ressources déjà dévolues à l'innovation. Conformément à la priorité donnée à la recherche, cette proportion s'élève à 23 % sur le périmètre de la stratégie nationale de recherche en IA. Aucun crédit budgétaire nouveau n'a été alloué au volet enseignement supérieur de la stratégie.

L'utilisation d'outils budgétaires existants pour le financement de la SNIA s'explique par la volonté de permettre son déploiement rapide, sans nécessiter la création de nouveaux dispositifs. La pérennité des dotations budgétaires n'est cependant pas garantie dans le temps, avec un risque de décalage entre la temporalité du financement et celle des leviers pluriannuels mobilisés par la stratégie.

Ces instruments relèvent de missions et programmes budgétaires différents. Il en résulte une architecture budgétaire complexe, qui implique plusieurs administrations, et rend délicat le suivi de l'exécution des financements. Selon les données transmises par le coordonnateur national de la stratégie, et Inria en charge du suivi de l'exécution budgétaire (cf. tableau n° 6), 554,64 M€ de financements étatiques ont *in fine* été engagés pour le volet recherche de la stratégie entre 2018-2022. Parmi ces moyens, il n'a cependant pas été possible de mettre en évidence ceux réellement dépensés. Par ailleurs, les financements de la stratégie pour 2022 sont doublement comptabilisés, au titre de la première phase (période 2018-2022) et de la seconde.

Étant donné le champ d'évaluation retenu, la Cour a concentré son analyse sur les seuls financements alloués à la SNRIA, en tenant compte de l'évolution de son périmètre. En particulier, les appels à projets portés par Bpifrance (y compris les « Grands défis³⁹ ») ont été exclus de l'analyse car relevant du volet « économie » de la stratégie. Dans les données suivies par le coordonnateur, il a pu néanmoins être constaté que l'ensemble des projets identifiés comme impliquant de l'IA dans le système d'information de BPI France sont considérés comme financés par la stratégie. Dans les faits, ces projets ne relèvent pourtant pas tous d'appels à projets mis en œuvre et financés par le plan IA. Il est dès lors impossible de déterminer avec rigueur la réalité de l'exécution de la SNIA sur ce volet et, de ce fait, au niveau agrégé.

Tableau n° 6 : crédits de la stratégie nationale de recherche intelligence artificielle *in fine* engagés sur la période 2018-2022 en M€

	2018	2019	2020	2021	2022 (p)	2018-2022
<i>Crédits budgétaires nouveaux</i>	0	17	43,48	37,58	36,87	134,93
<i>Redéploiements budgétaires</i>	38,52	34,52	71,51	101,91	85,25	331,71
<i>Crédits PIA, FII, GPI, FTAP</i>	0	0	74,50	0	13,50	88,00
<i>Total</i>	38,52	51,52	189,49	139,49	135,62	554,64

Source : Cour des comptes d'après données du coordonnateur national pour l'intelligence artificielle et d'Inria

³⁹ Financés depuis 2019 à hauteur de 120 M€ par an par le Fonds pour l'innovation et l'industrie (FII), les appels à projet « Grands défis » ont pour objet de s'appuyer sur les entreprises et les laboratoires français pour répondre à des enjeux sociétaux dont la portée technologique et scientifique est encore peu explorée, et qui offrent des perspectives de débouchés commerciaux. Figurant dans le volet recherche à son lancement en raison de la part de recherche appliquée qu'ils impliquent, ces derniers ont *in fine* été suivis plus spécifiquement au sein du volet économie.

B - Des outils de financements de courte durée complétés par des financements génériques importants, français et européens

Une très grande majorité des financements de la SNRIA (soit 81,45 %) a été distribuée par des appels à projets (ou manifestations d'intérêts) spécifiques ou généraux (le budget de la SNRIA abonde l'enveloppe globale de l'AAP pour financer des projets en IA) (cf. tableau n° 7). Les financements ont donc été attribués sur des appels à projets sélectifs, ouverts à l'ensemble de la communauté. Ils sont principalement pilotés et suivis par l'ANR, et évalués par des jurys d'experts indépendants, parfois internationaux. Cela requiert un effort de la part des équipes de recherche de conception et de rédaction des projets, sans garantie de réussite. Les critères de qualité scientifique qui sous-tendent l'évaluation par un jury d'experts ont donc guidé les choix de mise en œuvre de cette stratégie. En revanche, les sources de financement sont en grande majorité ponctuelles, pour une durée de quatre ans.

Cette absence de visibilité sur la pérennité de ces guichets de financement peut créer des effets de rupture dans la formation des jeunes chercheurs (programme doctoral) ou la poursuite des programmes de recherche (chaire académique), ce qui porte un préjudice évident au déploiement de la stratégie.

Tableau n° 7 : outils mobilisés pour la SNRIA pour distribuer les financements

<i>Projets</i>	<i>Type</i>	<i>Date</i>	<i>Durée</i>	<i>Montants alloués (en M€)*</i>	<i>Taux de sélection</i>
<i>Institut 3IA</i>	AMI puis AAP	2019	Jusqu'en décembre 2023	82,5	25,00 %
<i>Chaire Hors 3IA</i>	AAP	2019	Jusqu'en 2024	36,65	24,85 %
<i>AAP Générique</i>	AAP	2019-2022	Projet pluri-annuel	73,8	23,63 %
<i>Coopération bilatérale</i>	3 AAP	2020 2021 2022	4 ans maximum	19,21	12,92 %
<i>PhD Program</i>	Appel à programme	2020	3 ans de recrutement jusqu'en 2025 (fin de doctorat)	43,49	56,41 %
<i>Labcom</i>	AAP	Annuel	5 ans	2,15	48,04 %
<i>Challenge IA</i>	AAP	2021	4 ans jusqu'en 2025	3,95	33,33 %

*Source : Cour des comptes à partir des données ANR
* Montant attribué au 28 février 2022.*

Au-delà des crédits spécifiques de la SNRIA, d'autres investissements sont réalisés dans l'écosystème de recherche par des financements génériques. En appliquant l'analyse sémantique (cf. annexe n° 5) aux résumés en français et en anglais des projets déposés auprès de l'ANR, il est possible d'estimer les montants alloués aux quatre catégories suivantes : l'AAP générique annuel, les appels multilatéraux, les bourses tremplin pour un projet ERC (jeunes chercheurs lauréats d'un appel *European Research Council*) et les appels dédiés aux Labcoms. Les financements de la SNRIA sont inclus dans ces tableaux.

Tableau n° 8 : montant alloué par l'ANR sur l'IA (en M€)

Année	Appel générique			Appel multilatéral			Bourse Tremplin ERC			Labcom		
	IA	Total	%	IA	Total	%	IA	Total	%	IA	Total	%
2014	60,60	546,2	11,09	1,10	1,53	72,09				1,80	14,10	12,77
2015	63,57	582,2	10,92	1,11	1,81	61,51				0,00	8,40	0,00
2016	71,19	720,1	9,89	0,00	1,38	0,00	9,30	3,74	24,87	1,50	5,70	26,32
2017	44,00	431,4	10,20	4,89	12,88	37,95	1,17	12,59	9,26	1,70	9,33	18,21
2018	49,58	421,8	11,75	7,21	46,17	15,62		2,34	0,00	1,20	6,30	19,05
2019	65,10	462,6	14,07	6,33	19,85	31,89	1,98	1,66	11,96	4,55	12,82	35,48
2020	66,74	479,1	13,93	1,43	3,95	36,12	7,00	1,95	35,89	3,15	7,00	45,00
2021	112,72	791,6	14,24	7,64	19,37	39,46	4,66	1,30	35,87	7,96	18,84	42,22

Source : Traitement de la Cour des comptes à partir de la base de données projets de l'ANR

Note de lecture : L'ensemble des abstracts des projets déposés auprès de l'ANR ont été soumis à l'analyse sémantique de la Cour (cf. annexe n° 5). Les colonnes « Total » donnent le montant demandé ou alloué de projets déposés ou retenus par l'ANR chaque année. Les colonnes « IA » donnent le montant demandé ou alloué de projets déposés ou retenus qui font apparaître des termes relatifs à l'IA selon l'analyse sémantique de la Cour. La colonne % donne le ratio entre les montants IA et le total. Ce filtre, appliqué aux résumés des projets, recouvre l'IA de cœur et les applicatifs et permet d'englober l'ensemble des usages de l'IA. La colonne « Appel générique » correspond aux appels annuels de l'ANR pour financer la recherche à partir de leur plan d'action. La colonne « Appel multilatéral » regroupe l'ensemble des appels mobilisant plusieurs pays de manière conjointe. Les bourses « Bourse Tremplin ERC » permettent d'accompagner les chercheurs dans leur demande de bourse européenne ERC. La Colonne « Labcom » propose les financements des laboratoires communs mutualisés.

L'investissement de l'ANR est croissant dans le secteur de l'IA quel que soit le vecteur de financement (cf. tableau n° 8). Depuis le lancement de la stratégie, le taux de projets utilisant de l'IA sur l'appel générique se situe autour de 14 % au lieu de 10-11 % précédemment. L'augmentation des financements de l'ANR liée au plan de relance se traduit par un doublement de l'enveloppe en 2021 et donc des financements de la recherche en IA. L'évolution est plus significative pour les Labcom et les appels multilatéraux sous l'effet de financements ou appels spécifiques de la SNRIA.

Les bourses tremplin en IA pour décrocher une bourse européenne⁴⁰ représentent sur les deux dernières années 35 %, ce qui révèle le potentiel français pour former des jeunes équipes de haut niveau. Ces financements dépassent largement ceux spécifiquement issus de la SNRIA, montrant que cette thématique de recherche est également alimentée par les outils classiques de financement de la recherche.

Tableau n° 9 : financements européens de la recherche française en IA (en Md€)

Années	Budget total				Contribution UE				Contribution Nette UE			
	IA	Total	%	% IA FR	IA	Total	%	% IA FR	IA	Total	%	% IA FR
2014	0,8	9,61	8,34	6,40	0,66	8,1	8,18	6,91	0,3	3,42	8,84	7,42
2015	0,81	9,79	8,29	7,10	0,67	8,11	8,31	7,78	0,28	3,93	7,02	11,49
2016	1,19	12,68	9,37	8,28	0,9	9,58	9,38	9,54	0,38	3,91	9,66	11,86
2017	1,34	11,66	11,50	8,17	1,1	9,55	11,55	8,47	0,42	3,99	10,59	10,74
2018	2,01	12,49	16,06	12,16	1,64	10,54	15,52	12,57	0,58	4,42	13,18	11,86
2019	2,64	13,91	18,95	14,32	2,03	11,62	17,40	12,69	0,69	4,62	14,93	13,17
2020	3,14	11,65	26,92	9,90	2,46	9,86	24,97	8,51	0,74	3,81	19,53	8,83

Source : Traitement Cour des comptes d'après la base d'E-corda

Note de lecture : L'ensemble des abstracts des projets déposés auprès de la commission ont été soumis à l'analyse sémantique de la Cour (annexe n° 5 Les colonnes « Total » donnent le montant demandé ou alloué de projets déposés ou retenus par l'ANR chaque année. Les colonnes « IA » donnent le montant demandé ou alloué de projets déposés ou retenus qui font apparaître des termes relatifs à l'IA selon l'analyse sémantique de la Cour. La colonne % donne le ratio entre les montants IA et le total. La colonne %FR donne la part captée par la France sur les financements de l'IA. Ce filtre recouvre l'IA de cœur et les applicatifs et permet d'englober l'ensemble des usages de l'IA. Le budget total représente le coût du projet, la contribution de l'Union européenne le financement de l'UE pour le projet, la contribution nette celle spécifiquement dédiée à la France en tant que responsable du projet. Ce tableau ne mesure pas les contributions obtenues de la part de la France dans des projets où elle n'est pas responsable du projet vis-à-vis de la Commission. Pour la part revenant à la France (% FR), un projet est considéré uniquement si le coordonnateur est français.

Le secteur de la recherche a également été financé à l'échelle européenne notamment au travers du programme *Horizon 2020* pour la période 2014-2020. Avec la même analyse sémantique, il est observé une croissance de l'investissement en IA passant de 660 M€ à 2,46 Md€ sur la période, représentant un quart des financements globaux européens (cf. tableau n° 9). La part captée par la France est passée de 7,13 % en 2014 à 13,17 % en 2019, soit une contribution nette annuelle moyenne de 75,18 M€ sur la période de la SNRIA (2018-2020).

En comparaison avec les autres pays bénéficiaires, la France obtient une part significative et de premier rang dans les financements européens, notamment du fait du déclin du Royaume-Unis (passé de 19 % des financements captés en 2016 à 9 % en 2020) et de l'érosion progressive de l'Allemagne.

⁴⁰ ERC European Research Council.

Pour la première année de mise en œuvre d'Horizon Europe, la Commission européenne estime de son côté un financement direct de l'IA à hauteur de 32 M€ pour la France en 2021 représentant 6,79 % de l'investissement européen (cf. tableau n° 10). Ces financements complètent donc de manière significative ceux de l'écosystème français. La différence observée entre l'estimation de la Cour et de la Commission est due aux différences de classification des projets par la Commission et la Cour, notamment en termes de périmètres IA de cœur ou applicatifs. Les analyses sont néanmoins convergentes dans les ordres de grandeur et l'importance de ces financements complémentaires.

Tableau n° 10 : montant des financements alloués en IA et captés par la France pour Horizon Europe en 2021 (en M€)

	Montant alloué sur l'IA	Montant capté par la France approx.
<i>Cluster 1- Santé</i>	60,00	2,10
<i>Cluster 2 – Culture, créativité et société</i>	8,40	0,12
<i>Cluster 3 – Sécurité civile pour la société</i>	19,60	2,11
<i>Cluster 4 – Numérique, industrie et espace</i>	345,74	24,20
<i>Cluster 5 – Climat, énergie et mobilité</i>	19,00	3,01
TOTAL	452,74	31,55

Source : Retraitement Cour des comptes à partir des données de la Commission européenne

Au-delà des financements identifiables comme contribuant à la recherche en IA, il est difficile d'estimer plus précisément l'emploi des crédits budgétaires classiques dans ce domaine, tant sur la masse salariale (rémunération des enseignants et des chercheurs, doctorants) que sur les investissements des établissements. Les effets de leviers sur l'investissement des collectivités locales et des entreprises sont également complexes à évaluer au-delà des rapports obligatoires annexés à certains appels à projets.

C - Un suivi non exhaustif de l'exécution financière de la stratégie

La SNRIA ne résume pas l'ensemble des actions, et donc des financements de la recherche en IA. Il n'a pas été possible d'évaluer de manière exhaustive les financements publics pour l'IA en raison de leur caractère morcelé. L'estimation produite par la Cour porte de ce fait sur les financements consacrés à l'intelligence artificielle pour les trois opérateurs principaux (CNRS, Inria et CEA).

Le budget annuel d'Inria alloué à l'intelligence artificielle (cf. tableau n° 11) est en accroissement constant, avant même l'élaboration de la SNRIA, passant de 4,01 M€ à 29,95 M€ entre 2014 et 2021. La période 2016-2017 marque une rupture puisqu'en 2018, l'investissement en IA était déjà de 16,62 M€. La part des financements de la SNRIA obtenus par Inria sur la thématique IA est également en croissance, passant de 10,02 % en 2019 à 37,13 % en 2021. Depuis 2018, Inria investit dans l'IA en moyenne 18,61 M€ de ressources – hors financement spécifique SNRIA, alors que la part sur trois ans de financement SNRIA représente 19,33 M€.

Il est à noter comme sources de financements supplémentaires, la montée en puissance de la fondation Inria, dont l'un des objectifs est le développement d'une société numérique de confiance, une augmentation substantielle de la subvention pour charge de service public c'est-à-dire les crédits budgétaires pour le pilotage de la SNRIA (indiqué dans le COP 2019-2023) tout comme les crédits du 3^e PIA.

Tableau n° 11 : financement d'Inria sur la thématique de l'IA (en M€)

Type de financement	2014	2015	2016	2017	2018	2019	2020	2021	Total
Crédits PIA	0,64	0,82	0,92	1,39	2,01	1,82	3,38	3,95	14,93
Crédits issus des appels à projet	0,46	1,01	1,81	2,55	2,83	3,71	4,42	5,30	22,08
Subvention pour charge de service public	0,68	0,72	0,69	0,61	1,04	3,86	5,06	7,84	20,50
Ressources propres des organismes de recherche	0,98	1,69	2,53	3,90	4,36	3,61	5,24	5,80	28,11
Fondations d'établissement	0,00	0,00	0,00	0,00	0,00	0,11	0,11	0,11	0,33
Crédits européens	1,23	2,08	3,14	6,01	6,25	7,01	8,59	6,76	41,08
Autres	0,02	0,16	0,12	0,18	0,14	0,18	0,11	0,19	1,10
Total	4,01	6,48	9,21	14,64	16,62	20,29	26,91	29,95	128,11
% de financement SNRIA	0,00	0,00	0,00	0,00	0,00	10,02	22,97	37,13	

Source : Données Inria, retraitement Cour des comptes

Concernant le CNRS, l'établissement a reçu directement de la SNRIA, 9,67 M€ pour l'accompagnement du réseau d'ingénieurs en appui du déploiement, de l'accompagnement et de la maintenance du super ordinateur dit « Jean Zay » ainsi que pour le développement des infrastructures de calcul (5,67 M€), le recrutement de talents (3 M€) et la création du centre « IA pour les sciences et sciences pour l'IA » (1 M€).

Ce montant de près de 10 M€ ne doit pas masquer la réalité d'un investissement plus important au sein des laboratoires dont le CNRS est tutelle (cf. tableau n° 12). Il s'évalue pour les projets en IA portés par les deux instituts thématiques connexes particulièrement concernés INS2I et

INSMI à 17,97 M€ annuels en termes de nouveaux contrats depuis 2018, soit un triplement par rapport à 2014⁴¹. La part des collectivités et des entreprises évolue peu et reste marginale. En revanche, les financements distribués par l'ANR au CNRS ont été multipliés par 3,7 entre 2018 et 2021, pour partie provenant des financements spécifiques de la SNRIA sans toutefois que leur part ne soit quantifiable de la part de l'opérateur.

Tableau n° 12 : financement du CNRS sur la thématique de l'IA (en M€)

Type de financeurs/année	2014	2015	2016	2017	2018	2019	2020	2021	Total
ANR	1,69	1,27	2,47	2,31	1,69	3,16	5,72	6,31	24,62
Collectivités	0,17	0,25	0,21	0,67	0,13	0,27	0,54	0,33	2,56
Entreprise	0,48	0,93	1,96	2,20	3,47	1,21	0,96	1,16	12,36
Internationaux	0,32	1,50	3,87	3,89	2,48	3,03	4,98	2,56	22,63
Fondations et associations	0,24	0,17	1,97	0,22	0,06	0,02	0,39	0,11	3,17
Ministères et opérateurs	0,93	1,46	0,49	1,90	7,20	0,95	1,96	5,67	20,55
Organismes de recherche	1,44	0,86	1,85	0,80	1,47	3,29	3,67	2,79	16,19
Universités	0,13	0,01	0,29	1,42	1,47	1,40	0,83	0,21	5,76
Total	5,40	6,45	13,10	13,42	17,97	13,32	19,04	19,15	107,84

Source : Données CNRS à partir d'une enquête auprès des unités du CNRS du périmètre des instituts INS2I et INSMI (qui représentent près de 90 % des chercheurs/projets dans les domaines cœur et intégration de l'IA), retraitement Cour des comptes

Le CEA (cf. tableau n° 13) assure le suivi financier de son plan interne d'actions déterminées par les COP. Avant la SNRIA, le CEA investissait entre 20 et 25 M€ par an sur la thématique alors que depuis 2018, cet engagement financier a cru jusqu'à 36 M€ en 2021. Cette croissance est portée essentiellement par la subvention pour charge de service public, les crédits européens et dans une moindre mesure, les ressources propres.

Dans le cadre de la SNIA, le CEA a bénéficié d'un financement sur trois à quatre ans à hauteur d'1 M€ par an pour un réseau national d'ingénieurs de support aux équipes IA. À ce jour, le CEA a contribué à ce réseau mais pas à la hauteur de ses ambitions initiales, en raison de son plafond d'emplois. Dans le cadre des chaires IA et pour quatre ans, deux projets dans le domaine des neurosciences ont été lauréats à l'institut des sciences du vivant- Frédéric Joliot.

⁴¹ Par souci de normalisation, un contrat est intégré dans les estimations, l'année de début de contrat pour l'ensemble de la somme. Ce proxy permet d'estimer la capacité du CNRS à capter des financements.

**Tableau n° 13 : financements des recherches sur la thématique IA
du CEA (en M€)**

<i>Sources de financement/année</i>	2014	2015	2016	2017	2018	2019	2020	2021
<i>PIA 1</i>	0,1	0,2	0,0	0,7	0,0	0,0	0,0	0,0
<i>PIA 2</i>	1,9	2,2	1,6	0,8	1,8	0,7	0,1	0,1
<i>PIA 3</i>	0,0	0,2	0,2	0,2	0,0	0,5	1,0	1,1
<i>Appel à projets</i>							0,6	
<i>Subvention pour charge de service publique</i>	3,9	2,9	2,7	2,7	2,9	5,2	5,7	8,2
<i>Ressources propres des organismes de recherche</i>	12,3	15,0	17,2	18,2	21,5	19,1	17,0	20,7
<i>Crédits européens</i>	2,4	2,8	3,5	3,2	3,0	5,6	7,3	6,3
Total	20,7	23,2	25,3	25,7	29,2	31,1	31,8	36,4

Source : Cour des comptes d'après les données du CEA

Note de lecture : Les financeurs via les crédits des PIA1, 2 et 3 sont l'ANR, l'ADEME, la SATT Paris Saclay (dont le CEA est depuis 2020, observateur pour le compte des ONR au conseil d'administration), Bpifrance et la Caisse des dépôts. L'ANR finance les appels à projet.

Cette mise à plat montre la difficulté de disposer d'une vision exhaustive et cohérente des financements de l'écosystème de recherche en IA, tant dans l'exécution des crédits de paiements que le décaissement effectif des fonds. La synthèse des financements de l'écosystème en IA repose sur plusieurs logiques :

- une logique budgétaire de suivi des crédits de l'État menée par Inria dans le cadre de son rôle de pilote de la SNIA ;
- une logique de comptabilité analytique dans les opérateurs en fonction de leur intérêt sur la thématique.

Néanmoins, cette dernière n'est pas réalisée par les opérateurs de manière exhaustive. Il en va de même au sein du Secrétariat général pour l'investissement (SGPI) au travers des différents PIA qui remontent une information non normalisée de l'investissement en IA que ce soit ou non en lien avec la SNIA.

Il existe ainsi une difficulté réelle de recellement et de suivi de l'ensemble des crédits consentis à l'IA. Cette carence ne permet pas de mesurer un investissement global de la puissance publique sur l'intelligence artificielle.

Il serait possible de traduire cette politique publique de souveraineté numérique en un « jaune budgétaire » qui traduirait l'effort financier consenti dans le domaine au sein d'un document unique, ou bien, en un « orange budgétaire » qui répondrait à l'exigence d'évaluation de la performance par rapport aux objectifs, indicateurs et valeurs du projet annuel de performance.

III - Un positionnement scientifique de la France en IA en retrait depuis 2018

Le renforcement du positionnement de la France en IA était la principale ambition de la SNIA à son lancement, qui s'illustre par le double objectif assigné au volet recherche :

- installer durablement la France dans les cinq pays experts en IA, à l'échelle mondiale ;
- faire de la France le chef de file européen de la recherche en IA.

Selon les entretiens avec les autorités compétentes au moment de son élaboration, ces objectifs ont eu un objectif principal de communication, sans chercher à mesurer les effets globaux escomptés de la stratégie, et, par conséquent, sans identification précise de la métrique ou du classement international à retenir⁴². Chaque métrique ou classement susceptible d'être retenu a d'ailleurs une sensibilité différente à l'évolution de la production scientifique. La présence des chercheurs français dans les conférences les plus prestigieuses (indicateur retenu à ce jour par Inria dans le cadre de la SNRIA) est, par exemple, plus réactif à court terme de la qualité scientifique, le nombre de citations d'articles de recherche ayant une portée pouvant davantage s'apprécier sur plusieurs années. Différents critères ont ainsi été utilisés par la Cour pour évaluer le positionnement de la France.

A - Une production scientifique globalement constante sur la période

L'ensemble des indicateurs disponibles au niveau international a été rassemblé pour identifier, sans *a priori*, les principaux axes de différenciation des pays sur la scène internationale et établir sur cette base des classements multicritères (les données et méthodes employées sont détaillées en annexes n° 3 et 4). À partir de cette analyse, deux axes thématiques discriminants majeurs ont pu être identifiés :

- la production de la recherche en IA, à titre principal. Cette action, qui s'assimile à une mesure de performance globale, se mesure par des indicateurs de volumes, tels que le nombre d'articles, de conférences ou de citations associées. En pratique, ces différents indicateurs sont très fortement corrélés entre eux et à des variables d'échelle telles que le produit intérieur brut (PIB) ;

⁴² À la connaissance de la Cour, le premier rapport qui analyse le rang scientifique de la France en fonction de différentes données et indicateurs est une étude scientométrique indépendante relative à l'IA réalisée en décembre 2019 pour le compte des pilotes de la stratégie. Les objectifs de la stratégie ont été assignés dès 2018.

- l'efficacité de la recherche en IA, à titre secondaire. Cet axe se mesure par des indicateurs analogues à ceux du premier axe mais normalisés par le PIB des pays. Pour le dire autrement, certains pays « réussissent » même avec moins de moyens.

La recherche joue un rôle prépondérant dans le bon positionnement des différents pays, loin devant la valorisation de la recherche, et probablement la formation. Ce diagnostic conforte le choix fait de centrer le premier volet de la stratégie française sur des enjeux de recherche fondamentale.

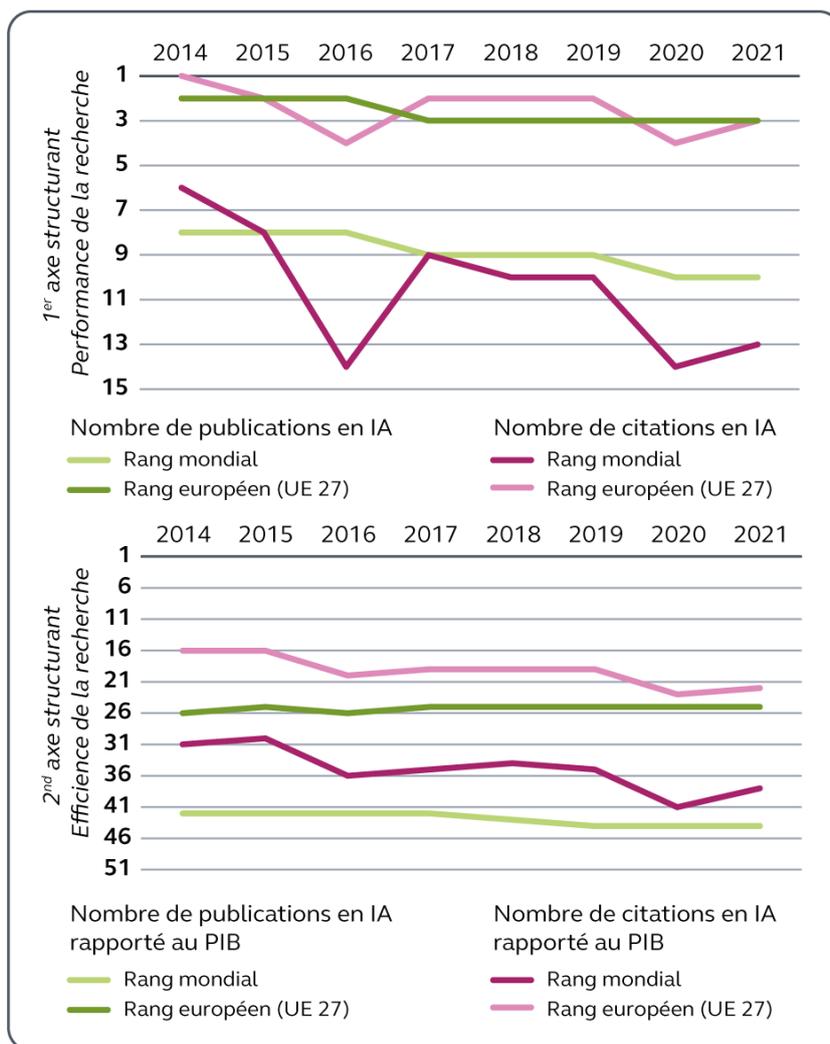
Si la France apparaît assez performante en matière de recherche en IA (10^e rang mondial et 2nd rang européen en 2021 en nombre de publications en IA sur un total de 47 pays comparés), sa recherche dans ce domaine apparaît peu efficace au regard du produit intérieur brut de la France (44^e rang mondial et 25^e européen en 2021 pour le même critère rapporté au PIB).

Entre 2014 et 2021, son rang apparaît en diminution quel que soit l'indicateur analysé (cf. graphique n° 2). Ce résultat est cohérent avec les résultats d'une étude scientométrique indépendante relative à l'IA réalisée en 2021 pour le compte des pilotes de la stratégie et fondée sur des données et indicateurs différents de ceux ici analysés. Celle-ci indique également une diminution du rang de la France en matière de production scientifique, avec un passage de la 10^e place en 2012 à la 13^e place en 2019 après un « pic » à la 7^e place entre 2014 et 2016. L'outil *Web of Science* conduit à des constats analogues (voir annexe n° 3), avec, s'agissant du principal schéma de catégorisation des productions en IA, une tendance à la baisse en matière de citations pour la France depuis 2020. À l'exception de l'Allemagne, des Pays-Bas, de la Suisse et du Japon (voir analyse du « top 20 » mondial), ce phénomène n'est pas observé pour les autres pays en pointe dans ce secteur. L'emploi d'autres méthodes proposées dans *Web of Science* pour identifier les productions scientifiques au cœur de l'IA conduit à des constats analogues sur l'évolution du rang de la France (voir annexe n° 3 et documents en ligne).

Ces éléments indiquent de façon concordante que le lancement de la stratégie en 2018 a été opportun pour permettre, à terme, à la France de ne pas décrocher de cette compétition. Les effets concrets de la stratégie sur la production scientifique (publications dont articles scientifiques et actes de conférences) ne sont cependant pas encore pleinement intervenus. Le temps de la recherche étant long, ils ne pourront être mesurés de façon fiable à travers des indicateurs quantitatifs avant un délai de plusieurs années.

Si le fait d'adopter une stratégie en IA joue un rôle marginal en regard d'autres variables explicatives (voir annexe n° 4), une analyse économétrique tenant compte de l'hétérogénéité structurelle entre pays permet néanmoins de montrer que la mise en place d'une stratégie nationale augmente légèrement la productivité d'un pays de 5,60 % en matière de publications et de 4,35 % en matière de publications corrigées du PIB. Cela suggère un impact positif de la formalisation d'une stratégie pour la production scientifique en IA, effet renforcé dans les pays les plus développés.

Graphique n° 2 : évolution des rangs de la France dans le monde et en UE à 27 suivant des critères complémentaires et différentiant sur la scène internationale



Source : Cour des comptes

Note de lecture : Les méthodes et données mobilisées sont détaillées en annexe n° 3. S'agissant des indicateurs relatifs à l'efficience de la recherche, il est à noter que la crise de la covid 19 a pu impacter les PIB des pays à compter de 2020. Le suivi des rangs des pays permet de limiter les biais associés à la normalisation par les PIB annuels, comme le montre la relative stabilité des deux séries temporelles concernées sur la période 2019 à 2021

Une stratégie n'est pas ainsi un facteur intrinsèque déterminant de la production scientifique, mais vient renforcer des leviers existants et peut favoriser des effets positifs sur la production scientifique telles que les mesures en faveur de l'attractivité, de créations de postes de chercheurs, les règles éthiques ou la valorisation. Le diagnostic sur la qualité, de l'efficacité et de l'efficience de la stratégie française en IA passe donc par une analyse de chaque mesure et des résultats atteints, ce qui sera l'objet des développements proposés ci-après.

B - Une amélioration depuis 2018 de la valorisation de la recherche en IA

Outre le renforcement des capacités de recherche, la priorité donnée à ce domaine a *in fine* pour objectif l'application des résultats aux secteurs public et privé. À partir des données transmises par l'Institut national de la propriété industrielle (INPI) et des publications de France Digitale, une amélioration de la valorisation de la recherche en IA se constate depuis 2018, avec des brevets déposés et des start-up créées dans le domaine en croissance. Si le lancement de la stratégie a pu avoir un effet d'entraînement sur la diffusion de l'IA dans l'économie, il est encore trop tôt pour le démontrer formellement.

1 - Un retard de la France en matière de dépôt de brevets liés à l'IA qui semble s'être résorbé

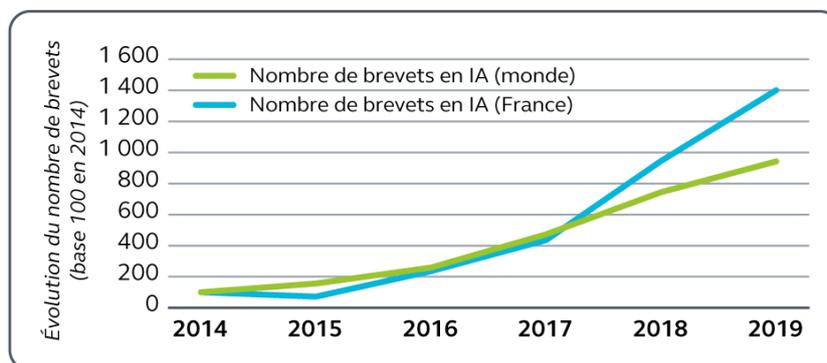
Compte tenu des définitions de l'IA et des périmètres retenus, qui peuvent varier d'un organisme ou d'un pays à l'autre⁴³, les comparaisons internationales en matière de brevets liés à l'IA doivent être interprétées avec prudence. Un certain nombre d'algorithmes créés le sont en open source, réduisant *de facto* les possibilités de valorisation des recherches sous forme de brevets⁴⁴.

⁴³ L'une des particularités s'agissant de l'IA est en effet de savoir ce qui peut être breveté (l'algorithme, la machine et/ou l'application), la réponse pouvant varier selon les pays. En France, comme aux États-Unis et au Japon, une machine ne peut pas être inventeur, contrairement à d'autres pays comme l'Australie ou l'Afrique du Sud.

⁴⁴ Les innovations en IA étant difficilement brevetables, il peut exister un risque de divulguer des informations, en déposant un brevet fragile au risque de se faire copier, ce qui explique que les chercheurs hésitent à breveter leurs recherches en IA.

Selon les données de l'INPI, et après retraitement selon leur propre méthodologie, une croissance exponentielle du volume de dépôts de brevets liés à l'intelligence artificielle est observée à l'échelle mondiale depuis le début des années 2000, passant de 236 à 29 545 entre 2000 et 2020. Jusqu'en 2016, une telle évolution n'était cependant pas observée en France, le nombre de brevets liés à l'IA augmentant ou diminuant selon les années. À partir de 2016, le nombre de brevets liés à l'IA déposés en France a connu une hausse significative (passant de 8 à 26 par rapport à l'année précédente), et une progression constante est observée depuis (cf. graphique n° 3⁴⁵). Le lancement de la stratégie nationale en IA est concomitant avec le renforcement de cette tendance, avec une augmentation très nette en 2018, année à partir de laquelle le nombre de brevets liés à l'IA déposés en France a dépassé la centaine. La proportion dans le total des brevets de ceux liés à l'IA est désormais supérieure en France par rapport à celle constatée à l'échelle mondiale (cf. graphique n° 4).

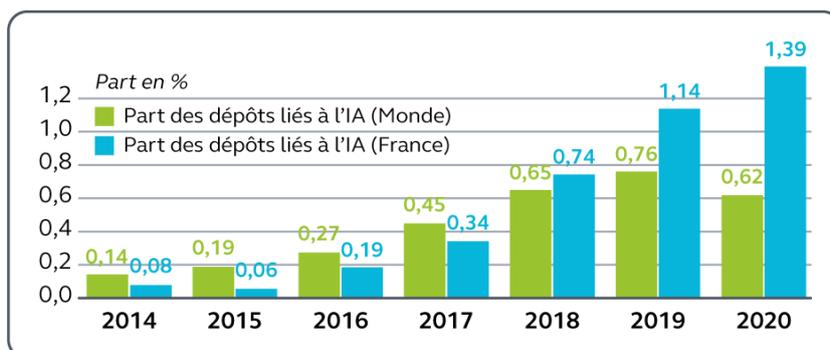
Graphique n° 3 : volume de dépôts de brevets liés à l'IA dans le monde et en France sur la période 2014-2019



Source : Cour des comptes d'après les données transmises par l'INPI et arrêtées au 31 mars 2022
 Note de lecture : Le nombre de brevets est représenté en base 100 avec comme référence l'année 2014 (3 136 brevets IA dans le monde et 11 en France). La courbe représente l'évolution du nombre de brevets comparé à cette référence 100. Ainsi en 2019, l'indice pour le monde atteint 942 (soit 9,42 fois plus de brevet qu'en 2014) alors que l'indice pour la France atteint 1 400 (soit 14 fois plus qu'en 2014). * L'année présentée correspond à la première année de dépôt

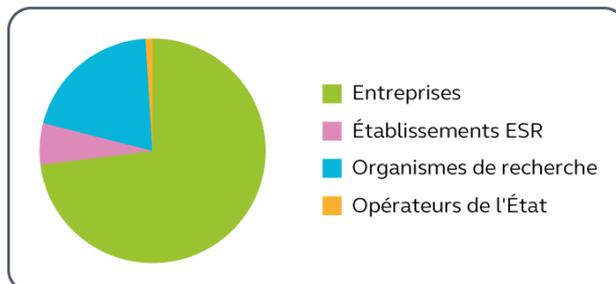
⁴⁵ À la date de transmission à la Cour, seul un tiers des données était disponible pour l'année 2020 avec un nombre de dépôts de brevets ne correspondant pas au total des brevets déposés. La tendance à la hausse par rapport aux années précédentes se confirmait néanmoins.

Graphique n° 4 : part des brevets mondiaux et français liés à l'IA sur la période 2014-2020



Source : Cour des comptes d'après les données transmises par l'INPI et arrêtées au 31 mars 2022
Note de lecture : À la date de transmission à la Cour, seul un tiers des données était disponible pour l'année 2020.

Graphique n° 5 : répartition des dépôts de brevets en IA par des entités françaises sur la période 2014-2020



Source : Cour des comptes d'après les données transmises par l'INPI et arrêtées au 31 mars 2022

Les déposants français (cf. graphique n° 5) les plus importants sont, s'agissant des organismes de recherche, le CEA⁴⁶ et le CNRS, et, dans une moindre mesure, l'Institut national de la santé et de la recherche médicale (Inserm), l'AP-HP dans le secteur public hospitalier et s'agissant du secteur privé, des sociétés comme Thalès, Safran, Amadeus, Orange, PSA, Atos, Continental ou encore STMicroelectronics.

⁴⁶ 93 brevets ont été déposés par le CEA depuis 2016 avec une pointe de dépôt en 2021 à hauteur de 26 brevets et 16 en 2020.

La contribution de la France au dépôt de brevets liés à l'IA à l'échelle mondiale reste cependant faible, légèrement supérieure à 0,5 % en 2019, par rapport à 1,13 % pour l'Allemagne où depuis 2018 sont déposés près du double de brevets liés à l'IA.

2 - Une croissance du nombre de start-up en intelligence artificielle

À l'instar des brevets, l'évolution du nombre de start-up en IA dépend également du périmètre retenu. L'association France Digitale a établi une cartographie des start-up en IA à partir de méthodologies propres⁴⁷. 318 start-up en IA sont recensés en 2018, 432 en 2019, 453 en 2020 et 502 en 2021, soit une forte hausse. Selon l'association, les fonds levés par ces dernières sont également en augmentation, passant de 708 M€ en 2020 à 1,6 Md€ en 2021.

La cartographie établie montre également une pénétration de l'IA dans l'ensemble des secteurs de l'économie, avec des start-up en IA particulièrement représentées dans les secteurs de la robotique, de la santé, de l'énergie, de la finance, de l'agriculture ou encore des transports.

⁴⁷ La liste des start-up en IA est élaborée notamment à partir de segmentations IA proposées par certaines plateformes de données (*Crunchbase*) et d'analyses de bases de données internes de France Digitale. En 2020, cette méthodologie a permis à l'association de ne retenir que 453 start-up en IA, lorsque plus de 700 start-up déclaraient utiliser l'IA dans leurs solutions.

Cartographie des start-up en IA par Bpifrance

L'établissement public Bpifrance a établi une cartographie des start-up IA en France, qu'il est susceptible d'accompagner et de financer. À partir de questionnaires adressés à ces start-up IA, un certain nombre d'enseignements peuvent être tirés sur la capacité de la France à valoriser sa recherche en IA, d'autant plus qu'il n'existe pas au sein de l'ESR de traçabilité des projets de recherche transférés sous la forme de création d'entreprise.

- Depuis 2019, Bpifrance concentre son action de financement sur les innovations de rupture. Les 471 start-up IA répertoriées par Bpifrance représentent le premier domaine technologique (18 %) des start-up *deeptech*.
- 223 start-up *deeptech* fondées sur des technologies IA ont levé en montant cumulé 3,7 Md€ entre 2017 et mi 2022, via 391 tours de table, dont 2,3 Md€ en 2021.
- Les start-up IA qui créent des applications représentaient en 2019 80 % du total des start-up IA en France, dont 32 % d'applications fonctionnelles et 68 % d'applications sectorielles, alors que celles qui proposent des outils « cœur IA » ne représentent que 19 % du total.
- Les start-up IA étaient domiciliées en 2019 principalement à Paris où résident 215 start-up sur les 327 start-up franciliennes et dans les autres sites 3IA, Grenoble (22), Toulouse (28) et Nice (13). La stratégie française de concentrer les recherches dans les quatre sites 3IA, qui regroupent plus de la moitié des start-up, peut être considérée comme un succès, même si 42 % des start-up sont domiciliées en dehors des agglomérations de ces quatre sites.
- Les start-up *deeptech* IA sont nées de recherches développées et valorisées directement à 62,5 % dans des laboratoires de recherche situés dans des Organismes nationaux de recherche (ONR) et à 38,5 % grâce au travail de valorisation des sociétés d'accélération du transfert de technologies (SATT). Le CNRS arrive largement en tête des ONR (68 % des réponses), devant Inria (25 %) l'INRAE (5 %) et PSL (2 %). Pour les start-up IA dont les recherches ont été valorisées par l'intermédiaire des SATT, 67 % des recherches ont été faites dans les universités et 33 % dans une ONR (au CNRS pour 27 % et à Inria pour 6 %).

CONCLUSION ET RECOMMANDATION

La France est l'un des premiers pays à s'être dotés d'une stratégie nationale en IA, en 2018. Elle donne la priorité au renforcement de la recherche fondamentale, à l'instar de tous les pays ayant choisi de formaliser une stratégie en la matière. Il s'agit de renforcer les capacités nationales de recherche en IA dans un premier temps, pour accélérer sa diffusion dans un second temps. L'absence de stratégie aurait pu mener au décrochage de la France en matière de recherche scientifique, d'autant plus que la très grande majorité des pays en ont adopté une.

Bien que les investissements financiers dans la recherche aient été importants dans cette première phase, le manque d'information sur l'exécution des crédits et l'absence de consolidation des différentes sources mobilisées en matière d'IA ne permettent pas de mesurer globalement les efforts financiers étatiques consentis et d'apprécier l'efficacité de la stratégie. En outre, la politique publique de l'IA ne couvre pas l'ensemble des moyens d'intervention. La multiplication des sources de financement de court et moyen termes est susceptible de nuire au déploiement de la stratégie en occasionnant des effets de rupture dans la formation des jeunes chercheurs ou la poursuite de programmes de recherche.

La Cour ne peut ainsi qu'apporter une réponse en demi-teinte à la première question évaluative qui a guidé ses travaux : l'objectif de placer la France dans les cinq premiers pays experts en IA à l'échelle mondiale affiché par la stratégie n'est pas atteint. Sur la période analysée, la France conserve difficilement une place au 10^e rang à l'échelle mondiale, en nombre de publications en IA sur un total de 47 pays comparés. Au niveau européen, elle se maintient au 2^e rang.

Les analyses économétriques faites par la Cour montrent néanmoins que le lancement d'une stratégie en 2018 a été opportun pour renforcer les leviers de la production scientifique et pour permettre à la France de ne pas décrocher de la compétition sur la scène scientifique internationale.

À terme, alors que l'IA irrigue de nombreux secteurs, que certaines politiques publiques portent sur des enjeux connexes, et que d'autres ont pour objectif sa diffusion dans des domaines précis de politique publique, c'est une vision coordonnée et intégrée de l'ensemble des actions prises et des politiques publiques adoptées qui doit être recherchée.

La Cour formule la recommandation suivante :

- 1. traduire la politique publique sur l'intelligence artificielle dans un document budgétaire de synthèse qui permettra de la suivre et d'en mesurer les effets (MEFSIN).*

Chapitre II

Un écosystème à mieux structurer et à pérenniser

La stratégie nationale de recherche se fonde sur des individus ayant investi dans le développement de l'IA, des laboratoires de recherche et des institutions, notamment de l'enseignement supérieur. La stratégie vise à la fois à structurer un réseau d'acteurs et à faire émerger des pôles d'excellence de niveau mondial pour être en capacité de porter les ambitions de celle-ci. Cela apparaît comme des conditions *sine qua non* pour permettre à la France de conserver son rang mondial.

Cette section de l'évaluation vise à répondre aux deux problématiques suivantes :

- la stratégie nationale de recherche a-t-elle accéléré la structuration de l'écosystème français en IA ? ;
- la stratégie nationale de recherche des pôles d'excellence est-elle efficace et efficiente ?

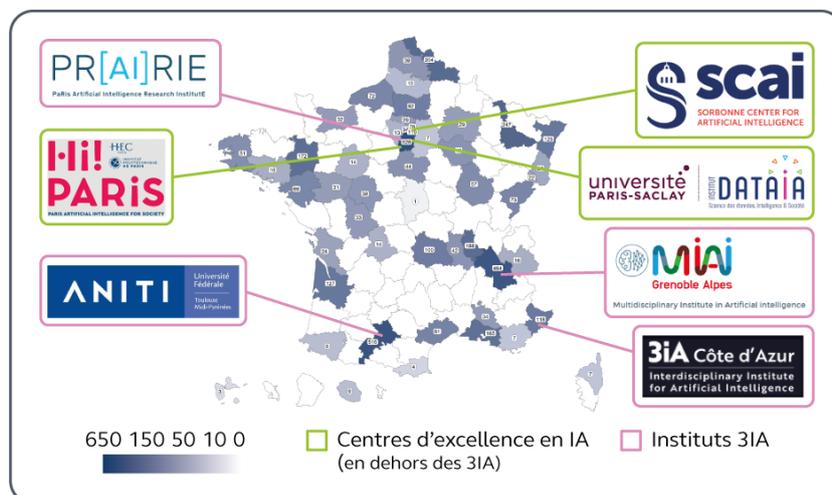
I - Une stratégie de pôles d'excellence pertinente qui doit gagner en maturité

La notion de pôles d'excellence en IA

La constitution de « pôles d'excellence » est un axe structurant de la SNRIA. Ces pôles regroupent aujourd'hui trois types d'entités :

- les quatre instituts interdisciplinaires en IA (instituts 3IA) identifiés au cours de la première phase via un AAP spécifique de l'ANR ;
- les 43 chaires individuelles portées par un chercheur identifié au travers d'un autre AAP spécifique de l'ANR ;
- les trois autres centres d'excellence, dits « hors 3IA » identifiés par le coordonnateur de la stratégie en 2021, sans faire l'objet d'un AAP spécifique et dont les membres peuvent détenir une chaire individuelle.

Carte n° 1 : implantation des instituts thématiques en IA (3IA-PR[AI]RIE, MIAI, 3IA Côte d'Azur et ANITI) et des centres d'excellence (SCAI, DATAIA et Hi! PARIS) en regard des zones historiquement actives dans ce domaine.



Source : Cour des comptes

Note de lecture : L'activité historique en IA est mesurée à travers la répartition départementale des thèses en IA soutenues entre 1989 et 2019 dans les établissements français d'enseignement supérieur à partir des données ouvertes de theses.fr (ABES). Les thèses sont données par année de soutenance. La détection des thèses en IA est réalisée suivant la méthode sémantique développée par la Cour. Les logos des 3IA et des centres d'excellence en IA sont tirés de leurs sites web officiels.

A - Le renforcement des bassins actifs en IA par la labellisation d'instituts 3IA et de chaires individuelles

La labellisation d'instituts interdisciplinaires en intelligence artificielle (3IA) est une mesure emblématique de la SNRIA. Leur sélection s'est déroulée en deux phases : quatre pôles de recherche, de formation et d'innovation (parmi 12 dossiers déposés) ont d'abord été présélectionnés par un jury international et indépendant⁴⁸ au cours d'un appel à manifestation d'intérêt clos le 28 septembre 2018, et ensuite labellisés pour quatre ans à la suite d'un AAP clos le 19 février 2019 :

- 3IA Côte d'Azur à Nice, porté par l'Université Côte d'Azur (UCA) ;
- ANITI (*Artificial and Natural Intelligence Toulouse Institute*) à Toulouse, porté par l'Université fédérale de Toulouse Midi-Pyrénées (UFTMiP) ;
- MIAI (*Multidisciplinary Institute in Artificial Intelligence*) à Grenoble, porté par l'Université de Grenoble Alpes (UGA) ;
- PRAIRIE (*PaRis Artificial Intelligence Research InstitutE*) à Paris, porté par Paris Sciences et Lettres (PSL) et Paris Cité.

Les instituts 3IA se situent dans les zones déjà actives historiquement dans le domaine de la recherche en IA (voir carte n° 1). Ce maillage territorial reflète des bassins où sont implantés des centres universitaires de taille critique et de renom, des antennes locales d'organismes de recherche (centres du CNRS, d'Inria et du CEA, notamment), des structures adossées de valorisation de la recherche (Labcom et IRT), un tissu économique et industriel⁴⁹ (grands groupes, petites et moyennes entreprises, start-up) et enfin des collectivités locales impliquées dans le développement et le soutien à l'innovation.

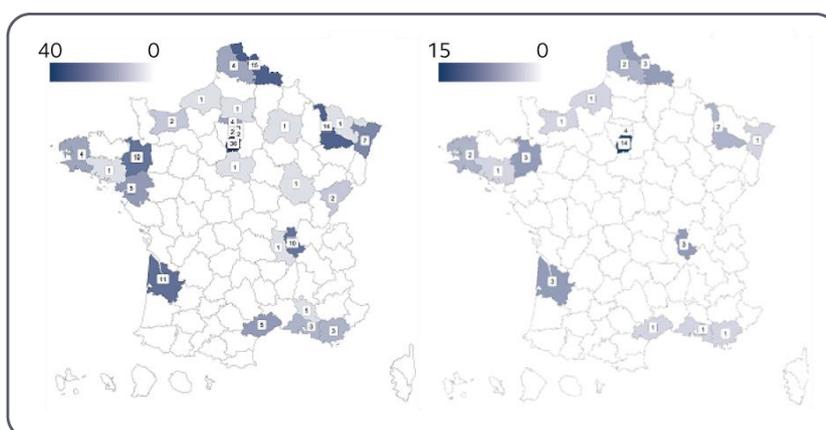
Un appel à projets dit « Chaires hors 3IA » a été mis en place en complément pour valoriser des chercheurs ou petites équipes disposant d'une expertise de haut niveau sans être intégré à un centre disposant d'une masse critique pour bénéficier d'un rayonnement international. Les

⁴⁸ Présidée par Mme Aude Billard, présidente de l'EPFL et de membres éminents venus des États-Unis, de la Grande-Bretagne, du Canada, du Japon, de la Corée du Sud, d'Allemagne et d'Israël.

⁴⁹ Paris et son tissu économique, industriel et entrepreneurial dense et diversifié, Toulouse et son secteur d'activités aéronautique, spatiale et de transport, Nice Sophia-Antipolis, première technopole française et européenne d'innovation et Grenoble-Rhône-Alpes et son développement économie fondé sur la recherche et les technologies de pointe (microélectronique, imagerie, capteurs, biopuces, robotique, etc.).

titulaires de chaires sont majoritairement des enseignants-chercheurs d'universités ou des grandes écoles, ainsi que des chercheurs d'Inria. Le taux de sélection de cet appel à projets a été de 24,86 %, soit un niveau de sélection légèrement supérieur mais dans la moyenne des autres AAP de l'ANR⁵⁰. La diversité territoriale des projets *in fine* lauréats est similaire à celle d'émanation des candidatures (cf. carte n° 2).

**Carte n° 2 : répartition spatiale des candidatures (gauche)
et des lauréats de l'AAP (droite) des chaires IA hors Instituts 3IA**



Source : Traitement Cour des comptes à partir de la base de données ANR. Carte de gauche, répartition géographique des dépôts de candidatures. Carte de droite, répartition des lauréats

Une concentration spécifique de chaires individuelles s'observe dans trois écosystèmes, identifiés par la suite par le coordonnateur national comme des centres d'excellence, en dehors de tout AAP compétitif spécifique impliquant un jury indépendant expert en IA :

- Institut DATAIA à Saclay, porté par l'Université Paris-Saclay (11 chaires) qui est également un institut de convergence (issu d'un AAP compétitif pluri-disciplinaire) depuis 2017⁵¹ ;
- SCAI à Paris porté par Sorbonne université (deux chaires) ;
- Hi! PARIS à Palaiseau porté par l'Institut Polytechnique de Paris et HEC Paris de création plus récente (cinq chaires).

⁵⁰ L'objectif est de 23 % pour l'appel annuel à projets générique dans le cadre du plan de relance et de la loi de programmation de la recherche.

⁵¹ DATAIA a bénéficié d'un financement sur 10 ans dans le cadre de la deuxième vague de l'AAP Institut de convergence.

Sept centres d'excellence en recherche en IA émergent ainsi (carte n° 1) et fédèrent une communauté de chercheurs en France. Des coopérations existent entre les instituts 3IA et les autres, notamment au niveau francilien. Ainsi, l'ensemble des centres d'excellence de la région (DATAIA, Hi! PARIS, PRAIRIE et SCAI sous le pilotage d'Inria Paris) a proposé un projet structurant autour de l'IA auprès de la région Île-de-France pour les cinq prochaines années (DIM4IA). Il permettra d'encourager les programmes doctoraux, notamment en finançant des thèses et en créant une communauté de doctorants, et de renforcer les capacités d'équipement, notamment une plateforme expérimentale robotique de haut niveau. Cette dynamique converge avec les priorités nationales, notamment sur la consommation des ressources et la frugalité. Ce projet permet de renforcer la visibilité de l'écosystème IA francilien en attirant les talents et en favorisant la valorisation économique et en renforçant les collaborations avec les industriels.

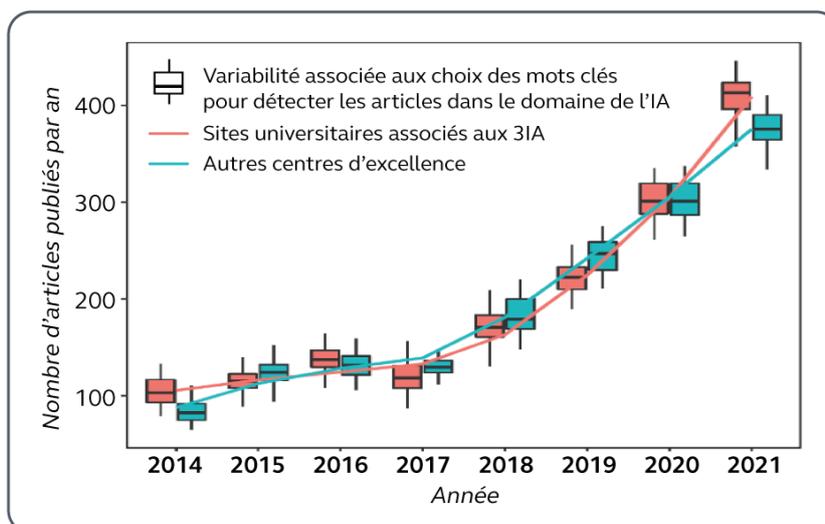
Les centres d'excellence hors 3IA sont explicitement considérés par la stratégie d'accélération en IA (phase 2) comme des leviers aux côtés des 3IA pour renforcer la recherche, l'innovation et la formation⁵². Ces centres n'ayant pas répondu à un AAP spécifique, les attendus de la puissance publique vis-à-vis d'eux sont moins explicites, et leur modèle de développement est moins contraint par les conditions de financement.

Dès lors, il serait utile de clarifier les missions des sept centres d'excellence et les conditions d'accès à ce statut et d'étudier l'opportunité d'un élargissement de la labellisation sans nécessairement passer par un nouvel AAP compétitif.

⁵² Comme évoqué dans le chapitre de ce rapport consacré à la formation (cf. chapitre III), les sept centres d'excellence (3IA et hors 3IA) ont ainsi pu déposer un dossier de candidature pour la première vague pré-fléchée de l'appel manifestation compétences et métiers d'avenir lancé dans la stratégie d'accélération et de France 2030.

B - Une performance des 3IA et des autres centres d'excellence comparable en matière de production scientifique

Graphique n° 6 : production scientifique comparée des principaux sites universitaires associés aux 3IA et aux centres d'excellence



Source : Cour des comptes, à partir des données scientométriques issues de HAL retraitées à travers l'application de filtres sémantiques

Note de lecture : Une méthode probabiliste, issue d'un processus formalisé de consultation des experts du comité d'accompagnement de l'évaluation, a ici été employée (analyse de sensibilité sur la base de 40 répliques aléatoires ; cf. annexes n° 5 et n° 6).

Afin de comparer la production des centres d'excellence et d'examiner les synergies éventuelles avec les sites associés, des analyses sémantiques et scientométriques appliquées à la recherche des sites en matière d'IA ont été menées (cf. graphique n° 6 et annexes n° 5 et 6).

D'après le graphique n° 6, la mise en œuvre de la SNIA se traduit par un accroissement des taux de production scientifique en IA pour l'ensemble des sites. Le graphique montre également que le niveau de production est sensiblement identique⁵³ pour les sites associés aux 3IA,

⁵³ Les différences ne sont pas significatives au sens statistique (emploi de tests de Student) en tenant compte de la variabilité associée aux mots clés utilisés pour détecter les articles scientifiques dans le domaine de l'IA. Les métriques employées sont le nombre d'articles scientifiques en IA publiés sur la période 2018-2021 ainsi que le ratio entre le nombre d'articles scientifiques en IA publiés sur la période 2018-2021 et le nombre d'articles scientifiques en IA publiés sur la période 2014-2017.

d'une part, et les autres centres d'excellence, d'autre part. Les sites universitaires présentent tous des productions scientifiques en IA en forte progression depuis 2018 (cf. annexe n° 6) avec un doublement de la production annuelle moyenne sur la période 2018-2021 par rapport à 2014-2017. L'analyse des données de WoS conduit à des constats convergents, avec toutefois une dynamique haussière moins marquée, ce qui pourrait notamment s'expliquer par des différences dans les définitions et méthodes employées pour détecter les productions dans le domaine de l'IA.

Le site de Saclay pris dans son ensemble (Université Paris Saclay, Institut Polytechnique de Paris et HEC Paris) se place à la 1^{ère} position en nombre d'articles publiés en IA, suivi des sites parisiens ainsi que de l'Université Grenoble-Alpes, puis des autres sites répartis sur le territoire national. Le rôle de la région francilienne est prépondérant en matière d'IA, une analyse que confirme l'outil WoS quel que soit le schéma de classement employé⁵⁴.

Au sein des 3IA, les performances observées en matière de production scientifique diffèrent. Sur la période 2020-2021, les taux de croissance des publications des 3IA ont été hétérogènes. Par ordre décroissant : + 49 % pour le 3IA MIAI à Grenoble, + 18 % pour le 3IA ANITI à Toulouse, + 14 % pour le 3IA PRAIRIE à Paris, et + 3 % pour le 3IA Côte d'Azur.

Pour mesurer l'impact sur les sites universitaires associés, deux métriques de performance des 3IA ont été calculées⁵⁵. Bien qu'étant fortement corrélées, les données issues de WoS et HAL conduisent à des hiérarchisations différentes de l'impact des quatre instituts 3IA sur leurs sites respectifs. Ces mesures de « valeur-ajoutée » ou d'effet additif de la stratégie nationale doivent en outre être interprétées avec prudence car les 3IA ont été lancés récemment et l'impact de la recherche est difficilement perceptible sur une durée inférieure à quelques années.

Sur les 792 articles⁵⁶ publiés par les instituts 3IA lors de la période sous revue, seulement trois articles ont été co-signés par des chercheurs issus d'au moins deux 3IA⁵⁷, soit moins de 1 % de la production agrégée. Les mutualisations entre 3IA appréciables à l'aune de ce seul critère apparaissent très négligeables. La question du renforcement de la mutualisation entre 3IA est donc cruciale car certains 3IA travaillent sur

⁵⁴ Voir annexe n° 6.

⁵⁵ Le rapport entre le nombre de publications en IA publiées sur la période 2018 à 2021 rapportée au nombre de publications en IA sur la période 2014 à 2017 (1^{ère} mesure de performance) ainsi que l'écart entre le taux de croissance 2020-2021 de la production du 3IA et celui de la production en IA de l'ensemble du site dont le 3IA ressort (2^e mesure). Voir annexe n° 6.

⁵⁶ Nombre d'articles uniques au sens du champ « docid » dans HAL.

⁵⁷ Les trois articles cosignés l'ont été au titre de deux instituts 3IA seulement. Aucun article retracé dans HAL n'implique trois ou quatre instituts 3IA.

des domaines d'application de l'IA qui leur sont communs ; « la santé et l'IA » est le secteur le plus fédérateur en rassemblant les 3IA à l'exception d'ANITI, mais également « l'industrie 4.0 », la robotique, ou encore « l'énergie et le développement durable ». Des actions de type activités événementielles, conférences, ateliers, sont à l'œuvre (cf. monographies des quatre 3IA en annexes n° 16 et suivantes). Cependant, la coopération scientifique peine singulièrement encore à produire des « papiers de recherche » et une production scientifique publiés de manière commune.

C - Une visibilité des instituts sur la scène internationale encore à construire

La montée en puissance de la recherche française en IA au niveau international était l'un des objectifs de la labellisation d'instituts interdisciplinaires. Selon les termes de l'AAP, les instituts labellisés devaient établir des partenariats étroits avec d'autres établissements à l'étranger disposant d'une recherche de pointe en IA⁵⁸. Le renforcement d'un nombre ciblé de centres de compétences transdisciplinaires pour la recherche en IA est en effet une mesure existante dans de nombreuses stratégies nationales⁵⁹, et pour le cas des pays étudiés dans le cadre de cette évaluation, au Canada⁶⁰, aux États-Unis⁶¹, en Allemagne⁶², au Royaume-Uni⁶³, au Japon⁶⁴ et en Finlande⁶⁵.

⁵⁸ Selon l'AAP, les instituts devaient « proposer une stratégie convaincante visant à développer cette reconnaissance internationale de premier plan ».

⁵⁹ Le parangonnage montre qu'à chaque fois la concentration s'opère à partir de l'existant afin de ne pas créer de nouvelles structures juridiques mais de concentrer une expertise transdisciplinaire. Cette approche est conforme à celle suivie par la stratégie française.

⁶⁰ Trois centres de recherche et innovation en IA ont été constitués dès 2017 par l'Institut canadien de recherches avancées (CIFAR) chargé de mettre en œuvre la stratégie canadienne, dans les trois grands centres canadiens de l'intelligence artificielle : l'*Alberta Machine Intelligence Institute* (Amii) à Alberta, le MILA à Montréal et l'Institut vecteur à Toronto.

⁶¹ Sept "AI institutes" ont été créés en 2020, 11 en 2021, répartis dans des universités situées dans différents États, avec des financements alloués par la National Science Foundation.

⁶² Outre le Centre allemand de recherche en intelligence artificielle (DFKI), le ministère fédéral de l'éducation et de la recherche soutient cinq autres centres de compétence pour la recherche en intelligence artificielle, situés à Berlin (BIFOLD), Dresde/Leipzig (ScaDS.AI), Dortmund/St. Augustin (ML2R), Munich (MCML) et Tübingen (Tübingen AI Center).

⁶³ L'*Alan Turing Institute* fondé en 2015 à l'initiative du gouvernement britannique est le principal centre d'excellence de la recherche dans ce domaine.

⁶⁴ La stratégie japonaise a fixé comme objectif la structuration d'un réseau rassemblant les institutions engagées dans la R&D en IA (universités et instituts de recherche). Les trois instituts composant son cœur sont : le *RIKEN Center for Advanced Intelligence Project*, le *National Institute of Advanced Industrial Science and Technology center* et le \Rightarrow .

⁶⁵ Le principal centre sur lequel s'appuie la stratégie finlandaise est le Centre finlandais d'intelligence artificielle (FCAI) créé par l'Université d'Aalto et l'Université d'Helsinki et réunissant des universités, des acteurs industriels et des acteurs publics.

Les instituts 3IA sont rapidement parvenus à mettre en place des collaborations ou partenariats internationaux. Ils sont également tous intégrés à des réseaux européens ou internationaux de coopération scientifique⁶⁶. S'agissant des écosystèmes français et allemand, par exemple, la première phase de la stratégie avait pour objectif de renforcer plus particulièrement les coopérations entre les centres pluridisciplinaires des deux pays. L'analyse présentée dans la partie II.IV.B montre en effet un certain dynamisme des collaborations institutionnelles entre les instituts des deux pays. Elle met cependant aussi en évidence des effets structurants des 3IA limités au-delà de l'échelon local, et le faible renvoi de citations des uns aux autres, en particulier de PRAIRIE, MIAI ou 3IA Côte d'Azur qui renvoient peu sur les autres 3IA en est une illustration.

L'identification des 3IA sur la scène internationale reste encore faible, et inégale selon les instituts. Il est vrai que leur labellisation est récente. La crise sanitaire a de son côté contribué à mettre un coup d'arrêt à la mise en place des coopérations. Les effets limités des 3IA sur la structuration de l'écosystème à l'échelle nationale ainsi que leur faible collaboration scientifique doivent faire l'objet d'une attention particulière. Engager les centres d'excellence dans une valorisation réciproque de leurs travaux plus systématique contribuerait à renforcer leur visibilité tant sur la scène nationale qu'internationale.

D - Des modes de financement des centres d'excellence reposant sur une logique de trop court terme

À l'exception de l'institut PRAIRIE⁶⁷, les modalités de financement des instituts 3IA sont semblables. Conformément aux obligations de l'AAP, ils doivent financer au moins deux tiers de leur activité par des moyens propres et des apports industriels, ce qui a pu mettre en difficulté le modèle au moment de traverser la crise sanitaire de la covid 19. La SNRIA ne peut donc contribuer au maximum que pour un tiers de leur financement total (cf. tableau n° 14 colonne PIA 3 pour les instituts 3IA). Des financements croisés existent entre des partenaires publics et privés, intéressés pour ces derniers par les développements applicatifs de l'IA et dont la place est croissante dans le plan de financement des instituts.

Les trois autres centres d'excellence n'ont bénéficié d'un financement spécifique de la stratégie IA que pour des chaires ou programmes doctoraux. Ils ont mobilisé par ailleurs des ressources issues d'autres dispositifs issus des programmes d'investissements d'avenir ainsi que des collectivités territoriales et des entreprises.

⁶⁶ À l'instar d'ELLIS, CLAIRE ou encore TAILOR.

⁶⁷ Les tutelles, et un club de partenaires (entreprises privées) sous forme de mécénat avec un ticket de participation au montant identique quelle que soit la taille de l'entreprise qui adhère à ce club.

Ce système a le mérite d'assurer que l'investissement public n'est poursuivi qu'après une évaluation circonstanciée de son usage et de l'activité des centres. Il comporte cependant plusieurs difficultés. La temporalité de quatre ans est courte au regard de la durée nécessaire à l'apparition de résultats dans le secteur de la recherche en IA. À l'approche de l'échéance de l'année 2022, l'absence de certitudes sur le maintien de ces centres et de leurs financements publics a pu limiter leur capacité à démarcher de nouveaux financeurs privés sans qu'il soit possible de chiffrer ce phénomène. Ce modèle a pu en fin de période rendre difficile le lancement de doctorats d'une durée de trois ans en règle générale.

Tableau n° 14 : financements cumulés des centres d'excellence depuis 2019 (en M€)

<i>Instituts</i>	PIA 3	PIA 4 France relance France 2030	Tutelles	Crédits européens	Autres crédits publics	Collectivités	Entreprises	Total
<i>PRAIRIE</i>	19,43	8,84	4,00			1,74*	6,62	38,63
<i>ANITI</i>	18,00	5,10	0,06	1,58	0,12	1,67	10,74	37,27
<i>MIAI</i>	19,00	12,92	2,68	10,37	4,18	0,49	22,75	69,396
<i>Côte d'Azur</i>	16,00	15,21	1,42	10,26	2,26	1,48	10,89	54,52
<i>Hi! PARIS</i>		7,78		**	1,37	0,65*	12,80	22,60
<i>DATAIA</i>	8,34	11,45	0,35			1,24*	0,61	21,99
<i>SCAI</i>	0,24	7,01	3,90			1,01*		12,16

Source : Données issues des différents instituts d'excellence transmis en juin 2022 retraitées par la Cour
 Note de lecture : PRAIRIE, DATAIA, Hi! PARIS et SCAI ont été lauréats d'un programme (DIM4IA) porté par la région Île-de-France (budget 12,44 M€). Les montants « Collectivités » qui sont inclus dans le tableau correspondent à deux années de financement, identiques à 2022 pour laquelle le projet a reçu 2,3 M€. En septembre 2022, à la suite de l'évaluation de mi-parcours positive des instituts, PRAIRIE a bénéficié d'un financement complémentaire de 2 M€ alors que celui de MIAI et Côte d'Azur a été de 3 M€⁶⁸. Ce tableau ne valorise pas les apports des partenaires publics en mise à disposition de plateformes ou affectation sur projets. La méthode de collecte fondée sur du déclaratif ne permet pas une comparaison précise entre chaque institut. Il y a notamment des pratiques hétérogènes en matière de valorisation des contrats de recherche obtenus par les chercheurs membres des instituts. ** Cinq professeurs liés à Hi! PARIS ont reçu des bourses « ERC » financées par le Conseil européen de la recherche.

⁶⁸ La labellisation des instituts ayant été initialement prévue pour une première période de quatre ans, les financements issus de la SNRIA ont été attribués en deux tranches, la seconde étant conditionnée à une évaluation à mi-parcours.

Le choix de mener une politique d'investissement de la recherche sur quatre ans est en décalage avec les financements des Instituts de convergences ou des Labex (d'ailleurs souvent à l'origine du déclenchement d'un projet 3IA), construits sur une durée de cinq ans renouvelables à l'issue d'une période probatoire. S'agissant, par exemple, des centres interdisciplinaires en IA allemands, co-financés et cogérés par l'État fédéral et les *Länder* depuis le 1^{er} juillet 2022⁶⁹, la durée pendant laquelle ce financement est assuré est dans un premier temps de sept ans, puis un comité d'experts internationaux évalue de nouveau les centres et détermine la suite à donner aux financements.

II - Un investissement en capital humain pour la recherche en IA ciblé sur les chercheurs juniors

Pour renforcer la communauté de recherche en IA, il est nécessaire d'assurer un flux entrant de jeunes chercheurs formés et compétents en capacité de mener des recherches dans l'ensemble des disciplines scientifiques et de sciences humaines et sociales. D'après la consultation menée dans le cadre de l'enquête, les jeunes chercheurs (69,50 %) et les doctorants (58 %) sont des profils particulièrement en tension en matière de vivier disponible et de qualité. Le doublement du nombre de docteurs formés en IA (500 thèses classiques et 100 thèses CIFRE supplémentaires) est un des objectifs de la SNRIA, articulé autour de trois mesures : le financement de doctorants au sein des instituts 3IA, celui de programmes doctoraux sur appel à projets et celui de l'adjonction de 50 thèses CIFRE supplémentaires.

Appel à programmes « Contrats doctoraux en Intelligence Artificielle – Établissement »

Les programmes doctoraux proposés par appel à programme et financés en 2020 par l'ANR ont pour objectif de développer la formation en IA. Ils doivent s'inscrire dans une ou plusieurs écoles doctorales existantes, accréditées par le MESR, auxquelles les équipes de l'établissement-candidat doivent être rattachées. Ces programmes doctoraux sont institués hors des instituts 3IA. Leur action doit s'inscrire dans celles menées dans le cadre de la stratégie nationale de recherche coordonné par Inria. Cet appel vise à co-financer 200 contrats doctoraux en IA au niveau national.

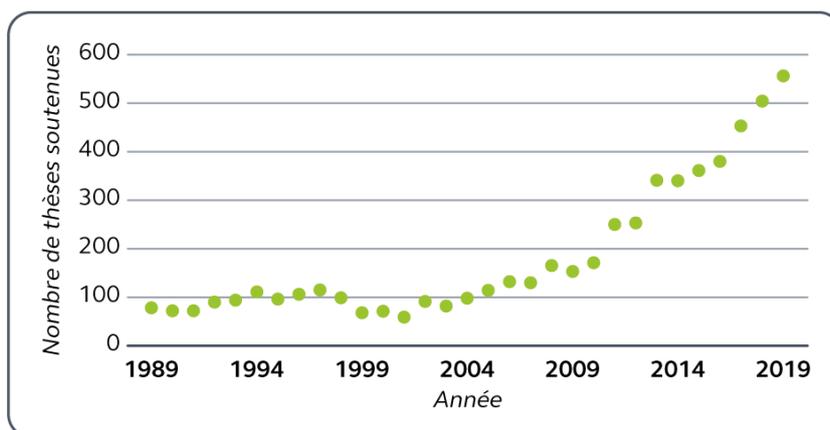
Vingt-deux programmes ont ainsi été financés sur 39 candidatures pour un montant de 43,49 M€ avec un ticket moyen de 1,12 M€.

⁶⁹ Le co-financement était une condition préalable au soutien du ministère fédéral de l'éducation et de la recherche après 2022. Les *Länder* où sont situés les centres de recherche se sont engagés à participer à la même hauteur que l'État fédéral. Chaque centre recevra entre 8,5 et 12,5 M€ par an en fonction de sa taille, pour un budget total de 50 M€ de la part de l'État fédéral. Le DFKI devrait recevoir 11 M€ par an.

Le doctorat étant le diplôme nécessaire à la poursuite d'une carrière académique en recherche, l'évolution au long cours des thèses soutenues en IA constitue un indicateur du vivier potentiel de jeunes chercheurs dans le domaine en France⁷⁰. En appliquant des analyses sémantiques à la base de données ouverte « theses.fr » maintenue par l'agence bibliographique de l'enseignement supérieur, un accroissement considérable des travaux dans ce domaine s'observe dès les années 2000, soit bien avant le lancement de la stratégie (cf. annexe n° 7, ainsi que le graphique n° 7).

Entre 1989 et 2019⁷¹, les thèses en IA représentent environ 1,5 % des thèses publiées. L'accroissement observé sur la période est de +3,6 points, les thèses soutenues en lien avec l'IA représentant 0,7 % de celles publiées en 1989 et 4,3 % en 2019. L'accroissement de leur nombre dans ce domaine a été de + 613 % contre + 18 % pour l'ensemble des thèses, soit un taux 34 fois plus important.

Graphique n° 7 : évolution du nombre de thèses françaises soutenues en IA sur la période 1989-2019



Source : Cour des comptes à partir des données ouvertes de theses.fr (ABES)

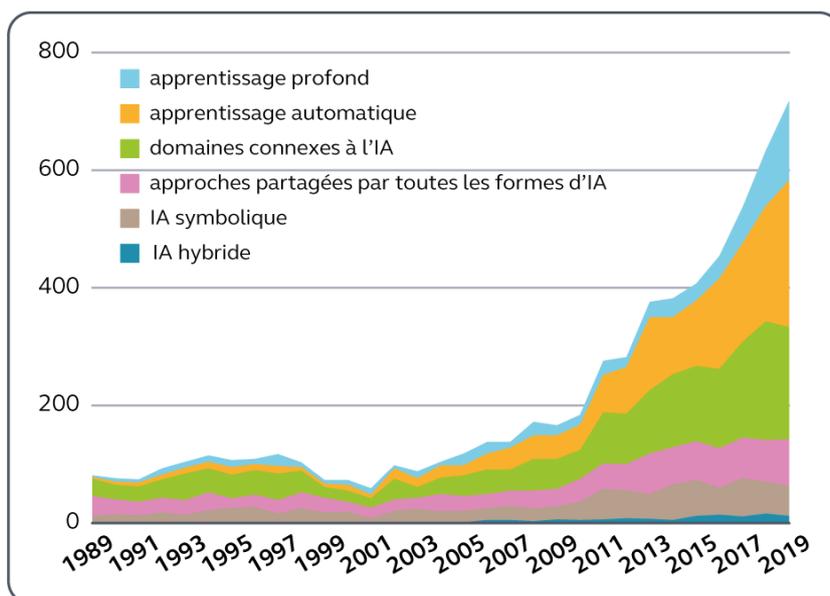
Note de lecture : Les thèses sont données par année de soutenance. La détection des thèses en IA est réalisée suivant la méthode sémantique développée par la Cour. Les étiquettes sur les points indiquent les années théoriques de lancement des thèses, avec un décalage avec la soutenance supposé égal à trois ans en moyenne.

Ces évolutions françaises relatives aux thèses en IA sont concordantes avec les tendances retracées au niveau international par les données du *AI Index Report* de l'Université Stanford qui indiquent une progression des publications scientifiques en IA de 1,3 % en 2011 à 3,8 % en 2019.

⁷⁰ Il est bien noté qu'il existe des mobilités géographiques entrantes et sortantes à la fin du doctorat entre la France et les autres pays. Ces flux sont plus difficilement quantifiables et sont traités dans le chapitre 3.

⁷¹ Années où les données de « theses.fr » peuvent être considérées comme fiables (cf. annexe n° 7).

Graphique n° 8 : évolution des sous-domaines de l'IA mesurée à travers la présence de mots clés dans les thèses sur la période 1989-2019



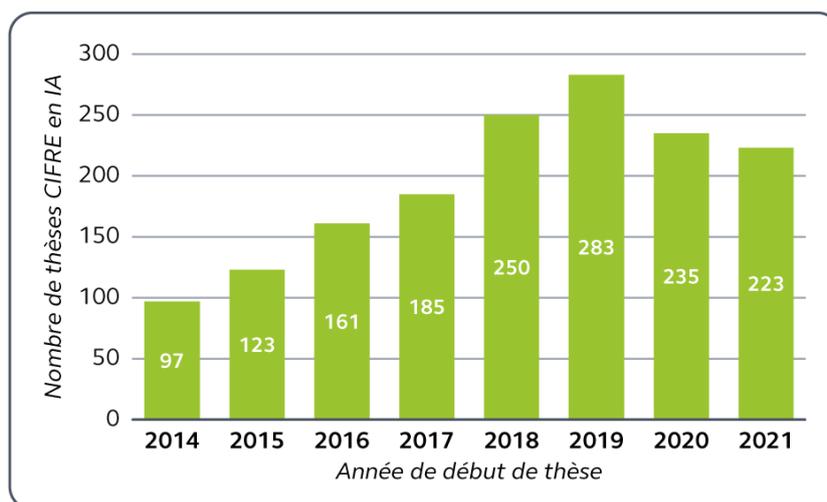
Source : Cour des comptes à partir des données ouvertes de theses.fr (ABES)

Note de lecture : Les thèses sont données par année de soutenance. La détection des thèses en IA est réalisée suivant la méthode sémantique développée par la Cour à travers les sous-corpus de mots clés identifiés. Dans la mesure où une même thèse peut citer des mots clés de sous-corpus distincts, elle peut être citée plusieurs fois ; il s'ensuit un léger décalage des chiffres « vers le haut » par rapport à la figure précédente.

La plupart de ces thèses sont réalisées en sciences « dures », au premier chef desquelles l'informatique, ce qui suggère une diffusion encore faible de l'IA dans les autres domaines scientifiques (cf. annexe n° 7). Sur la période 2009-2019, l'apprentissage profond en particulier, et l'apprentissage automatique plus largement, sont les domaines de l'IA qui ont connu une croissance massive (taux de croissance de + 694 % et + 525 %, respectivement). Ils sont suivis des domaines connexes à IA (+ 275 %), qui regroupent les principaux champs d'applications de l'IA par apprentissage, et enfin des autres formes d'IA (+ 163 % pour les approches partagées par toutes les formes d'IA ; + 132 % pour l'IA symbolique ; + 100 % pour l'IA hybride). Ces évolutions sont concordantes avec la forte montée en puissance de l'IA par apprentissage observée au niveau international et indiquent également l'existence de synergies avec les autres formes d'IA qui ont toutes prospéré sur la période, bien qu'à des degrés moindres (cf. graphique n° 8).

Pour la période la plus récente, les données issues de *theses.fr* ne sont pas encore fiabilisées et les données issues de l'association nationale recherche technologie (ANRT) relatives aux thèses CIFRE⁷² ont donc été privilégiées. Ces thèses particulières représentent environ 10 % de l'ensemble de celles en préparation⁷³. Elles ont pour spécificité d'être réalisées en collaboration entre la recherche publique et les acteurs socioéconomiques et favorisent le transfert des recherches fondamentales vers le secteur économique.

Graphique n° 9 : évolution du nombre de thèses CIFRE en IA débutées sur la période 2014-2021



Source : Cour des comptes d'après les données ANRT. La détection des thèses dans le domaine de l'IA a, dans le cas présent, été réalisée par l'ANRT

Sur la période 2014-2021 (cf. graphique n° 9), le nombre de thèses CIFRE en lien avec l'IA attribuées chaque année a plus que doublé, avec un rythme de croissance bien plus élevé que l'ensemble des thèses CIFRE (+ 130 % contre + 19 %, avec une part des CIFRE en IA passée de 7 % à 14 %). L'évolution a été croissante entre 2014 et 2019, avec depuis une stabilisation autour de 235 thèses annuelles. En 2019, la SNRIA a financé

⁷² Conventions industrielles de formation par la recherche, dispositif financé par le ministère de l'enseignement supérieur et de la recherche depuis 1981 et géré par l'Association nationale recherche technologie (ANRT). L'entreprise reçoit une subvention de 14 000 € sur trois ans.

⁷³ Cf. le tableau 39.04 « Le financement des doctorants inscrits en première année de thèse (2009-10 à 2020-21) » du MESRI-DGESIP/DGRI-SIES disponible sur : https://publication.enseignementsup-recherche.gouv.fr/eestr/FR/EESR15_R_39/le_doctorat_et_les_docteurs/

50 thèses CIFRE fléchées spécifiquement sur l'IA au sein des dossiers déposés auprès de l'ANRT qui gère le dispositif des thèses CIFRE pour le compte du MESR⁷⁴. Cet appel a permis de renforcer le nombre de thèses en IA sans rencontrer de difficulté de recrutement. La baisse consécutive en 2020 et 2021 est concomitante à la crise de la covid 19 mais reste limitée. Cela montre à la fois la disponibilité d'un vivier de talents pour répondre à l'appel à projets.

L'analyse économétrique (cf. annexe n° 7) souligne que la mise en œuvre de la SNIA a été concomitante à l'accroissement des thèses CIFRE en IA significativement supérieur à celui des thèses CIFRE en général. Les écosystèmes issus des établissements d'enseignement supérieur porteurs des centres d'excellence accueillent significativement plus de thèses CIFRE (plus de 40 %), sans néanmoins bénéficier d'un effet additionnel sur leur attractivité.

En rapprochant les deux bases de données *theses.fr* et CIFRE au niveau de granularité des établissements et des territoires (cf. annexe n° 7), les thèses CIFRE en IA débutées postérieurement à la mise en œuvre de la SNIA ont été allouées aux centres d'excellence en IA tels que l'Université Paris Saclay, Sorbonne Université ou encore l'Université Grenoble Alpes, sans bénéficier cependant d'un effet additif de la stratégie. Cette allocation des CIFRE a été strictement proportionnelle à l'activité en IA telle qu'observée sur des bases historiques à partir des données de *theses.fr*. Il n'y a donc pas eu de stratégie de surconcentration des ressources dans les centres d'excellence, en particulier les instituts 3IA, au-delà de leur capacité historique. Une raison potentielle, évoquée par certains acteurs, est la capacité limitante d'encadrement par des personnes habilitées à diriger des recherches.

Tableau n° 15 : évolution du nombre de doctorants et post-doctorants dans les laboratoires dont le CNRS est tutelle

Type de personnel	2016	2017	2018	2019	2020	2021	Écart 2018/2021 (%)
Nombre de doctorants	1 326	1 393	1 425	1 573	1 762	1 913	+ 34,25
Dont doctorants étrangers	612	614	652	695	772	761	+ 16,72
Nombre de post-doctorants	268	250	280	322	348	382	+ 36,29
Dont post-doctorants étrangers	147	132	144	198	205	219	+ 52,08

Source : Traitement Cour des comptes à partir des données CNRS

⁷⁴ Programme 172 - Recherches scientifiques et technologiques pluridisciplinaires de la Mission Interministérielle Recherche et Enseignement Supérieur.

À l'échelle nationale, le CNRS a pu répertorier les doctorants et post-doctorants dans les différents laboratoires dont il est tutelle sur les thématiques de l'IA (cf. tableau n° 15). La croissance des effectifs est significative, avec une hausse de 35 % environ pour les deux catégories de personnel. Pour les profils internationaux, c'est néanmoins au niveau des post-doctorants que la croissance est la plus marquée (+ 52,08 %), alors même que les entretiens avec les acteurs de l'écosystème ont mis en évidence la plus forte difficulté et la compétition accrue entre organismes pour recruter ces profils. Cela confirme une croissance des recrutements juniors (doctorants et post-doctorants) et une attractivité spécifique pour les profils étrangers : 16,72 % pour les doctorants et 52,08 % pour les post-doctorants.

Cette analyse, confirmée par les entretiens dans les différents centres d'excellence, montre l'attractivité croissante du doctorat en IA qui confère une expertise spécifique nécessaire pour poursuivre soit une carrière académique, soit une carrière dans le secteur privé. Cela signifie également que les laboratoires français ont la capacité et l'attractivité nécessaire pour former les experts en IA.

III - Un accès renforcé aux infrastructures souveraines de calcul

A - Une appropriation rapide par la communauté scientifique du supercalculateur Jean Zay

Pour faire des projets en IA, différents types d'infrastructures de calcul peuvent être utilisées par les scientifiques en fonction de leurs besoins⁷⁵ allant de leur ordinateur personnel, à des infrastructures de calculs de plusieurs dizaines de pétaflops⁷⁶. Conformément aux préconisations du rapport Villani, la SNRIA a prévu la mise en place d'un supercalculateur conçu pour des applications d'intelligence artificielle. Ce projet a été confié à l'organisme du Grand équipement national du calcul intensif (GENCI), opérateur public créé en 2007 afin de démocratiser l'usage de la simulation numérique par le calcul haute performance (HPC).

⁷⁵ GENCI distingue les infrastructures de calcul au niveau régional, dont la puissance est de l'ordre de 1 pétaflop et qui sont généralement exploitées dans les centres universitaires, les centres nationaux de calcul qui en France cumulent plus d'une quarantaine de pétaflops, et les moyens de calcul au niveau européen qui offrent en totalité une capacité de 150 pétaflops.

⁷⁶ Unité de mesure qui renvoie au nombre d'opérations de calculs numériques par seconde. 1 pétaflop = 1 million de milliards d'opérations par seconde.

Le choix de l'opérateur a été de ne pas acheter de machine spécifiquement conçue pour les applications d'IA, mais d'adapter une procédure d'appels d'offres en cours, afin d'y intégrer les besoins spécifiques à l'IA. Conçu par la société Hewlett-Packard Enterprise (HPE), le supercalculateur Jean Zay a ainsi été mis en place sous la forme d'un supercalculateur « convergé », ce qui signifie que sa composition matérielle et logicielle a été élaborée pour permettre tant des simulations numériques (HPC) que des simulations en IA (les deux types ne nécessitant pas les mêmes architectures). Hébergé et opéré par l'Institut du développement et des ressources en informatique scientifique (IDRIS/CNRS), le supercalculateur est entré en service en novembre 2019. Son renouvellement est prévu en 2026.

À l'origine, la mise en place de la machine a donc été décidée indépendamment du programme national de recherche en IA. L'apport de la stratégie réside dans l'ajout d'un volet IA à un supercalculateur HPC, afin de répondre aux besoins spécifiques de la recherche en IA et de permettre une convergence des communautés.

Les accès dynamiques, une procédure flexible très utilisée pour les projets en IA

Pour un projet de recherche donné, deux types d'accès permettent d'obtenir des heures de calcul :

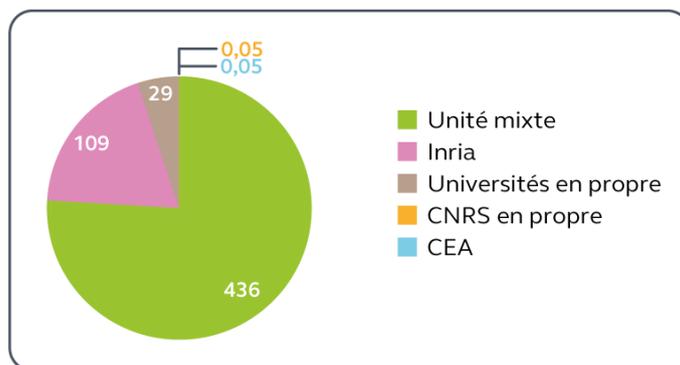
- les accès dynamiques, lorsque les ressources de calcul cumulées sollicitées sont inférieures à 50 000 heures *Graphics Processing Unit* (GPU) ou à 500 000 heures *Central Processing Unit* (CPU) ;
- les accès réguliers, lorsque les ressources sollicitées dépassent ces seuils.

Portant sur des projets de plus petite taille, les accès dynamiques font l'objet d'une procédure de sélection simplifiée et agile⁷⁷. Elle permet un accès quelques jours seulement après le dépôt du dossier, lorsque ce délai peut atteindre jusqu'à cinq mois pour les accès réguliers. Plus souple, ce mécanisme dynamique d'attribution a été plébiscité par la communauté en IA. Le nombre de demandes d'accès dynamiques est ainsi en progression, s'élevant à 370 pour les projets IA en 2020, et 574 en 2021 (cf. graphique n° 10).

⁷⁷ Alors que dans le cas des accès réguliers, les projets font l'objet d'une évaluation par un collège d'experts au sein de comités thématiques lors de campagnes de dépôts organisées deux fois par an, les demandes d'accès dynamiques sont ouvertes en permanence et ne nécessitent qu'une validation du directeur de l'IDRIS.

Les conditions de mise en place de la plateforme ont cherché à faciliter son appropriation par la communauté scientifique. L'accès au supercalculateur est gratuit, sous condition de recherche ouverte avec publication des résultats. Dans le cadre de la stratégie nationale de recherche en IA, des ressources humaines ont été mises à disposition des utilisateurs de la plateforme pour les accompagner face aux problématiques qu'ils pourraient rencontrer⁷⁸. Des procédures simplifiées d'accès aux ressources ont été mises en place par l'IDRIS *via* l'allocation d'« accès dynamiques ».

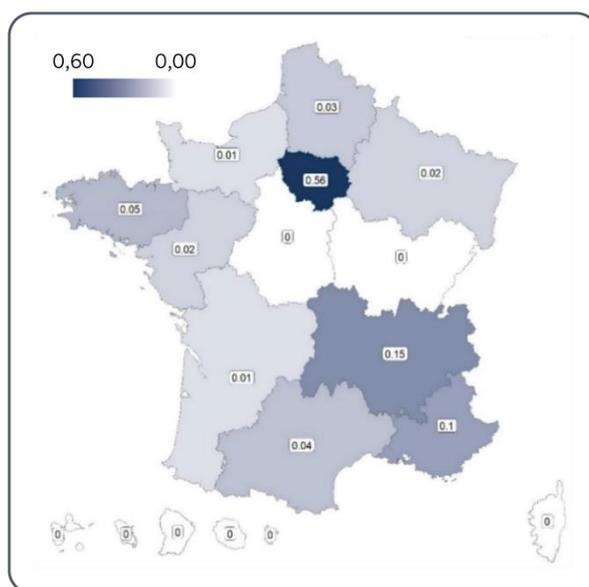
Graphique n° 10 : part des porteurs de projets parmi les organismes de recherche publique associés sur les 574 dossiers de demandes d'accès dynamiques pour des projets IA traités en 2021



Source : Cour des comptes d'après les données fournies par GENCI

⁷⁸ En avril 2022, l'équipe de support consacrée à l'IA créée à l'IDRIS était constituée de 13 ingénieurs.

Carte n° 3 : poids des régions métropolitaines et ultra marines dans le recours aux calculs par le supercalculateur Jean Zay



Source : Cour des comptes d'après données fournies par GENCI

La montée en charge de l'usage de la plateforme a été rapide : sur les plus de 1 400 projets de recherche ayant bénéficié d'une allocation sur « Jean Zay » depuis 2019, 70 % sont en lien avec l'IA. La machine est arrivée à saturation après seulement 18 mois de mise en service⁷⁹. Pour répondre à cette forte demande, trois extensions ont été réalisées afin d'augmenter les capacités de calcul la machine. Une dernière mise à niveau importante a été effectuée au début de l'année 2022 (+ 8,2 Pflops/s) permettant à « Jean Zay » d'atteindre une puissance de calcul de 36,85 Pflops/s.

Les données transmises par GENCI témoignent d'une diversité des domaines d'usage de l'IA à partir du supercalculateur⁸⁰. Dans leur grande majorité, les porteurs de projets sont des laboratoires de recherche académique. Le supercalculateur compte néanmoins une soixantaine de structures de recherche privées parmi les utilisateurs (start-up, petites et moyennes entreprises, grands

⁷⁹ Depuis mars 2021, la consommation des chercheurs en IA dépasse (et représente parfois près du double) les ressources du supercalculateur consacrées à l'IA (les cycles non utilisés par les projets HPC étant alors reversés aux projets IA, et inversement le cas échéant).

⁸⁰ De manière non exhaustive, on retrouve notamment des projets dans les domaines de la vision, du traitement automatique des langages, de l'aide à la décision, de la santé, des « villes intelligentes », de la robotique, des sciences sociales ou encore de la chimie.

groupes). « Jean Zay » permet ainsi à l'ensemble de ces acteurs, y compris les industriels dans des domaines stratégiques, de disposer d'une solution souveraine de calcul. Il existe enfin une relative dispersion géographique des projets déposés, avec une concentration marquée en Île-de-France, en régions AURA⁸¹, PACA⁸², Occitanie et Bretagne reprenant sensiblement les implantations géographiques des pôles d'excellence (cf. carte n° 3).

Ces éléments montrent que la mise en place d'un tel supercalculateur dans le cadre de la stratégie nationale en IA était une mesure pertinente, répondant à un réel besoin. Les résultats d'une consultation menée dans le cadre de l'enquête auprès de la communauté de chercheurs en IA confirment ce bien-fondé. L'ordinateur personnel reste l'outil privilégié des scientifiques en IA dans le cadre de leur activité de recherche, mais 20 % des répondants utilisent le supercalculateur. Les chercheurs qui ont eu recours à cette infrastructure en sont globalement satisfaits (93 %). L'information autour de la plateforme pourrait encore être améliorée, plus de 30 % des répondants indiquant ne pas avoir connaissance du supercalculateur. Une telle acquisition, ou l'augmentation de leurs capacités de calcul lorsqu'ils en sont déjà dotés, est d'ailleurs une constante des stratégies des pays examinés dans le cadre de cette évaluation.

B - Les effets du supercalculateur sur l'attractivité et la crédibilité de la France en matière d'IA

Sans que le phénomène soit précisément quantifiable, la mise en place du supercalculateur « Jean Zay » a pu contribuer à renforcer l'attractivité de la France en matière d'IA. En 2021, 26 % des porteurs des 574 projets IA déposés en accès dynamiques sont de nationalité étrangère, travaillant dans des laboratoires ou structures de recherche française. Le supercalculateur permet également de répondre aux besoins d'initiatives internationales d'ampleur, impliquant des modèles complexes qui nécessitent de grandes capacités de calcul. Le projet *BigScience*, porté par plus de 1 000 chercheurs, a ainsi reçu une allocation de cinq millions d'heures de calcul pour entraîner un des plus importants modèles de langues au monde, ouvert et multilingue (42 langues dont le français).

⁸¹ Auvergne-Rhône-Alpes.

⁸² Provence-Alpes-Côte d'Azur.

Conformément aux exigences de sécurité induites par la politique de sécurité des systèmes d'information de l'État⁸³ et par le dispositif de protection du potentiel scientifique et technique de la nation⁸⁴ (l'IDRIS étant une zone à régime restrictif), l'accès au supercalculateur doit cependant respecter certaines contraintes de sécurité. Cet environnement règlementaire, dont le bien-fondé n'est pas remis en cause pour les infrastructures sensibles de ce type, est néanmoins susceptible de limiter *de facto* son usage par des chercheurs étrangers. L'utilisation de la machine comporte donc des limites comme levier pour attirer les talents.

« Jean Zay » a également pu renforcer la crédibilité de GENCI comme potentiel partenaire de l'entreprise commune européenne pour le calcul à haute performance, EuroHPC⁸⁵. Dans le cadre d'un AAP d'EuroHPC, la France s'est en effet positionnée afin d'accueillir un supercalculateur exaflopique (dit de classe *exascale*)⁸⁶. L'État a déjà sécurisé 200 M€ pour cette opération, notamment *via* deux contributions extrabudgétaires de 80 M€ (PIA 3) et 57,6 M€ (issus de la mesure « Équipement » de la loi de programmation de la recherche pour les années 2021-2027). Ce supercalculateur *exascale* aura également une architecture et des usages convergés. L'expérience de « Jean Zay » a ainsi d'ores et déjà permis en France une plus grande convergence des communautés HPC et IA et de leurs besoins.

Un « Programmes et équipements prioritaire de recherche » (PEPR) exploratoire « Numérique Hautes Performances pour l'Exascale » (NUMPEX) piloté par le CEA, le CNRS et Inria a par ailleurs été validé à hauteur de 40,3 M€ en juillet 2022. L'objectif est d'accompagner le développement de logiciels adaptés qui équiperont les futures machines *exascale*, et de préparer les grands domaines applicatifs aussi bien pour la recherche scientifique que pour le secteur industriel.

⁸³ Circulaire n° 5725/SG du 17 juillet 2014.

⁸⁴ Dispositif visant à protéger les accès aux savoirs, savoir-faire et technologies les plus sensibles des établissements publics et privés émanant du cadre du juridique issu notamment des articles R. 413-7 et R. 413-5-1 du code pénal, du décret n° 2011-1425 du 2 novembre 2011, de l'arrêté du 3 juillet 2012 relatif à la protection du potentiel scientifique et technique de la nation et de la circulaire n° 3415/SGDSN/AIST/PST du 7 novembre 2012.

⁸⁵ Structure juridique et financière qui a pour but de permettre à l'Union européenne, aux pays participants et à des partenaires privés de coordonner leurs efforts afin de renforcer les capacités de l'Europe en matière de calcul à haute performance.

⁸⁶ Telle qu'envisagée actuellement par les pouvoirs publics, quelle que soit son issue, cette candidature ne remet pas en cause le renouvellement de Jean Zay qui devrait intervenir en tout état de cause en 2026.

Les infrastructures de calcul illustrent la nécessité d'inscrire davantage la stratégie française dans une approche européenne, évoquée au chapitre 3. Le calcul intensif est un domaine où les progrès sont continus, et extrêmement rapides : alors que son niveau de performance n'a cessé de croître depuis sa mise en place⁸⁷, « Jean Zay » est passé de la 42^e à la 114^e place dans le Top 500 des supercalculateurs les plus puissants au monde entre juin 2019 et juin 2022⁸⁸. Si la candidature française à l'AAP EuroHPC aboutit, elle conduira, couplée à l'initiative NUMPEX, à la constitution d'un ensemble diversifié d'outils, de logiciels et d'applications (incluant la formation) en matière d'IA.

IV - Un écosystème national encore structuré autour des acteurs académiques et des administrations nationales

L'intelligence artificielle n'étant pas une discipline académique, la structuration formelle de la communauté de recherche a constitué un des objectifs de la stratégie. L'objectif est de renforcer les relations entre les différentes disciplines scientifiques et avec les acteurs de la sphère socio-économique.

Pour analyser ces relations, la Cour s'est largement appuyée sur des analyses de réseaux à partir de sources complémentaires (voir encadré).

Mesurer la structuration de la stratégie nationale sur l'IA par des analyses de réseaux

Compte tenu de l'objectif de la SNRIA d'assurer une visibilité de la recherche au niveau international et de bâtir un écosystème en IA, des analyses de réseaux ont été effectuées à partir de deux sources complémentaires : les liens scientifiques entre institutions co-autrices d'articles scientifiques et les liens des sites web institutionnels des acteurs de l'IA en France.

L'emploi de réseaux permet de cartographier les écosystèmes de l'IA sous différents angles tels que les collaborations scientifiques ou les relations d'influence sur le web et de donner des mesures de l'importance des acteurs et groupes qui les constituent sur des fondements aussi objectifs que possible. Les sources, modalités de construction détaillées de ces réseaux, et les indicateurs associés sont précisées en annexe n° 8.

⁸⁷ Depuis sa mise en marche, le supercalculateur est passé d'une capacité de 13,9 à 36,8 pétaflops.

⁸⁸ <https://www.top500.org/>

S'agissant de la construction des réseaux web, la Cour a pu bénéficier de l'accompagnement du Media Lab de Sciences Po Paris, dont le directeur scientifique a été associé au comité d'accompagnement de l'évaluation. Le Media Lab a développé un outil appelé Hyphe qui permet de réaliser des fouilles automatisées et systématiques des pages internet afin de récupérer tous les sites web vers lesquels ils renvoient via des liens.

Les sites internet nationaux en lien avec la SNRIA ont été classés par la Cour en différentes catégories (administrations nationales, structures de recherche, sites d'excellence, instituts 3IA, etc.) et l'outil Hyphe a permis de disposer de l'ensemble des liens mutuels entre ces sites. Hyphe permet d'identifier, d'une part, des nouveaux acteurs dans le domaine de la recherche, de la formation et de la valorisation participant à la structuration de la communauté nationale d'IA et, d'autre part, les redirections vers des acteurs internationaux.

À titre de comparaison, la même méthodologie a ensuite été suivie pour établir le réseau allemand en partant des cinq centres équivalents aux instituts 3IA français. La sphère d'influence géographique locale de chacun des instituts 3IA a également été représentée.

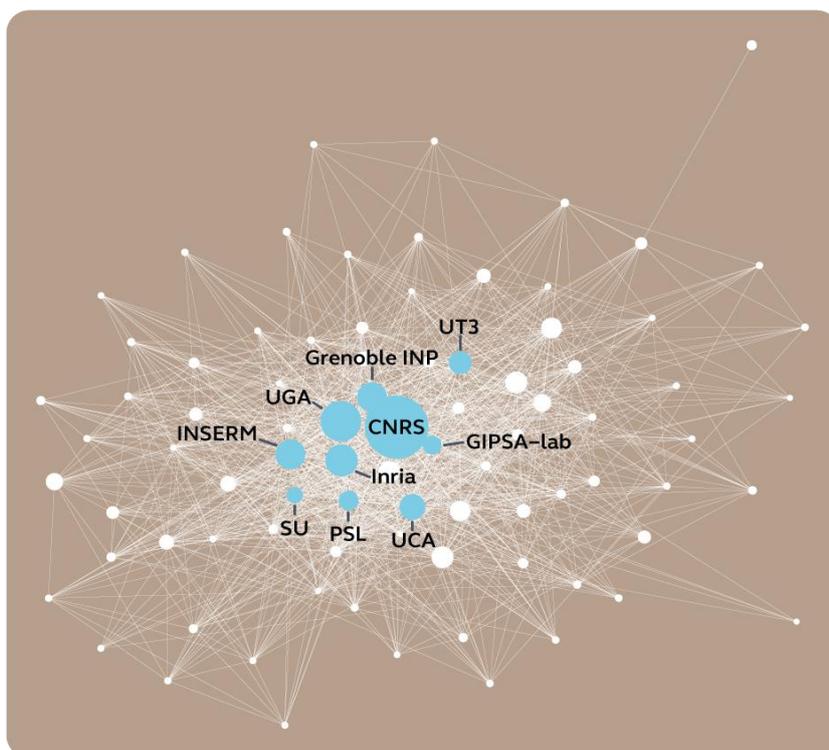
Les réseaux résultants sont représentés graphiquement sous deux formes :

- une représentation graphique 2D disponible dans le corps et les annexes du rapport ;
- une représentation graphique 3D disponible sur le site de la Cour des comptes.

A - Un rôle encore prépondérant des organismes de recherche, sites universitaires et administrations dans la structuration de l'écosystème

Les effets structurants des 3IA ont été examinés à travers une analyse du réseau de collaborations scientifiques retracées dans les quatre archives ouvertes pluridisciplinaires HAL des 3IA (cf. graphique n° 11 et tableau n° 16 ainsi que l'annexe n° 8).

La figure et le tableau des indicateurs associés ci-dessous montrent que les organismes nationaux de recherche (CNRS, Inria, Inserm) occupent une place centrale au sein du réseau de collaboration scientifiques des quatre instituts 3IA.

Graphique n° 11 : réseau de collaborations scientifiques des 3IA

Source : Cour des comptes, à partir des données HAL des 3IA transformées sous la forme d'un réseau de collaborations scientifiques

Note de lecture : La surface des nœuds et la taille des noms des entités reflètent leur fréquence dans les articles scientifiques. Par souci de lisibilité, seuls les 10 % des entités les plus importantes au sens du nombre d'articles copubliés sont représentées. Les entités en cyan correspondent à celles indiquées dans le tableau n° 15. Les autres entités sont indiquées en gris.

Tableau n° 16 : les dix entités les plus centrales au sens du nombre de liens de collaborations parmi les 852 entités répertoriées dans les archives HAL des 3IA

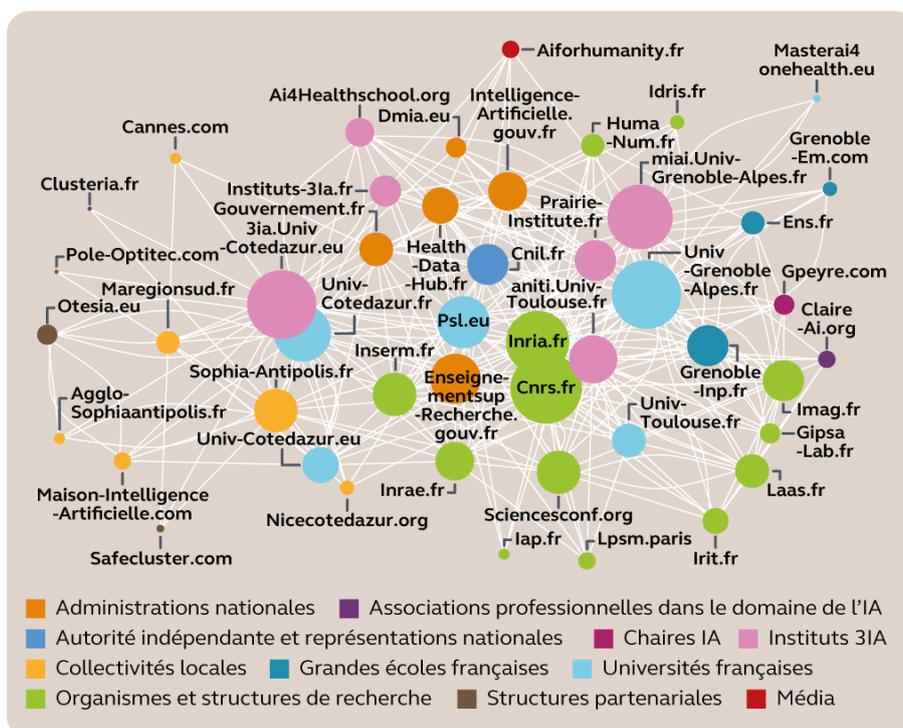
<i>Entité répertoriée dans HAL</i>	Nombre de liens dans le réseau	Part des liens	Intensité des liens du réseau (nombre d'articles copublés)	Part de l'intensité des liens
<i>CNRS</i>	778	3 %	175	12 %
<i>Université Grenoble Alpes (UGA)</i>	437	2 %	89	6 %
<i>Inria</i>	435	2 %	61	4 %
<i>Institut polytechnique de Grenoble</i>	369	2 %	53	4 %
<i>Inserm</i>	340	1 %	53	4 %
<i>Paris Sciences Lettres</i>	296	1 %	25	2 %
<i>Université Toulouse 3 (UT 3)</i>	216	1 %	32	2 %
<i>Sorbonne université (SU)</i>	215	1 %	17	1 %
<i>Université Cote d'Azur (UCA)</i>	211	1 %	41	3 %
<i>Ensemble du réseau</i>	23 402	100 %	1 474	100 %

Source : Cour des comptes, à partir des données HAL

Pour analyser les relations institutionnelles ainsi que la structuration de la visibilité de l'écosystème aux échelles nationale et internationale, une analyse de réseau a été effectuée à travers les liens hypertextes entre les sites web officiels de ces institutions (cf. annexe n° 8) ainsi qu'un sous-réseau ne résultant que des instituts 3IA à fins de comparabilité internationale (cf. graphique n° 12).

Le graphique n° 12 montre les interactions à l'œuvre entre la recherche, la formation et les pouvoirs publics, ainsi que le rôle clé joué par le CNRS et Inria.

Graphique n° 12 : représentation du sous-réseau français résultant des instituts 3IA



Source : Cour des comptes d'après le web et l'outil Hyphe de Sciences Po Paris

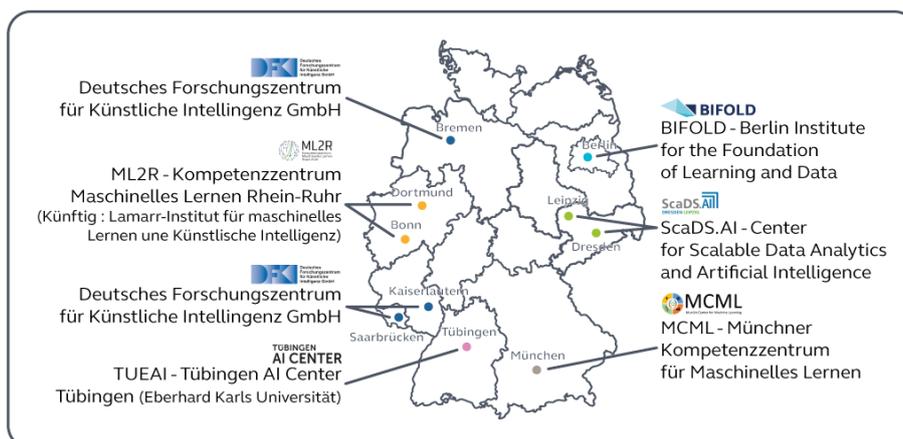
Note de lecture : Les indicateurs employés pour quantifier les importances globale et locale sont précisés en annexe n° 8.

Les indicateurs de centralité calculés à partir de ce réseau et présentés en annexe n° 8 indiquent que les organismes de recherche ont une position centrale dans l'écosystème français de l'IA (contribution à hauteur de 33 % dans la structuration d'ensemble du réseau), aux côtés des universités et des grandes écoles (contribution cumulée de 28 %) ainsi que de certaines administrations (contribution de 11 %). Les instituts 3IA (6 %) et les trois autres centres d'excellence en IA (3 %) jouent pour l'heure un rôle secondaire dans la structuration de la communauté. Les relations avec le monde économique demeurent faibles (3 %).

B - Les instituts interdisciplinaires allemands, plus récents et plus efficaces en matière de structuration de l'écosystème d'IA

Le développement et le renforcement d'instituts interdisciplinaires (ou « centres de compétence ») allemands était l'un des objectifs de la stratégie allemande. Outre le Centre allemand de recherche en intelligence artificielle (DFKI), cinq centres interdisciplinaires pour la recherche en IA ont été identifiés (cf. carte n° 4).

Carte n° 4 : carte d'implantation des six centres de compétences en intelligence artificielle allemands et du DFKI (centre de recherche fondamentale en intelligence artificielle équivalent à Inria)



Sources : Ministère fédéral allemand pour l'éducation et la recherche. Berlin : BIFOLD ; Dresde-Leipzig : ScaDS.AI ; Dortmund-St-Augustin : ML2R ; Munich : MCML ; Tübingen : AI center et le DFKI à Brême, Kaiserslautern et Saarbrücken

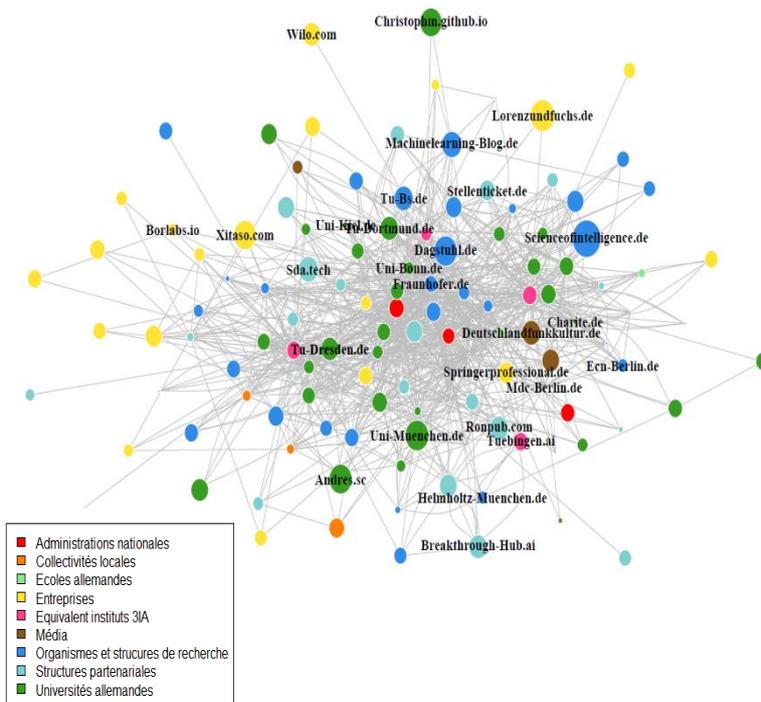
Pour caractériser les effets de structuration de l'écosystème induits par les 3IA et à titre de comparaison, les écosystèmes allemands ont également été analysés et cartographiés par des analyses de réseaux. Le réseau allemand apparaît ainsi plus connecté à l'international que le réseau français et plus efficace en matière de structuration de l'écosystème en IA (cf. graphique n° 13 en regard du graphique n° 12). La comparaison des réseaux institutionnels de la France et de l'Allemagne, au périmètre strict des 3IA et de leurs homologues allemands (cf. annexe n° 8), met en évidence que l'écosystème français ne représente qu'un tiers environ de la taille de l'écosystème allemand, tant en nombre d'acteurs qu'en nombre de liens (cf. tableau n° 17).

Tableau n° 17 : comparaison des réseaux français et allemand au périmètre des instituts 3IA et de leurs homologues allemands

Critère de comparaison	Réseau français	Réseau allemand	Poids relatif de la France/Allemagne
Nombre d'acteurs	47	127	37 %
Nombre de citations entre acteurs	278	992	28 %

Source : Cour des comptes

Graphique n° 13 : représentation du réseau allemand résultant des instituts équivalents aux 3IA



Source : Cour des comptes d'après le web et l'outil Hype de Sciences Po Paris

À l'échelle des catégories d'acteurs, les liens avec les structures partenariales et les entreprises sont davantage développés en Allemagne⁸⁹, ce qui témoigne d'échanges plus importants entre le monde académique et le monde socio-économique.

⁸⁹ Centralité globale de 18 % en Allemagne contre 7 % en France pour les structures partenariales. Centralité globale de 9 % contre 0 % pour la France s'agissant des entreprises.

Pour mesurer les liens entre les instituts français et les centres allemands, les citations entre sites web institutionnels impliquant les 3IA et les centres allemands ont été analysées. Conformément à l'objectif de renforcement des coopérations bilatérales entre la France et l'Allemagne affiché par la SNRIA, un certain dynamisme des collaborations institutionnelles s'observe entre les instituts 3IA français et les instituts interdisciplinaires allemands (cf. graphique en annexe n° 8).

Pour ce qui concerne le rayonnement international et d'influence, les instituts 3IA français sont moins bien positionnés que leurs homologues allemands. Comme le montrent les graphiques pour la France et pour l'Allemagne en annexe n° 8), les liens sortant du réseau de sites webs institutionnels allemand sont bien plus nombreux que ceux du réseau français.

C - Des « communs » mis en place pour renforcer la communauté

De manière simultanée au déploiement de la SNRIA, des « communs »⁹⁰ – ou ressources partagées – sont mis en place et peuvent également à terme renforcer la communauté en IA. Cette approche « outils-ouverts » est développée notamment par les opérateurs de recherche. À titre d'exemple peuvent être évoquées certaines initiatives menées par le CNRS ou par Inria.

Concernant la recherche, le CNRS a lancé en 2021 un centre « IA pour la Science et Sciences pour l'IA » (AISSAI) afin de renforcer la contribution de l'IA à la performance de sa recherche. L'objectif est de participer activement à la construction des interactions entre l'IA et les sciences pour accompagner la mobilisation de l'IA dans tous les domaines scientifiques. Accessible pour toute la communauté du CNRS, il pourrait permettre une plus grande fertilisation entre l'IA et les autres sciences.

Centre IA pour la science et sciences pour l'IA - CNRS

Gouvernance

- Comité scientifique international (ISAB) : formé de 10 à 15 scientifiques sur le spectre thématique du centre (un à deux représentants par institut du CNRS). L'ISAB fournit les conseils, les orientations et les recommandations scientifiques au bureau exécutif et au comité de pilotage sur les principales orientations et priorités scientifiques et programmatiques d'AISSAI.

⁹⁰ Les « communs » ou *commons* en anglais renvoient à des ressources matérielles ou immatérielles partagées et gérées de manière collective dans la perspective de faciliter leur utilisation ou leur accès par le plus grand nombre.

- Comité de pilotage (Copil) : composé de membres nommés par chaque institut du CNRS, et qui sert d'interface entre les instituts et leurs communautés de recherche et le centre. Installé depuis 2021, le Copil tient des réunions bimensuelles.
- Bureau exécutif : composé d'une direction, d'une direction adjointe et d'un chargé de projet, il assure la gestion opérationnelle au jour le jour du centre, en lien étroit avec le Copil et l'ISAB.

Animation scientifique

- Deux équivalents temps plein travaillé (ETPT) supports Ingénieurs IA mis à disposition du centre dans le cadre de la SNRIA.
- Deux programmes thématiques d'un mois chacun programmés en octobre et novembre 2022.
- Une campagne de recrutement de « *fellows* ou sociétaires ».

Concernant la valorisation de la recherche, Inria a, pour sa part, lancé en 2019 *Inria startup studio*, qui a pour but d'accompagner des projets de start-up technologiques intensives en logiciel issues de la recherche publique. Ce programme s'inscrit dans un partenariat avec Bpifrance et permet de financer et d'accompagner des projets issus ou non de l'institut. A partir de 2023, Inria a l'ambition de soutenir *via* ce dispositif plus de 100 projets de startups *deeptech* numériques par an.

Concernant les outils communs, l'opérateur prend également part de manière active à l'initiative *Scikit-learn*⁹¹, bibliothèque en *open source* d'apprentissage statistique en Python⁹². Cinq personnes à temps plein au sein d'Inria sont mobilisées sur le projet. Ce logiciel libre développé par de multiples contributeurs experts en statistiques, algorithmique ou encore en réalisation logiciel est aujourd'hui utilisé pour de nombreuses applications en IA ou en sciences des données. D'après un sondage indépendant réalisé par une filiale de Google auprès d'environ 24 000 répondants dans 173 pays⁹³, *Scikit-learn* se place en position de leader des librairies en apprentissage automatique.

Dans le cadre de la première phase, ces dispositifs n'étaient pas directement reliés à la stratégie, mais venaient en appui. Recentrée sur la valorisation et le transfert, la seconde phase de la stratégie – ou stratégie

⁹¹ L'initiative a été lancée en 2007 par des membres de la communauté Python. Elle est aujourd'hui développée par de nombreux contributeurs issus du monde académique, et utilisée quotidiennement par des millions de *data scientists*. *Scikit-learn* figure au 4^e rang des étoiles GitHub.

⁹² Langage informatique de programmation.

⁹³ Voir <https://www.kaggle.com/kaggle-survey-2022>. Kaggle, filiale de Google, constitue une plateforme permettant aux data scientifiques de publier et réemployer des jeux de données et de participer à des compétitions en science des données.

d'accélération en IA – lancée pour la période 2022-2025 pourrait avoir un effet de levier sur ces initiatives, sans que des financements spécifiques leur soient pour autant directement alloués. Le dossier de presse de la stratégie d'accélération fait ainsi directement référence au projet *Scikit learn*.

V - Une gouvernance affaiblie avec le temps et fragmentée à laquelle il faut mieux associer les parties prenantes

Les enjeux, la multiplication des outils de financement publics et privés et la complexité de l'écosystème d'acteurs rendent cruciale la question de la gouvernance de la stratégie nationale en IA. Au niveau global, elle est pilotée et coordonnée par un coordonnateur national accompagné d'un adjoint. Cette fonction a d'abord été assurée par M. Bertrand Pailhès (entre février 2018 et novembre 2019), placé au sein de la direction interministérielle du numérique et du système d'information et de communication (DINSIC), puis par M. Renaud Vedel, placé au sein de la direction générale des entreprises (DGE) entre février 2020 et juillet 2022.

Le rôle du coordonnateur est de constituer le point de référence de la stratégie pour les interlocuteurs de l'État, de coordonner sa mise en œuvre, tant au niveau national qu'international, et de garantir la cohérence globale des actions conduites. Un comité de pilotage interministériel réunit par ailleurs le Secrétariat général pour l'investissement (SGPI), les neuf ministères plus particulièrement impliqués dans la mise en œuvre de la stratégie⁹⁴, un représentant de la direction interministérielle du numérique, Inria en tant que coordonnateur du volet « recherche » et un représentant du *Health data hub*⁹⁵. Les réunions de cette instance sont prévues à une fréquence mensuelle, ce qui permet un pilotage réactif de la stratégie.

⁹⁴ La lettre de mission de l'ancien coordonnateur national M. Vedel, qui date de février 2020, indique que la stratégie relève du champ de neuf ministères : le ministère de l'économie et des finances, le ministère de l'enseignement supérieur, de la recherche et de l'innovation, le ministère des solidarités et de la santé, le ministère de la transition écologique et solidaire, le ministère de l'action et des comptes publics, le ministère de l'Intérieur, le ministère du travail, le ministère des armées et le ministère de l'Europe et des affaires étrangères.

⁹⁵ Créé par la loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé, *Health data hub* est un groupement d'intérêt public (GIP) qui associe 56 parties prenantes, en grande majorité issues de la puissance publique (CNAM, CNRS, Haute Autorité de santé, France Assos Santé, etc.). Le GIP met en œuvre les grandes orientations stratégiques relatives au Système National des Données de Santé (SNDS) fixées par l'État et notamment le ministère de la santé et de la prévention.

Depuis octobre 2021 et le lancement de la stratégie d'accélération, un comité stratégique a été mis en place, composé également de représentants des administrations et des opérateurs impliqués. Jusqu'au départ du coordonnateur national en juillet 2022, appelé à d'autres fonctions, des réunions de pilotage et de suivi se tenaient de manière hebdomadaire entre le coordonnateur national et la coordinatrice du programme national de recherche en IA d'Inria. Des entretiens entre la coordinatrice du volet recherche pour Inria et le MESR ont également lieu chaque semaine. Il existe donc des canaux de communication directs et réguliers entre ces trois acteurs.

Le coordonnateur bénéficie certes d'un appui des administrations, les services du Premier ministre, BPI France et l'ANR entre autres. Mais la question du rôle et des moyens de la coordination nationale est posée. Il n'est pas en mesure, par exemple, de produire un rapport annuel d'étapes du déploiement de la stratégie, de suivre les financements ou encore de mesurer les effets de la stratégie.

Il devait assurer le suivi et la consolidation d'indicateurs de mise en œuvre, mais l'utilisation de la maquette de performance a progressivement été abandonnée, sans que d'autres mesures de performance soient élaborées en substitution. Il est pourtant indispensable que soit élaborée une liste d'objectifs prioritaires du plan IA et les indicateurs de résultats associés à suivre en comité de pilotage, en tenant compte des priorités européennes et en s'appuyant sur les administrations, le SGPI⁹⁶, Bpifrance France et l'ANR entre autres.

Il résulte de ces observations que le pilotage de la stratégie nationale s'est affaibli avec le temps et qu'il court le risque d'être inopérant.

S'agissant du volet recherche coordonné par Inria, la lettre de mission confiée à son président directeur général (cf. annexe n° 2) précise que doit être mise en place une « *gouvernance adaptée pour le programme de recherche en IA* », avec notamment un « *comité stratégique* ». Le choix a finalement été fait de ne pas créer d'instance de gouvernance spécifique et cette fonction a été dévolue au comité de pilotage global de la stratégie. Inria y est cependant le seul opérateur de recherche représenté et aucun établissement d'enseignement supérieur et de recherche n'y est inclus.

⁹⁶ Le SGPI est le pilote des financements du 4^e programme d'investissements d'avenir et France 2030.

Sans remettre en cause le pilotage et la coordination du volet recherche par Inria, la seconde phase de la stratégie doit être l'occasion d'une gouvernance renouvelée et plus inclusive de ce volet afin de l'articuler avec les enjeux de formation. Cet enjeu est d'autant plus prégnant que les COP de plusieurs organismes, aux premiers rangs desquels Inria⁹⁷, le CNRS⁹⁸ et le CEA⁹⁹ énoncent clairement leurs intentions en matière d'IA. Ces trois acteurs ont par ailleurs convenu de renforcer leur coopération dans le cadre d'Allistène et pilotent conjointement le Programme et Équipement Prioritaire de Recherche de la stratégie d'accélération en IA.

Compte tenu des liens entre recherche et formation, il est essentiel d'être en capacité d'associer les universités et les grandes écoles, notamment en associant France Universités dans la gouvernance du dispositif. En retrait lors de la première phase de la stratégie, le ministère de l'enseignement supérieur et de la recherche devrait désormais introduire la thématique de l'IA dans son dialogue stratégique et de gestion avec les établissements d'enseignement supérieur.

⁹⁷ COP 2019-2023 entre l'État et Inria - mention du rôle de coordination du volet recherche de la SNIA, du rôle de prescription dans les moyens de calcul, de partenariat notamment avec l'Allemagne, de coordination légère des instituts 3IA respectueuse des politiques d'établissement.

⁹⁸ COP 2019-2023 entre l'État et le CNRS - mentions introductives sur le rôle du CNRS dans le numérique et énumération des six grands défis sociétaux auxquels le CNRS veut contribuer (IA cité en 3^e position)

⁹⁹ COP 2021 – 2025 - Objectifs assignés (postérieurement à la SNIA) au CEA de préserver les intérêts stratégiques économiques et de souveraineté de la France et de répondre aux préoccupations sociétales, pour contribuer à la transition numérique.

CONCLUSION ET RECOMMANDATIONS

La constitution de pôles d'excellence dans le cadre de la SNRIA a donné lieu à la labellisation d'instituts interdisciplinaires en IA et de chaires individuelles, tandis que d'autres centres d'excellence ont en parallèle été identifiés. Leur performance en matière de production scientifique est comparable au moment où tous les centres sont impliqués dans la seconde phase de la stratégie, ce qui interroge légitimement sur la similitude de leurs missions. En outre, la stratégie a permis de structurer une communauté interdisciplinaire de recherche sur la thématique et de créer des liens, à renforcer, entre les mathématiques, les sciences informatiques et les sciences humaines et sociales, ainsi que de diffuser les applications de l'IA dans l'ensemble des disciplines.

Le supercalculateur convergé « Jean Zay », mis en service en 2019, a connu une appropriation rapide par la communauté scientifique, même si son usage ne répond aux besoins que d'une partie de la communauté. Il a permis une plus grande convergence des communautés HPC et IA, tandis que la France est candidate afin d'accueillir un supercalculateur européen de type exascale dont l'architecture sera aussi convergée.

En appui de la stratégie, des « outils ouverts » ont été déployés par les principaux organismes de recherche. Largement mis à disposition de l'écosystème en IA, ils pourront à terme participer du renforcement de la communauté en IA. Cela explique qu'ils soient pour partie identifiés dans le cadre de la seconde phase comme leviers d'accélération.

À la deuxième question évaluative, la Cour est dès lors amenée à répondre de façon positive : alors que l'IA n'est pas une discipline en tant que telle, la SNRIA a permis d'accélérer la structuration d'un écosystème en la matière qui nécessite encore un travail d'inclusion de tous les acteurs. Elle repose encore largement sur les organismes de recherche, Inria et CNRS. En particulier, le rôle des instituts 3IA et des autres centres d'excellence apparaît encore secondaire dans cette structuration. En comparaison, les instituts interdisciplinaires allemands, plus récents, semblent être plus centraux dans leur écosystème national. La deuxième phase pourrait avoir pour ambition de poursuivre et de rendre plus résiliente la structuration de l'écosystème, notamment autour des centres d'excellence.

Ces éléments conduisent dès lors la Cour à répondre de manière plus nuancée à la troisième question évaluative, qui portait sur l'efficacité et l'efficience des pôles d'excellence. La stratégie a permis de mettre l'accent sur le recrutement de jeunes chercheurs français, avec notamment l'augmentation significative du nombre de thèses en IA. Mais les synergies

entre les centres d'excellence sont encore à renforcer. Une valorisation mutuelle et réciproque de leurs travaux gagnerait à être conduite de manière plus systématique, ce qui contribuerait à accroître leur visibilité, tant au niveau national qu'au niveau international. En raison de l'identification récente des centres d'excellence, et de leur modèle de financement encore à consolider auquel il est important de s'attacher, il est prématuré de donner une appréciation sur l'efficacité de cette mesure.

Une meilleure coordination de l'ensemble de la communauté scientifique, notamment l'association formelle des universités, au travers de France Universités, peut permettre un renforcement de l'écosystème, un meilleur pilotage et une plus grande synergie entre recherche et formation. L'appréciation de la Cour tient également à l'absence originelle de vision intégrée et coordonnée de cette politique publique portée par des chefs de filat, à l'heure où l'articulation avec le volet formation en IA et les effets économiques, industriels, mais également éthiques, sociétaux et frugaux de l'IA sont recherchés et attendus.

La Cour formule les recommandations suivantes :

- 2. préciser les missions respectives des centres d'excellence 3IA et hors 3IA, et clarifier en conséquence les financements pluriannuels qui leur sont alloués (MESR, SGPI) ;*
- 3. établir de manière partagée les objectifs et les indicateurs prioritaires de la politique publique en IA, en lien avec la stratégie européenne (MEFSIN, SGPI) ;*
- 4. créer un comité scientifique et de pilotage auprès d'Inria, co-présidé par France Universités, pour suivre de manière concertée la mise en œuvre de la stratégie et définir les futures orientations stratégiques (MESR, Inria).*

Chapitre III

L'accélération de la stratégie, avec une approche européenne, frugale et de confiance à approfondir

Alors que la première phase de la stratégie a permis de structurer un écosystème de recherche et de mobiliser les institutions nationales en faveur du développement de l'IA, la seconde phase ou « stratégie d'accélération » vise désormais un transfert accru des résultats de la recherche vers le monde économique et une plus grande convergence entre la formation et la recherche. Si le premier volet de la stratégie, sans atteindre les objectifs assignés, a évité un décrochage scientifique de la France depuis 2018, le second volet est déterminant pour améliorer le positionnement de la France en IA dans la compétition mondiale. Concernant plus spécifiquement les enjeux de formation et de recherche, elle affiche notamment comme objectifs de :

- former et financer une cible d'au moins 2 000 étudiants en DUT¹⁰⁰ / licence / licence pro, 1 500 étudiants en master et 200 thèses supplémentaires par an en régime de croisière ;
- placer au minimum un établissement d'excellence dans les meilleurs rangs internationaux (sans en préciser les critères d'évaluation) ;
- recruter 15 scientifiques étrangers d'envergure mondiale d'ici janvier 2024 au travers du dispositif *Choose France*.

¹⁰⁰ Diplôme universitaire technologique délivré par un institut universitaire technologique (en deux ans).

**Tableau n° 18 : financement du deuxième volet de la SNIA –
stratégie d'accélération (en M€)**

<i>En M€</i>	Programme de Recherche	IA décentralisée et embarquée	IA de confiance	Diffusion de l'IA & démonstrateurs d'IA responsable	Compétences et talents	Total
<i>Financement public</i>	134	265	111	259	776	1 545
<i>PIA 4</i>	73	263,5	97,5	123		557
<i>France 2030</i>					700	700
<i>Autres Crédits État et collectivités</i>	61	1,5	13,50	136	76	288
<i>Financement privé</i>		310	105	86	5	506
<i>Union européenne</i>		60	10	16		86
TOTAL	134	635	226	361	781	2 137

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

Trois dispositifs d'investissement dans l'innovation, en complément d'un fléchage de crédits budgétaires étatiques classiques, sont mobilisés (tableau n° 18) :

- le 4^{ème} programme d'investissements d'avenir (PIA 4) lancé en 2020, dont une partie participe du plan de relance (11 Md€ sur les 20 Md€), à hauteur de 557 M€ ;
- le plan d'investissement « France 2030 » (34 Md€ au total) lancé le 12 octobre 2021, à hauteur de 700 M€ ;
- les programmes pluriannuels européens Horizon Europe¹⁰¹ et Digital Europe¹⁰².

Les financements publics alloués à la stratégie d'accélération pour la période 2021-2025 s'élèvent à 1 545 M€, ce qui correspond à un investissement dans l'écosystème similaire à celui de la première phase. Néanmoins, seuls 8,67 % des crédits prévus portent sur le volet recherche fondamentale (au lieu de 29,14 %), tandis que le volet formation en représente désormais 50,23 % (au lieu de 8,38 %).

¹⁰¹ Programme-cadre de l'Union européenne pour la recherche et l'innovation pour la période allant de 2021 à 2027 finançant des projets à hauteur de 95,50 Md€.

¹⁰² Programme de financement de l'UE axé sur l'intégration de la technologie numérique aux entreprises, aux citoyens et aux administrations publiques à hauteur de 7,50 Md€.

Ce changement d'équilibre est largement induit par la mise en œuvre du plan *France 2030* : au-delà des 10 thématiques stratégiques identifiées¹⁰³, le soutien à l'émergence de talents et l'adaptation des formations aux besoins de compétences des nouvelles filières et métiers d'avenir ont été identifiés comme des leviers transversaux de réussite. Il en résulte un investissement dans la formation nettement réhaussé par rapport aux premiers arbitrages de la stratégie, réalisés avant le lancement de ce plan. Les financements totaux qui lui sont accordés représentent en effet 776 M€, dont 500 M€ consacrés à l'excellence et 276 M€ à la massification.

Le volet « excellence » doit permettre de renforcer le positionnement d'établissements d'enseignement supérieur en pointe ou d'excellence sur la thématique de l'IA, et de proposer une « École française de l'IA » (EFELIA, portée par les instituts 3IA) attractive pour les étudiants nationaux et internationaux jusqu'au doctorat. Le volet « massification » vise à développer les compétences en mathématiques et en informatique nécessaires pour traiter de l'IA, ainsi que l'ensemble des approches interdisciplinaires de l'IA. Ce dernier s'adresse à l'ensemble des organismes de formation sur le territoire pour généraliser des offres de formation en IA. Cette approche permet de couvrir l'ensemble de la chaîne de formations, initiale et continue.

S'agissant du volet recherche, le mandat de coordination d'Inria est reconduit. Le principal instrument mobilisé repose sur un PEPR¹⁰⁴, piloté conjointement par Inria, le CNRS et le CEA, et financé à hauteur de 73 M€ par le PIA 4, autour des quatre axes suivants :

¹⁰³ Les 10 objectifs stratégiques : « Favoriser l'émergence d'une offre française de petits réacteurs modulaires (SMR) d'ici 2035 et soutenir l'innovation de rupture dans la filière » ; « Devenir le leader de l'hydrogène vert et des énergies renouvelables en 2030 » ; « Décarboner notre industrie afin de respecter notre engagement de baisser de 35 %, entre 2015 et 2030, les émissions de gaz à effet de serre dans ce secteur » ; « Produire en France, à l'horizon 2030, près de 2 millions de véhicules électriques et hybrides chaque année » ; « Produire en France, d'ici 2030, le premier avion bas-carbone » ; « Investir dans une alimentation saine, durable et traçable afin d'accélérer la révolution agricole et alimentaire » ; « Produire en France au minimum 20 biomédicaments, en particulier contre les cancers et les maladies chroniques, et créer les dispositifs médicaux de demain » ; « Placer la France en tête de la production des contenus culturels et créatifs, et des technologies immersives » ; « Prendre toute notre part dans l'aventure spatiale » ; « Investir dans le champ des grands fonds marins ».

¹⁰⁴ Au sein du volet dit « dirigé » de France 2030, dit « Financement des investissements stratégiques », une action est consacrée au financement de la recherche la plus fondamentale : les programmes et équipements prioritaires de recherche (PEPR). Ces PEPR visent à construire ou consolider un *leadership* français dans des domaines scientifiques liés ou susceptibles d'être liés à une transformation technologique, économique, sociétale, sanitaire ou environnementale et qui sont considérés comme prioritaires au niveau national ou européen. Les PEPR adossés aux stratégies nationales d'accélération, afin d'accompagner une transformation déjà engagée avec des produits, services, usages et acteurs bien identifiés doivent alors permettre la levée de barrières ou de verrous scientifiques liés à cette stratégie. L'État désigne les pilotes associés et le financement alloué. Le ou les pilotes scientifiques décrivent concrètement leur programme dans un document de cadrage.

- l'IA embarquée (composants nanoélectroniques, couches logicielles et interfaces) et l'IA frugale (ou sobre) (en données / en efficacité énergétique) ;
- l'IA décentralisée (complémentarités et alternatives entre les architectures décentralisées dans les composants et le *cloud*) ;
- l'IA de confiance (enjeux d'explicabilité ou d'interprétabilité, limitation des biais ou enjeux de cybersécurité) ;
- l'approfondissement des fondements mathématiques de l'IA permettant de renforcer les interfaces entre les mathématiques et les autres disciplines.

Le PEPR devra également accroître l'implication des industriels nationaux et favoriser l'émergence de start-up, dans la continuité de la première phase. L'écosystème de la recherche et de sa valorisation, comme Allistène¹⁰⁵, a été sollicité après une consultation du milieu socio-économique pour identifier les besoins et les défis scientifiques à résoudre. Le PEPR a fait l'objet d'une note de cadrage formalisée sous une lettre de mission adressée en décembre 2021 aux pilotes (Inria, CNRS et CEA). Des remontées de projets portés par l'écosystème IA ont été réalisées. Les pilotes ont ensuite structuré leur réponse soumise à un comité d'experts (comité scientifique et technologique de programme). Cette proposition devra être validée par le comité exécutif de *France 2030* et le comité de programme ministériel sur le numérique.

Les enjeux sociétaux prennent une importance particulière tant sur les perspectives éthiques que sur l'impact environnemental de l'IA. Cette section vise à répondre à la question évaluative suivante « La stratégie nationale de recherche a-t-elle amélioré la prise en compte des enjeux éthiques (IA frugale et de confiance) ? ».

¹⁰⁵ Allistène, actuellement piloté par Mme Isabelle Herlin, déléguée générale et par ailleurs coordinatrice de la SNRIA au sein d'Inria, a pour but d'assurer une coordination des différents acteurs de la recherche dans les sciences et technologies du numérique, afin d'élaborer un programme cohérent et ambitieux de recherche et de développement technologique. Elle permet d'identifier des priorités scientifiques et technologiques communes et de renforcer les partenariats entre les opérateurs publics (universités, écoles, instituts), tout en créant de nouvelles synergies avec les entreprises. Créée en décembre 2009, Allistène regroupe en tant que membres fondateurs la CDEFI, le CEA, le CNRS, France universités, Inria et l'Institut Mines-Télécom. Ses membres associés sont l'INRAE, l'INRETS et l'ONERA.

I - Des objectifs centrés sur la formation et l'attractivité des talents jusqu'à présent moins prises en considération

La formation dans le domaine de l'IA recouvre deux enjeux complémentaires :

- former les futurs talents de l'IA, tout particulièrement aux niveaux master et doctorat, qui ont vocation à faire carrière dans le monde académique ou dans les autres domaines de l'économie et de la société ;
- diffuser la culture et la pratique de l'IA au bénéfice de l'ensemble des apprenants, en formations tant initiale que continue.

En parallèle, l'attractivité des talents, c'est-à-dire la capacité de recruter les personnes ayant les meilleures compétences du secteur, constitue un élément clé pour maintenir le dynamisme de la recherche, assurer la bonne marche de l'économie, et conserver sur le territoire français et européen des capacités d'innovation et de production d'excellence.

A - Une croissance des talents formés en IA en progression mais qui demeure en deçà des enjeux

Le premier défi est d'être en capacité d'identifier les formations en IA qui n'est pas une discipline et irrigue de nombreux domaines de formation du simple vernis aux enseignements experts. Si des cartographies et codifications de formations en IA ont été utilement proposées par les acteurs interrogés dans le cadre de cette enquête¹⁰⁶, elles demeurent à homogénéiser et à être davantage mises en visibilité pour dépasser l'échelon local¹⁰⁷.

¹⁰⁶ Dans le cadre de son évaluation, un ensemble de cartographies existantes de formations en IA a pu être identifié : codification des formations en IA dans le cadre de « Mon compte formation » (source : voir annexe n° 11), cartographies réalisées par le MESR au niveau national à partir des plateformes Parcoursup et « Trouve mon master » (source : réponse au questionnaire de la Cour), cartographies proposées par les sites web des instituts et centres d'excellence en IA, cartographies réalisées par les établissements du supérieur dans le cadre de la réponse à l'appel à projets « AMI CMA IA fast track » et cartographie des centres d'excellence en IA.

¹⁰⁷ Les codes des formations de « Mon compte formation » sont publics, notamment pour ce qui concerne l'IA (voir www.francecompetences.fr). Quelques cartographies des formations en IA ont été publiées au niveau « local » sur les sites web des organismes concernés, comme dans le cas des 3IA, mais elles n'ont pas été construites sur des critères homogènes et elles ne sont pas fédérées.

Dans le cadre de l'évaluation, les formations tant initiales que continues dans ce domaine ont été identifiées, à partir de la méthode d'analyse par mots clés spécifiques (cf. annexe n° 11).

1 - Une formation initiale en IA en développement, qui ne repose pas exclusivement sur les centres d'excellence

Cette évaluation permet de réaliser un bilan des parcours de formations débouchant sur des compétences en IA et d'identifier la capacité de la France à former des talents et d'étudier le positionnement des centres d'excellence sur le volet formation de la stratégie.

Tableau n° 19 : évolution de l'offre de formation où l'IA est mentionnée de façon explicite aux niveaux Bac+0 (Parcoursup).

<i>Année</i>	Nombre de formations spécifiquement en IA	Nombre total de formations	Part des formations en IA (en %)
2020	2	16 990	0,01
2021	9	19 439	0,05
2022	8	21 347	0,04

*Sources : analyses sémantiques de la Cour des comptes d'après les données ouvertes du MESR relatives à Parcoursup (cartographie des formations de 2020 à 2022 ; mise à jour en décembre 2021)
Note de lecture : Les analyses sémantiques ont été appliquées aux noms longs des formations et aux spécialités et mentions de spécialisation.*

L'essentiel de l'offre de formation initiale ciblée sur l'IA est concentré au niveau du master. La disponibilité des formations au niveau postbac portant spécifiquement sur l'IA est faible (inférieur à 10 soit 0,04 % de l'offre en 2022, voir tableau n° 19).

Sur la période 2019-2021, de manière concomitante à la stratégie, la capacité d'accueil des formations en IA pour une entrée en master au niveau Bac+3 est en progression de + 34 % en moyenne par an, contre + 1 % seulement pour l'ensemble des formations. Cela démontre une dynamique de développement intéressante de l'offre de formation au niveau master, mais sûrement pas encore suffisante pour atteindre les objectifs du second volet de la stratégie en matière de formation de nouveaux étudiants. Elle représente ainsi en 2021 un total de 16 687 places, soit 1,7 % des places proposées toutes formations confondues (cf. tableau n° 20).

**Tableau n° 20 : évolution de l'offre de formation et des capacités d'accueil
dans les formations accessibles au niveau Bac+3 (« Trouver Mon Master »)**

Année	Nombre de formations spécifiquement en IA	Nombre total de formations	Part des formations en IA (en %)	Nombre de places proposées spécifiquement en IA	Nombre total de places proposées	Part des places (en %)
2019	119	8 752	1,36	9 241	993 786	0,9
2020	158	9 189	1,72	12 384	1 021 626	1,2
2021	185	9 078	2,04	16 687	1 011 042	1,7
Évolution annuelle moyenne (%)	25 %	2 %	23 %	34 %	1 %	38 %

Sources : Analyses sémantiques de la Cour des comptes d'après les données ouvertes du MESR relatives à « Trouver Mon Master » (informations destinées au portail « Trouver Mon Master » (TMM) au titre des années universitaires 2019-2020, 2020-2021 et 2021-2022 ; dernière mise à jour des données en juin 2022)

Note de lecture : Les champs suivants ont été employés pour l'analyse sémantique : intitulé de la mention de master, intitulé du parcours ainsi que l'ensemble des mots clés disciplinaires, sectoriels et métiers associés à la formation.

Néanmoins, il est nécessaire et important d'être en capacité d'assurer un flux de talents suffisant dans ces formations de niveau master qui repose sur des formations « amont » apportant les prérequis nécessaires à la pratique de l'IA.

D'après les données de « Trouver mon master », les licences d'informatique, de mathématiques ou d'autres domaines mobilisant fortement le calcul sont les premières conseillées par les établissements pour suivre les enseignements proposés dans le cadre des masters où l'IA est enseignée¹⁰⁸. Aux niveaux licence, licence professionnel et DUT qui regroupent 63 % des inscrits en informatique et mathématiques tous niveaux d'études confondus, les effectifs étudiants ont crû de 4 % par an en moyenne sur la période 2006 à 2020 pour environ 50 000 étudiants formés en 2020 (cf. annexe n° 10). Ceci témoigne un grand dynamisme des filières formant aux prérequis de l'IA qui devra être au moins maintenu voire renforcé pour répondre également aux objectifs du second volet.

La capacité de pourvoir les places dans ces filières de formation du supérieur avec des profils adaptés dépend également des choix faits par les lycéens depuis la réforme du baccalauréat de 2018. En particulier une attention doit être portée dans le futur aux évolutions du recours aux enseignements de mathématiques ainsi qu'au développement de la spécialité numérique et sciences informatiques (NSI) qui aborde de manière plus poussée des algorithmes d'IA (cf. encadré).

¹⁰⁸ Les cinq premières licences conseillées au titre de la session 2021 (dernières données disponibles) pour les masters détectés comme enseignant l'IA suivant la méthode de la Cour sont : informatique (22,1 % des licences conseillées), mathématiques (14,9 %), mathématiques et informatique appliquées aux sciences humaines et sociales (9,5 %), économie et gestion (5,1 %) et physique (4,5 %). Ces cinq licences regroupent plus de la moitié des licences conseillées pour suivre les enseignements des formations de master à l'IA.

Une diffusion croissante mais circonscrite de l'IA dans le secondaire, enjeu de compétence, d'attractivité et de vivier futur

Dans le cadre de la réforme du baccalauréat de 2018, à partir de la rentrée 2019, l'enseignement de sciences numériques et technologie (SNT) a été introduit en seconde générale et technologique, ainsi que la spécialité numérique et sciences informatiques (NSI) en première et en terminale générale. Un algorithme d'intelligence artificielle, « l'algorithme des k plus proches voisins », est au programme de la première NSI¹⁰⁹. D'après la direction générale de l'enseignement scolaire (DGESCO), le volume horaire consacré à l'étude de cette notion est de quatre à six heures, et l'utilisation et l'exploitation de méthodes d'IA peuvent être étudiées dans les projets élaborés par les élèves, soit 30 heures en première et 50 heures en terminale.

Par ailleurs, dans le cadre de l'enseignement obligatoire de terminale générale appelé « enseignement scientifique », l'IA est introduite de façon moins approfondie, avec toutefois une approche large des grands principes et enjeux du domaine qui peut contribuer à sa diffusion dans la société¹¹⁰. Des introductions à l'intelligence artificielle sont parfois proposées au collège et dans l'enseignement professionnel mais elles ne sont pas diffusées sur une base systématique¹¹¹.

La Cour observe que la spécialité NSI était encore très peu diffusée à la rentrée 2021 en comparaison des autres spécialités de sciences « dures » (cf. annexe n° 10). En outre, la spécialité mathématiques, et l'option « mathématiques expertes », matière phare pour l'entrée dans les filières du supérieur qui nourriront les disciplines du cœur scientifique et technique de l'IA, marquent un recul relativement important en 2021 par rapport à 2020 (respectivement - 3,7 points et - 6,7 points). Le déséquilibre femme-homme est également notable.

Ces éléments témoignent de façon concordante d'un risque d'épuisement des viviers des filières scientifiques qui nourrissent l'IA sur le temps long, avec de surcroît un risque majeur de renforcement des biais de genre. La DGESCO a indiqué qu'un ensemble de mesures (cf. annexe n° 10) ont été déployées ou le seront prochainement avec pour objectif d'endiguer ces phénomènes. La Cour veillera à leur mise en œuvre et à leur efficacité.

¹⁰⁹ Voir : <https://eduscol.education.fr/document/30061/download>

¹¹⁰ L'IA est introduite dans le cadre du sous-thème 3.5 relatif à l'intelligence artificielle. Voir : <https://eduscol.education.fr/document/25498/download>

¹¹¹ La DGESCO indique que « au cycle 4 (classes de 5^e, 4^e, 3^e), l'éducation aux médias et à l'information (EMI), en lien avec la pratique dans l'enseignement de technologie, peut proposer une découverte de l'IA. Cette initiation à l'IA est liée à la compréhension de la gestion de données et à l'élaboration d'un prototype d'objet communicant dans le programme de technologie, et à la sensibilisation aux questions éthiques associés à l'utilisation de l'IA pour l'EMI. Cette notion peut être proposée sur un volume horaire de 5 h. Dans l'enseignement professionnel, notamment en filière Systèmes numériques (SN), l'IA n'est pas présente dans les programmes, cependant elle est de plus en plus exploitée comme exemples d'applications industrielles (IA et voiture autonomes, IA et maintenance de systèmes automatisés). »

Ce parcours de formation doit permettre de former des talents d'un niveau de doctorat et ainsi d'assurer la convergence entre formation et recherche. De manière générale, la croissance des thèses en IA est un marqueur de la qualité du parcours amont (« licence – master » à l'université ou « classes préparatoires aux grandes écoles – grandes écoles »), filières qui sont souvent hybridées en IA¹¹²), en particulier dans les filières informatique et mathématique¹¹³.

Une part substantielle des places pour les formations en IA au niveau master est portée par des universités qui ne sont pas associées aux centres d'excellence (cf. tableau n° 21), ce qui montre que des dynamiques séparées peuvent exister entre le développement (voire la massification) de la formation et l'excellence de la recherche.

Tableau n° 21 : liste des dix premiers établissements contributeurs en matière de formés à l'IA au niveau Bac+3 en 2021

<i>Établissement</i>	Nombre de places de formations en IA	Part des places de formation en IA au niveau national (en %)	Nombre de formations en IA	Part des formations en IA au niveau national (en %)
<i>Université Paris-Saclay</i>	4 347	26,1	15	8,1
<i>Université de Lille</i>	1 768	10,6	17	9,2
<i>Université d'Aix-Marseille</i>	1 720	10,3	14	7,6
<i>Université de Bordeaux</i>	1 320	7,9	14	7,6
<i>Université de Strasbourg</i>	1 246	7,5	9	4,9
<i>Université Paris Cité</i>	642	3,8	5	2,7
<i>Université Paris-XII</i>	525	3,1	6	3,2
<i>Université Toulouse-III</i>	506	3,0	3	1,6
<i>Université Lyon-I</i>	296	1,8	3	1,6
<i>Université Rennes-I</i>	284	1,7	5	2,7
Total « top 10 » [A]	12 654	75,8	91	49,2
Total « top 10 » hors instituts 3IA et centres d'excellence [B]	7 159	42,9	68	36,8
<i>Ratio [B / A]</i>	57 %	57 %	75 %	75 %

Sources : analyses sémantiques de la Cour des comptes d'après les données du MESR relatives à « Trouver Mon Master » (informations destinées au portail « Trouver Mon Master » (TMM) au titre des années universitaires 2019-2020, 2020-2021 et 2020-2021 ; dernière mise à jour des données en juin 2022)

Note de lecture : les champs suivants ont été employés : intitulé de la mention de master, intitulé du parcours ainsi que l'ensemble des mots clés disciplinaires, sectoriels et métiers associés à la formation.

¹¹² 20 % des répondants à la consultation de la Cour déclare être passé par les deux systèmes.

¹¹³ 79,8 % des répondants se déclarent rattachés ou proches des sections du conseil national des universités suivantes : informatique (section n° 27), mathématiques appliquées et applications des mathématiques (n° 26) et génie informatique, automatique et traitement du signal (n° 61).

Avec près d'un quart des formés à l'IA chaque année, l'Université Paris-Saclay, hôte de l'Institut DATAIA (institut de convergence hors 3IA), apparaît comme le premier contributeur français, en matière de formation à l'IA au niveau Bac+3 (cf. tableau n° 20). Concernant les 3IA, l'université Paris Cité, co-partenaire de PRAIRIE se positionne au 6^e rang (3,8 %) et l'université de Toulouse 3, co-partenaire d'ANITI, au 8^e rang national (3,0 %). Les partenaires des centres d'excellence contribuent donc de manière hétérogène à la massification de la formation en IA.

Les partenaires développent de plus en plus leurs formations en IA, sans que les centres d'excellence, quel que soit leur statut (3IA ou hors 3IA), soient le catalyseur de cette dynamique, jusqu'à récemment en tout cas.

Ce constat montre l'intérêt de mobiliser lors de la seconde phase de la stratégie les centres d'excellence pour assurer la bonne convergence entre recherche d'excellence et formation de haut niveau, élément d'attractivité pour le programme *Choose France*.

2 - Une formation continue en IA déficiente avec des besoins probablement insuffisamment couverts et une faible mobilisation de l'enseignement public

La formation continue dans le domaine de l'IA a été appréciée à l'aune des données issues de « Mon compte formation » exploitées par la Cour en lien avec la Caisse des dépôts et consignations, qui ne couvrent certes pas l'ensemble du domaine de la formation continue. Mais selon ces données, la formation continue en IA constitue une lacune majeure (cf. annexe n° 11). L'IA représente moins de 0,1 % de l'offre et de la demande de formation continue proposée, tous secteurs confondus, soit moins de 1 000 personnes formées depuis janvier 2020.

Bien que les établissements d'enseignement supérieur publics contribuent à cette offre de formation continue, ils ne couvrent qu'un dixième des apprenants, avec une part des formés en érosion par rapport aux autres organismes de formation (cf. tableau n° 22). Pourtant, la forte progression de la demande (+ 302 % entre 2021 et 2022) suggère que les besoins de formation continue en IA sont loin d'être saturés.

La création des 3IA a pu dans certains cas améliorer le dialogue entre université et entreprises pour adapter l'offre de formation continue aux besoins locaux. Un effort par les salariés est cependant nécessaire, ce qui passe notamment par une amélioration de la connaissance des usages de l'IA.

Tableau n° 22 : évolution comparée de la demande de formation continue en IA (nombre de dossiers de formation validés par les apprenants) dans l'enseignement supérieur public relativement aux autres organismes de formation

<i>Statut des organismes de formation en IA</i>	1 ^{er} semestre 2020	1 ^{er} semestre 2021	1 ^{er} semestre 2022	Évolution 2021 à 2022
<i>Établissement d'enseignement supérieur public</i>	4 (100 %)	31 (20 %)	32 (6 %)	+ 3 %
<i>Autres organismes de formation (secteur privé)</i>	0 (0 %)	123 (80 %)	494 (94 %)	+ 302 %
<i>Tous</i>	4 (100 %)	154 (100 %)	526 (100 %)	+ 242 %

Source : Cour des comptes en lien avec la Caisse des dépôts et consignations à partir des données de « Mon compte formation »

Égalité des genres et formations en IA

Étant donné les répercussions en termes de biais sur les outils de diagnostic et d'aide à la décision déployés dans la société et l'économie, notamment lors des processus d'apprentissage, que cela peut induire dans le développement de l'intelligence artificielle, le déséquilibre structurel entre les femmes et les hommes formés à l'IA est susceptible de renforcer des inégalités de façon plus large dans la société.

Dans les filières d'élection préparant à l'IA dans l'enseignement supérieur, l'écart de représentation en licence est demeuré stable sur la période, à hauteur d'environ 18 % de femmes en moyenne. Les améliorations récentes constatées en matière de parité sur la période 2018 à 2020 (+ 0,6 point par an) sont en réalité entièrement attribuables à l'augmentation de la part des femmes parmi l'ensemble des étudiants toutes filières confondues (+ 0,7 point par an). Si les filières de prédilection de l'IA sont dynamiques, elles sont affectées par un biais de genre durablement installé dans l'enseignement supérieur. Des actions correctrices ne peuvent être efficaces qu'à un stade plus précoce, au niveau de l'enseignement scolaire qui commence à aborder les enjeux de l'IA. Les profils des apprenants formés tout au long de la vie en IA dans les établissements d'enseignement supérieur publics sont également biaisés en termes de genre (un quart de femmes formées seulement).

Les disparités de genre constatées en IA plongent leur racine dès le lycée, dans les principales sciences formelles et expérimentales, à l'exception notable des sciences de la vie et de la Terre (cf. annexe n° 10). La parité dans les filières permettant l'acquisition des compétences mathématiques nécessaires à l'IA présente une trajectoire préoccupante.

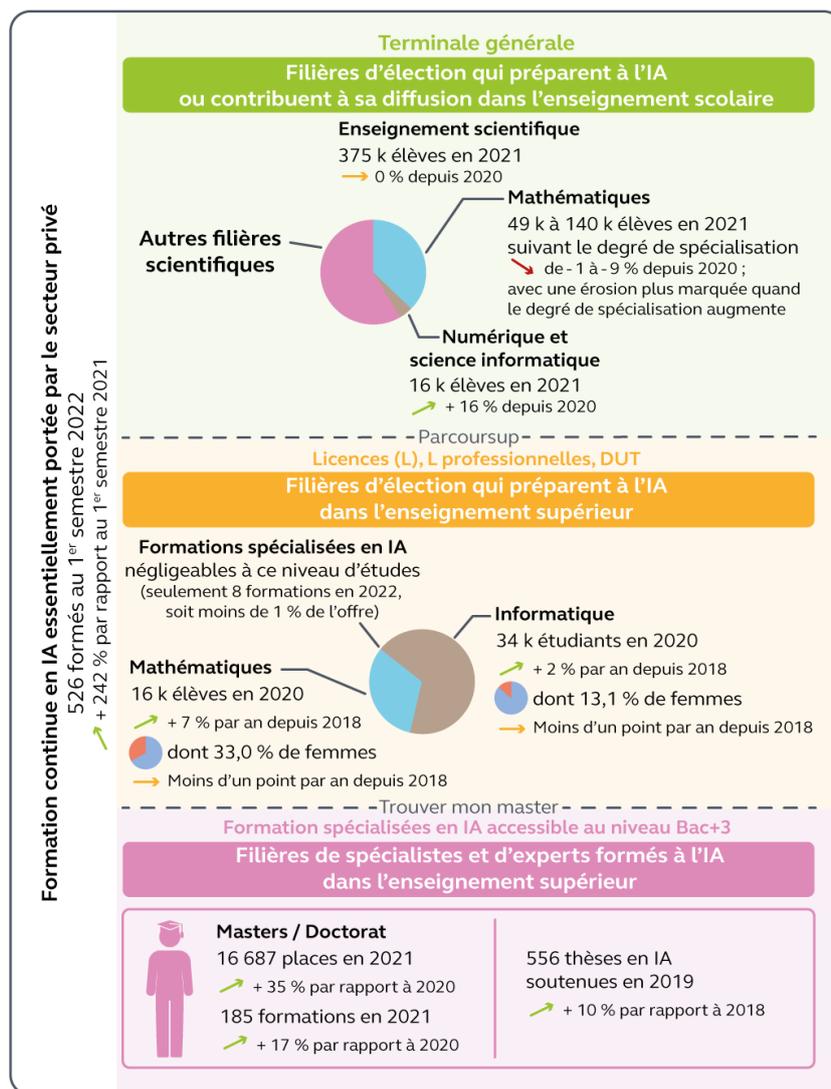
Afin de lutter contre les biais de genre dans l'enseignement scolaire, la DGESCO a indiqué qu'un ensemble de mesures ont été déployées ou le seront prochainement. La Cour sera attentive à leur efficacité.

Les formations initiales (masters, thèses) et continues (mobilisation du compte personnel de formation) à l'IA sont en croissance avec néanmoins une dynamique qui ne permet pas de répondre aux ambitions de la stratégie et le positionnement des centres d'excellence sur cet enjeu doit être renforcé (cf. schéma n° 3).

Si les filières d'élection qui conduisent à ses formations sont dynamiques au niveau licence, elles présentent des signes de fragilités au lycée, tout particulièrement en mathématiques, malgré des efforts du MENJS pour contribuer plus largement à l'enseignement de l'IA aux élèves et à sa diffusion plus large à l'école¹¹⁴. Les forts déséquilibre de genre à l'œuvre dans ce secteur demeurent et ne montrent pas de tendance à l'amélioration (cf. encadré ci-avant).

¹¹⁴ En lien avec la DNE du MENJS et avec le soutien du PIA, des prototypes d'IA ont été déployés avec succès dans l'enseignement scolaire (voir <https://eduscol.education.fr/1911/partenerariat-d-innovation-et-intelligence-artificielle-p2ia>). Le MENJS a indiqué à la Cour que ces prototypes d'IA, après une phase de retours d'expérience des utilisateurs en académie, sont désormais proposés au niveau national sous la forme de services numériques industrialisés.

Schéma n° 3 : cartographie des formations et croissance des effectifs d'apprenants formés à l'IA ainsi que dans ses filières « amont » d'élection



Source : Cour des comptes à partir des données du MENJS, du MESR ainsi que de la Caisse des dépôts et consignations, avec application de filtres sémantiques pour identifier les formations spécifiques à l'IA

Note de lecture : l'ordre de grandeur du nombre total d'experts académiques en IA est estimé à partir des éléments communiqués par les opérateurs de recherche et des universités dans le cadre de la consultation opérée par la Cour (cf. annexe n° 9) ainsi qu'à travers les effectifs de personnels enseignants-chercheurs dans les sections universitaires où l'IA prévaut (cf. annexe n° 12). Le terme « k » représente les milliers.

B - Les ambitions contrariées d'excellence de la recherche et de la massification

1 - Une première impulsion rapide pour augmenter la capacité de formation et faire converger recherche et formation de la deuxième phase

Afin d'organiser une nouvelle convergence entre la formation et l'excellence en recherche, l'appel à manifestations d'intérêt « Compétences et Métiers d'Avenir » (AMI CMA) constitue une des actions de la deuxième phase de la stratégie (cf. encadré).

L'appel à manifestations d'intérêt « Compétences et Métiers d'Avenir » : bras armé de la 2^e phase de la stratégie pour développer les formations en IA

Doté de 2,5 Md€, l'appel à manifestations d'intérêt « Compétences et Métiers d'Avenir » (AMI CMA) vise à répondre aux besoins en matière de formation et de compétences nouvelles pour les métiers d'avenir, les entreprises et les territoires en lien avec le programme *France 2030*. Il s'agit en pratique d'accroître l'offre de formations initiales ou continues, qu'elles soient sanctionnées par des titres, des certifications ou des diplômes, ainsi que leur accessibilité en termes d'information, d'attractivité et d'inscription pour tous les publics. Deux catégories de projets peuvent être soutenues :

- catégorie « Diagnostic » pilotée par la Caisse des dépôts et consignations pour effectuer les diagnostics de besoins en compétences et en formations tout au long de la vie et pour les mettre en perspective avec l'offre existante en France et à l'international. Cela doit déboucher sur un plan d'actions global pour accompagner les évolutions des emplois et des compétences et sur des recommandations auprès des différents acteurs ;
- catégorie « Dispositif de formation » pilotée par l'ANR déclinée en sous-catégories « Enseignement supérieur » et « Enseignement scolaire ou formation continue » en fonction des porteurs de projets et des publics cibles. Cette catégorie vise à mobiliser des outils développés dans le cadre des PIA pour développer des formations professionnalisantes initiales et continues afin d'accompagner les transformations des entreprises et des innovations et d'améliorer l'image et l'attractivité des métiers et des emplois dans la filière. Les projets peuvent mobiliser des formes pédagogiques telles que la digitalisation et l'hybridation pour massifier le nombre de formés.

Les formations en IA ont bénéficié d'un volet thématique spécifique au lancement de l'AMI CMA¹¹⁵. Les sept centres d'excellence ont ainsi pu déposer un dossier de candidature pour la première vague. L'ensemble des centres ont été lauréats à hauteur de 54,57 M€ par un jury de professionnel dédié à l'IA¹¹⁶. Dans ce cadre, le projet commun « École Française de l'IA » (EFELIA) porté par les quatre instituts 3IA qui vise à (i) la labellisation des formations nationales en IA, (ii) la mise en place et la gestion de plateformes de calcul et de challenges à destination des étudiants, (iii) la création de formations pilotes et (iv) la création de contenus pédagogiques (MOOC¹¹⁷) pour l'ensemble de l'écosystème d'enseignement supérieur ressort lauréat. Il est à noter qu'il s'agit de la collaboration de la plus grande envergure entre instituts, ce qui suggère une certaine maturité de ce réseau.

Retenir l'ensemble des centres d'excellence en IA permet aux coordonnateurs de l'appels à projet de s'appuyer sur la structuration amorcée depuis plusieurs années dans le domaine de la recherche en IA et d'abonder le volet formation qui, bien qu'inclus dans les objectifs initiaux de l'appel à projets « Instituts 3IA », ne disposait que de peu de crédits dans la première phase de la stratégie nationale. Cela révèle et conforte la qualité de ces institutions pour porter le développement de l'IA en France. Il sera néanmoins important d'assurer l'articulation et la cohérence entre le projet EFELIA porté par les instituts 3IA et les autres projets portés par les autres centres d'excellence.

En outre, l'AMI est opéré suivant un système de relèves régulières de projets pour une instruction « au fil de l'eau » de l'ensemble de projets déposés, y compris ceux relatifs à l'IA qui n'auraient pas été déposés dans le cadre du volet thématique initial. Une seconde vague a ainsi autorisé des dépôts de candidatures jusqu'au 5 juillet 2022 permettant la relève de six dossiers de type « diagnostic » et de huit dossiers de type « dispositif de formation » en IA¹¹⁸, soit 14 dossiers au total, ce qui permet d'identifier de nouveaux acteurs pour porter la massification de la formation en IA.

¹¹⁵ Ce volet a représenté 18 % des candidatures (voir annexe n° 13).

¹¹⁶ Les montants attribués sont par université : Université Côte d'Azur 8,01 M€ (EFELIA- 3IA Côte d'Azur), Université PSL 8,84 M€ (EFELIA-PRAIRIE), Université Grenoble-Alpes 6,38 M€ (EFELIA-MIAI), Université fédérale de Toulouse Midi-Pyrénées 5,10 M€ (EFELIA – ANITI), Université Paris-Saclay 11,45 M€ (SaclAI School), Sorbonne Université 7,01 M€ (Sorbonne.AI) et Institut Polytechnique de Paris 7,78 M€ (AccelAI Learning@Hi!PARIS). Il est à noter que deux autres projets n'ont pas été lauréats. Cela correspond à 21 % des montants alloués lors de la première vague. Cf. annexe n° 13.

¹¹⁷ *Massive Open Online Course* : formation en ligne ouverte à grande échelle.

¹¹⁸ Cela représente 7 % des dossiers déposés sur le volet « Dispositif de formations ». Cf. annexe n° 13.

Le SGPI a indiqué par voie de presse que le nombre total de dossiers déposés à l'AMI était très faible et que les acteurs de la formation étaient confrontés à des freins importants en matière de locaux ou de personnels¹¹⁹. Les facteurs visant à lever ces verrous et à garantir la réussite de la mise en œuvre de cet AMI CMA peuvent être identifiés :

- la capacité de recrutement de formateurs et de formation des enseignants quel que soit leur discipline dans des délais courts ;
- la capacité à augmenter de manière significative le nombre de formations (taux actuel de croissance de 34 % qui ne permet pas de répondre à l'objectif de doublement à court terme) ;
- le fait de disposer de locaux en quantité suffisante ou de dispositifs adaptés pour les formations à distance ou hybrides ;
- la capacité à cibler tous les publics en assurant une certaine diversité des profils des enseignants ;
- l'articulation et la continuité entre formations initiales et continues ;
- la reconnaissance des formations dans le RNCP (Répertoire national des certifications professionnelles) et le RSCH (Répertoire spécifiques des certifications et habilitations) pour favoriser la diffusion de l'IA dans les métiers.

2 - Des viviers de formateurs de haut niveau contraints en regard des besoins

Les ambitions de la SNIA en matière de formation sont contraintes par de fortes tensions en matière de ressources humaines, mobilisées en partie par l'effort de recherche. Les enseignants-chercheurs et enseignants du supérieur dans les domaines qui recouvrent préférentiellement l'IA ont augmenté dans des proportions bien moindres que les jeunes talents en IA. Sur la période 2010-2019, le taux d'accroissement des enseignants-chercheurs dans ces domaines a été de seulement + 4 % contre + 225 % pour les docteurs en IA. Les postes permanents sont stables autour de 7 000 depuis la mise en œuvre de la SNIA en 2018 (cf. annexe n° 12).

De façon concordante, la consultation auprès de la communauté scientifique en IA révèle qu'en matière de répartition du temps de travail, seulement un quart du temps de travail est consacré à des activités d'enseignement et un peu moins de la moitié à la recherche, notamment du fait de la forte proportion en IA de personnels des organismes de recherche

¹¹⁹ Article *News tank* n° 262201 citant le SGPI.

qui y consacre 10,40 % de leur temps en moyenne, révélant la tension entre recherche et enseignement. Pour les enseignants chercheurs, le ratio de temps passé à la recherche est de 1,15 contre 1 pour l'enseignement, il est de 6 pour 1 pour les chercheurs dans les organismes nationaux de recherche.

Le nombre de formateurs spécialisés dans l'enseignement supérieur public est ainsi sous-dimensionné par rapport aux besoins de formation en IA, tant en formation initiale qu'en formation continue. Pourtant, sur la période 2010-2020, les domaines qui recouvrent préférentiellement l'IA ont été favorisés pour les recrutements (+ 3 % contre - 1 % pour l'ensemble des personnels enseignants ; cf. annexe n° 12). Mais la part des jeunes enseignants-chercheurs est en diminution (- 2,9 points), comme celles des personnels étrangers avec une légère inflexion à la baisse depuis la mise en œuvre de la SNIA (+ 0,6 point depuis 2018).

Cette tendance démontre les limites de l'objectif d'attractivité des jeunes talents en recherche, y compris étrangers, affiché dans la première phase de la stratégie. Elle peut s'expliquer par une priorité trop faible donnée à la formation dans le premier volet à laquelle il importe de remédier dans le second.

En l'absence de personnels de haut niveau formés en quantité suffisante et dans les différents champs disciplinaires afférant à l'IA, les besoins en matière d'IA risquent de ne pas être satisfaits, tant en matière d'excellence que de massification. En particulier, les engagements des établissements dans le cadre de l'AMI CMA risquent d'être difficiles à tenir.

3 - Une gouvernance de la formation lacunaire, une coordination à renforcer au bénéfice de l'ensemble des apprenants

Si la coordination de la stratégie nationale en IA assurée par la DGE¹²⁰ a pour mandat le suivi des enjeux de formation de façon globale et consulte régulièrement les acteurs de la formation initiale¹²¹, il n'existe pas à ce jour un espace de dialogue et d'échange entre tous les acteurs (France Universités, CGE, universités, grandes écoles, ministères) pour construire un référentiel ou un label des formations en IA et mobiliser l'ensemble de l'écosystème de formation afin de massifier le nombre de formés. Les acteurs de la formation continue tels que France compétence, qui gère le répertoire national des compétences professionnelles (RNCP) et le

¹²⁰ Direction générale des entreprises.

¹²¹ Direction du numérique éducatif pour l'enseignement scolaire et Direction générale de l'enseignement supérieur et de l'insertion professionnelle du MESR pour les établissements d'enseignement supérieur.

répertoire spécifique des certifications et habilitations (RSCH), ou la Caisse des dépôts et consignation, qui administre la plateforme « Mon compte formation », pourraient être davantage associés pour assurer la bonne reconnaissance des compétences dans le secteur privé et la fluidité entre formation initiale et continue.

La coordination nationale pourrait également s'appuyer sur les acteurs chargés de la coordination du volet recherche pour définir une méthodologie harmonisée visant à cartographier l'ensemble des formations en IA et les acteurs qui proposent ces formations, y compris du secteur privé, en complément des acteurs institutionnels d'ores et déjà identifiés et mobilisés. Cette approche permettra une cartographie complète pour la coordination nationale. Les formations en IA pourraient être mise en lumière à travers un « label » qui définirait les grands enjeux et critères des formations en IA (cf. notamment l'un des objectifs du projet EFELIA ci-avant), tout en permettant l'autonomie des établissements de l'enseignement supérieur, et une vitrine commune, au bénéfice des apprenants tout au long de la vie, en formation initiale comme en formation continue.

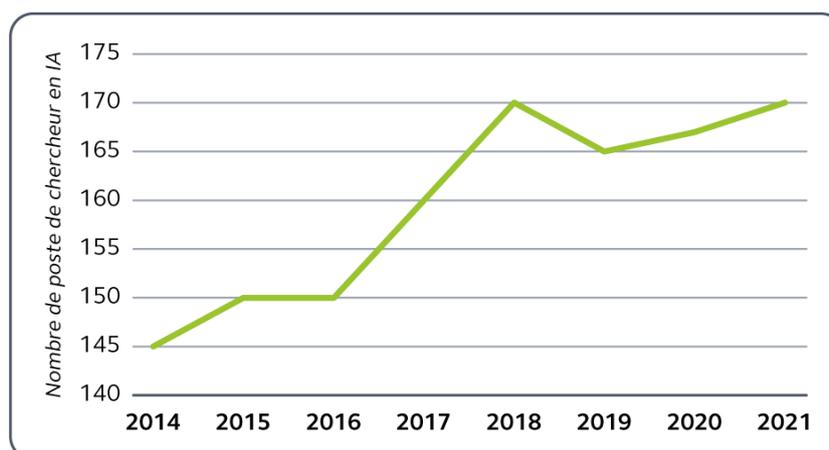
C - Une attractivité envers les chercheurs et étudiants étrangers encore fragile

La communauté scientifique en IA estime que l'attractivité « française » dispose d'atouts mais également de freins importants. Les points positifs de la recherche en France sont le cadre de vie (53 %), les compétences des étudiants (45 %) et l'accès aux infrastructures (39 %). Il existe certains freins : le faible nombre de postes permanents offerts aux concours par les organismes de recherche et l'absence de perspective de carrière, un accompagnement financier insuffisant, une allocation d'une équipe de recherche ne permettant pas de démarrer et de mener un projet¹²².

Dans la dynamique de la SNRIA, le graphique n° 14 illustre les efforts de recrutement consentis par le CNRS à l'image des organismes de recherche qui ont connu plus particulièrement en 2018 et 2019 des débauchages externes de chercheurs confirmés avant de retrouver en 2021 le niveau d'avant 2018.

¹²² C'est ainsi que deux chercheurs récemment promus ou recrutés au CNRS en 2021 ont rejoint l'EPFL en Suisse (École polytechnique fédérale de Lausanne) et que de 2013 à 2019, une trentaine d'enseignants-chercheurs ou chercheurs issus des UMR/CNRS ont quitté leur laboratoire, dont près des deux tiers pour rejoindre Google ou DeepMind, Facebook ou Amazon. Pour rappel, le laboratoire *Facebook Artificial Intelligence Research* (FAIR) a été créé au sein de Facebook en 2013 et, sur la période 2013-2019, ont été ouvertes à Paris les antennes de Huawei, Google Brain, DeepMind et du FAIR.

Graphique n° 14 : nombre de postes permanents de chercheurs en IA au CNRS



Source : Cour des comptes d'après les données du CNRS

Note de lecture : 2016 : quatre recrutements et quatre sorties ; 2017 : 11 recrutements et une sortie ; 2018 : 13 recrutements et trois sorties ; année d'implantation de laboratoires en IA de Cisco, Facebook, Deepmind, Fujitsu, Google, Hewlett Packard Enterprise, IBM, Intel, Microsoft, SAP, Naver labs ; 2019 : six recrutements et 11 sorties ; 2020 : six recrutements et quatre sorties ; 2021 : huit recrutements et cinq sorties.

L'attraction des « talents » étrangers s'inscrit dans une compétition salariale internationale entre pays « recruteurs » caractérisée par le niveau et par la qualité de l'environnement de recherche proposé au chercheur (paquet financier, moyens/équipements et équipe de recherche), la facilitation dans les démarches administratives d'installation en France pour le recruté et sa famille et dans l'obtention de visas, par exemple. Des systèmes de « guichet d'accueil » ou de « bureau d'accueil » (*Welcome desk*) sont inexistantes.

Des organismes de recherche comme le CNRS développent des mesures pour recruter spécifiquement sur l'IA en réservant systématiquement les postes (deux à cinq par an depuis 2016) au niveau du concours de recrutement chargé de recherche de classe normale (CRCN)¹²³ (cf. tableau n° 23 pour les niveaux de rémunération). Par ailleurs, a été créée en 2022 une nouvelle commission interdisciplinaire (CID) 55 « Sciences et données » destinée à recruter des candidats s'inscrivant dans une approche associant les sciences de données aux autres domaines scientifiques (physique des hautes énergies, science du climat, astronomie, science des matériaux, humanités numériques, etc.). Le premier concours de recrutement de cette CID (en cours) est doté de cinq postes de CRCN. La création de la CID a également pour objectif de faciliter la promotion

¹²³ Recruté au niveau du doctorat.

statutaire de chercheurs CRCN rattachés à l'ensemble des sections du comité national et présentant un profil interdisciplinaire associant l'IA à un autre domaine scientifique (ce concours de niveau DR2¹²⁴ est doté de cinq postes en 2022). Par ailleurs, sur les 25 postes de chaire professeur junior (CPJ) CNRS qui viennent d'être annoncés par parution au *Journal Officiel* d'avril 2022, trois postes mentionnent explicitement l'IA.

S'agissant du volet salarial, des efforts ont été entrepris de concert entre la DINSIC¹²⁵, la DGAFP¹²⁶ et la DB¹²⁷ : des grilles communes de rémunération des agents contractuels visent à limiter les concurrences entre ministères recruteurs et autoriser la dispense du visa du CBCM¹²⁸ en dessous d'un plafond de rémunération¹²⁹. Une réactualisation qui porte à 56 les métiers en tension dans le secteur numérique et informatique a été réitérée en 2021. Le personnel fonctionnaire ou permanent fait l'objet d'une revalorisation progressive mais continue depuis 2016 pour établir des niveaux de rémunération annuels qui se situent en fourchette haute des rémunérations de la fonction publique, eu égard au niveau, à l'expérience et aux responsabilités requis, sans pour autant rivaliser avec les rémunérations que le secteur privé peut proposer.

Tableau n° 23 : rémunérations brutes moyennes annuelles non chargées CNRS et Inria en 2021 du personnel fonctionnaire-permanent en IA

Rémunérations	CNRS	Inria
Directeur de recherche classe exceptionnelle	78 245,00 €	101 436,00 €
Directeur de recherche de 1 ^{ère} classe	69 264,00 €	82 985,00 €
Directeur de recherche de 2 ^e classe	63 161,00 €	67 446,00 €
Chargé de recherche hors classe	58 297,00 €	57 314,00 €
Chargé de recherche de classe normale	50 121,00 €	50 190,00 €
Ingénieur de recherche hors classe	69 363,00 €	nc
Ingénieur de recherche de 1 ^{ère} classe	55 071,00 €	nc
Ingénieur de recherche de 2 ^e classe	47 959,00 €	nc

Source : Cour des comptes d'après les données fournies par le CNRS et Inria

¹²⁴ Directeur (trice) de recherche de 2^e classe recruté(e) au niveau du doctorat et huit années de pratique à l'international.

¹²⁵ Devenue direction interministérielle du numérique (DINUM). La direction interministérielle du numérique est une direction de l'administration publique française, chargée de coordonner les actions des administrations en matière de systèmes d'information.

¹²⁶ Direction générale de l'administration et de la fonction publique-DRH de l'État.

¹²⁷ Direction du budget-Ministère de l'économie, des finances et de la souveraineté industrielle et numérique.

¹²⁸ Contrôleur budgétaire et comptable ministériel.

¹²⁹ Note DINSIC/DGAFP/DB du 19 juillet 2019 établissant le référentiel de rémunération des 15 métiers en tension dans la filière numérique et des systèmes d'information et de communication et les règles de gestion associées.

Enfin, le CNRS a perçu dans le cadre de la SNIA, à travers le programme *Choose France-CNRS Ai Rising Talents*¹³⁰ quelque 3 M€ pour attirer des chercheurs juniors prometteurs à l'international en capacité de décrocher des appels à projet de type ERC - *European Research Council-Starting grant*¹³¹. Leur recrutement est en cours sur trois postes en CDD¹³² dont la grille salariale est majorée de 50 % par rapport à la grille d'un chercheur du CNRS en CDD et bénéficie d'un environnement équivalant aux chaires individuelles financées par l'ANR. Les méthodes nouvelles et la diversification du recrutement mises en œuvre par le CNRS ne peuvent encore donner des résultats quant à leur efficacité et nécessiteront quoi qu'il advienne l'établissement d'une cartographie simplifiée et d'une communication sur les différentes voies d'accès et de recrutement au CNRS, *a fortiori* pour des candidats chercheurs non familiers du système de recherche français.

Depuis 2020, Inria a, quant à lui, multiplié par plus de deux le nombre de recrutements de permanents scientifiques (20 en moyenne par an sur la période 2010-2018, 48 par an depuis 2020), en ouvrant un nouveau statut, le statut ISFP (*Inria Starting Faculty Position*), un CDI¹³³ de droit public permettant d'avoir accès à des grilles de rémunération potentiellement plus attractives, sujet identifié comme dirimant pour l'attractivité des talents. Sur les 48 ISFP recrutés par Inria en 2020 et 2021 (24 par an), un nombre très significatif est dans le champ de l'IA (cœur de l'IA ou applicatifs). Inria met par ailleurs en œuvre des dispositifs de recrutement *Starting Research Position* (SRP) et *Advanced Research Position* (ARP). Ces dispositifs sont attractifs en termes de rémunération, particulièrement à l'international. Sur la période 2018-2021, le nombre d'étrangers employés en SRP-ARP est de 41 pour les équipes IA (58 pour toutes les équipes Inria). Par ailleurs, les équipes-projets Inria mènent une chasse aux talents à l'international pour les recrutements de chargés de recherche (CR) et de directeurs de recherche (DR). Sur la période 2018-2021, le nombre d'étrangers employés en CR et DR est de 142 pour les équipes de CR et DR en IA représentant 366 personnes au total, soit 39 %.

¹³⁰ Émergence de talents en IA.

¹³¹ Le CNRS estime que ce profil est recruté en avançant un salaire du niveau d'un directeur de recherche associé à une prime mensuelle de 4 000 € à 4 500 € et d'un paquet financier pour l'environnement de recherche de l'ordre de 500 000 € pour cinq années de recherche.

¹³² Contrat à durée déterminée.

¹³³ Contrat à durée indéterminée.

**Tableau n° 24 : évolution des ETPT entre 2016 et 2020
pour l'ensemble des équipes projets IA d'Inria**

	Type de contrat	2016	2020	Évolution
Inria	Chargé de recherche, directeur de recherche, équivalent CDI, SRP, ARP	73,8	162,3	+ 120 %
	Doctorants	77,6	162,3	+ 109 %
	Post-doctorants	21,8	44,3	+ 103 %
	Ingénieurs	59,2	118,9	+ 101 %
	dont nationalité hors française	20,4	45,5	+ 123 %
Hors Inria	Enseignants-chercheurs : professeur des univ., maître de conf.	152,6	266,3	+ 75 %
	Membres d'équipes non payés par Inria hors personnel académique (<i>faculty members</i>)	303,5	809,9	+ 167 %

Source : Cour des comptes d'après les données d'Inria

Note de lecture : CR : chargés de recherche, DR : directeurs de recherche, CDI : contrats à durée indéterminée, SRP : started research position, ARP : advanced research position.

L'attractivité des talents à l'étranger

La comparaison sur les stratégies nationales en IA menées par la Cour met en évidence des mesures adoptées très volontaristes. Dans le cadre de son programme de chaires, le plan Canadien prévoit que la moitié doit revenir à des chercheurs étrangers. La Finlande a développé une série d'initiatives spécifiques dont trois peuvent être notées : une procédure accélérée pour l'obtention des permis de travail et de résidence pour les spécialistes ; des mesures prises pour encourager les étudiants internationaux à rester en Finlande (extension de la validité du permis de séjour) ; le projet HEI LIFE (*Higher Education Institutions' Support for International Academics and Staff - Living, Integration, Family and Employment in Finland*) vise à créer des modèles nationaux pour les établissements d'enseignement supérieur finlandais afin de soutenir l'intégration des universitaires et du personnel internationaux et de leurs familles.

Le rapport produit par le *think-tank* Macro Polo, qui s'intéresse au *top tier AI researchers* (les 20 % les meilleurs selon certains critères quantitatifs), montre cependant que la grande majorité (60 %) de ceux-ci travaillent actuellement aux États-Unis, dans des universités ou des entreprises.

Dans le cadre de la SNRIA, Inria est également mandaté par le MESR pour mener l'expérimentation d'un programme « *Choose France* » destiné à permettre le recrutement de trois scientifiques, au plus haut niveau international, en accompagnant ces recrutements d'une chaire sur une durée de cinq ans. Les recrutements effectués dans ce cadre se feront dans le contexte des sites universitaires sur lesquels Inria est présent, à travers les « Centres Inria de l'Université ». L'objectif de cette expérimentation est de mettre en place un dispositif d'attractivité permettant de recruter des scientifiques, pas nécessairement issus du milieu académique, actuellement en activité à l'étranger, sans condition de nationalité. Le retour durable et effectif en France de scientifiques français de rang mondial est également un objectif explicite et une priorité du programme. La possibilité de proposer une rémunération attractive associée à un tel CDI est une condition indispensable pour la réussite de ce programme. La grille de rémunération des agents contractuels ne permet pas de répondre aux attendus initiaux de ce programme qui vise à accueillir des scientifiques au plus haut niveau international, dans le cadre d'une compétition pour les talents du numérique qui implique les grands acteurs académiques mondiaux et les géants de la Tech américains et chinois.

Inria envisage donc de proposer aux candidats une rémunération plus attractive que celle issue du cadre d'emploi classique. Bien que très en deçà des salaires proposés par les concurrents étrangers pour le type de profils souhaités, cela devrait néanmoins permettre à Inria d'accueillir des candidatures pertinentes. Ces propositions de rémunération seront discutées avec chaque candidat retenu puis soumises pour avis préalable du CBCM, qu'Inria a déjà fortement sensibilisé à l'importance du dispositif « *Choose France* ».

II - Une projection internationale et européenne de la stratégie à poursuivre

A - Des coopérations internationales à amplifier

Le renforcement des collaborations internationales était un objectif explicite de la SNRIA, avec une attention particulière portée sur la coopération bilatérale entre la France l'Allemagne (cf. partie II.IV.B pour les coopérations entre centres d'excellence). Dans le cadre du Traité d'Aix-la-Chapelle, les deux pays ont ainsi signé en octobre 2019 un partenariat commun visant à resserrer les liens des écosystèmes de recherche et d'innovation en IA. Ce dernier a notamment donné lieu en 2020 à un appel à projets en matière de recherche fondamentale¹³⁴ lancé de manière

¹³⁴ Un autre appel à projets visant à soutenir des projets innovants permettant de prévenir et de gérer les crises sanitaires, économiques et environnementales a été lancé en 2021 par les ministères français et allemand en charge de l'économie.

conjointe par le ministère de l'enseignement supérieur français et le ministère fédéral de l'éducation et de la recherche allemand. L'approfondissement des liens est également notable avec d'autres pays, à l'instar du Japon¹³⁵, de Singapour¹³⁶, du Canada¹³⁷ et de la Finlande¹³⁸.

En dehors du cadre de la SNIA bien qu'en cohérence avec ses objectifs, l'IA reste une composante importante des partenariats des organismes de recherche. Inria est ainsi impliqué dans le programme du *Canadian Institute for Advanced Research* (CIFAR¹³⁹) « Apprentissage automatique, apprentissage biologique », renouvelé en 2019 et codirigé par Yann Le Cun (Université de New York, Facebook) et Yoshua Bengio (Université de Montréal) et a signé en janvier 2020 un accord de partenariat stratégique avec le centre de recherche allemand pour l'intelligence artificielle (DFKI). Le CNRS a développé plusieurs laboratoires de recherche internationaux (IRL) impliquant de l'IA, comme en Australie¹⁴⁰, à Singapour, au Japon, au Royaume-Uni et au Canada¹⁴¹. Le programme de recherche DESCARTES, portant sur l'IA hybride et co-financé par la *National Research Foundation* de Singapour, a été lancé en 2022.

La seconde phase de la stratégie doit être l'occasion de poursuivre le développement des partenariats bilatéraux et multilatéraux en IA, tout en veillant à la bonne articulation des actions menées au niveau des organismes de recherche, entre elles, d'une part, et par rapport aux priorités définies par la stratégie, d'autre part. Le renforcement de la présence des laboratoires de recherche français dans les réseaux internationaux pourrait également être une priorité, au sein notamment des réseaux européens ELLIS (*European laboratory for learning and intelligent systems*) et CLAIRE (*Confederation of Artificial Intelligence Research Laboratories in Europe*) qui ont émergé à partir de 2018 à l'initiative des communautés scientifiques¹⁴². À ce jour, le réseau ELLIS compte 30 unités dans 15 pays

¹³⁵ Séminaires bilatéraux ou trilatéraux avec l'Allemagne, appels à projets bilatéraux de l'ANR avec l'agence japonaise pour la science et la technologie.

¹³⁶ Signature d'un *Memorandum of Understanding* le 13 juillet 2018 par l'Université nationale de Singapour, le CNRS, Inria et l'Inserm.

¹³⁷ Afin de notamment promouvoir le développement responsable et éthique de l'IA (section III.II).

¹³⁸ Une déclaration conjointe de coopération en IA du 30 août 2018 et des échanges sont en cours pour rapprocher les écosystèmes de la recherche et de l'innovation des deux pays. Un accord entre le *Health Data Hub* et *FinData* a été par ailleurs conclu en 2021.

¹³⁹ Le CIFAR est une organisation canadienne à but non lucratif, dont le siège est basé à Toronto qui soutient des programmes de recherches multidisciplinaires réunissant des spécialistes et des scientifiques internationalement reconnus afin de répondre à des questions revêtant une importance mondiale.

¹⁴⁰ Création en 2020 de l'IRL CROSSING autour de l'IA et de l'interaction humain machine, en partenariat avec Naval group.

¹⁴¹ Création en 2022 de l'IRL ILLS centré sur l'IA.

¹⁴² Les deux réseaux ont pour objectif de renforcer les interactions des chercheurs en IA. Le réseau CLAIRE vise à établir un réseau de centres d'excellence en IA dans toute l'Europe et un centre européen d'IA. ELLIS a pour but d'établir des instituts de recherche

dont sept en Allemagne et une en France, à Paris. Le réseau CLAIRE, présent dans 37 pays regroupe 445 laboratoires et institutions dont seulement 28 français (62 allemands). Les résultats de la consultation menée par la Cour corroborent également la faible présence française dans les réseaux professionnels en IA : seuls 5,6 % et 6 % d'entre eux sont impliqués respectivement dans ELLIS et CLAIRE.

B - Une nécessité de capitaliser davantage sur les financements européens

Dès sa première phase la stratégie nationale a été définie dans la perspective de s'inscrire dans une approche plus globale portée à l'échelle européenne. Les priorités européennes ont depuis été davantage formalisées (cf. chapitres 1 et 2), et associées à des financements spécifiques.

Dans le cadre du plan européen pour l'IA, divers programmes budgétaires européens de soutien à la recherche participent à favoriser le développement de l'IA – les deux principaux étant actuellement « Horizon Europe »¹⁴³ et « pour une Europe numérique »¹⁴⁴. Succédant à « Horizon 2020 », « Horizon Europe » est doté d'un budget total de près de 100 Md€ sur la période 2021-2027, dont une partie significative sera consacrée à des projets concernant l'IA. Pour le budget 2021-2022, la Commission avait ainsi estimé que 29 % des 4,3 Md€ prévus pour le cluster « Numérique, Industrie et Espace » de ce programme, soit 1,2 Md€, seraient destinés à la recherche en IA. S'agissant du programme « pour une Europe numérique », 7,5 Md€ sont prévus pour l'ensemble des actions sur la période 2021-2027. En 2021, la Commission a adopté trois programmes de travail dans ce cadre, pour un financement total de 1,98 Md€. Le principal, doté de 1,38 Md€, devait porter notamment sur des investissements dans le domaine de l'IA jusqu'à la fin de 2022.

L'élaboration de la stratégie nationale d'accélération a ainsi pris en compte les principaux objectifs de l'actuel plan coordonné européen et les moyens d'intervention associés. Des stratégies *ad hoc* ont par ailleurs été mises en place par certains opérateurs de recherche pour renforcer leur participation aux programmes européens. Un dispositif d'accompagnement des projets européens a été proposé d'après le SGPI dans le cadre du volet *Edge IA* de l'appel *Testing and Experimental Facility* du programme « Pour une Europe numérique ».

de pointe en IA, de renforcer la recherche fondamentale et de créer un programme européen de doctorat en IA pour favoriser l'excellence européenne dans le domaine.

¹⁴³ Programme-cadre de l'Union européenne pour la recherche et l'innovation pour la période 2021-2027.

¹⁴⁴ Programme de financement axé sur l'intégration de la technologie numérique aux entreprises, aux citoyens et aux administrations publiques.

Pour autant, il est nécessaire de capitaliser davantage sur les efforts menés à l'échelle européenne¹⁴⁵. Par rapport à la première phase de la stratégie, l'effet de levier des financements étatiques sur les financements européens pourrait être amélioré¹⁴⁶. À l'occasion de la mise en œuvre de la stratégie d'accélération, cet indicateur pourrait faire l'objet d'un suivi spécifique de la part la coordination, tout comme l'évolution du taux de succès des opérateurs français aux réponses des AAP européens.

III - Un positionnement stratégique autour des enjeux sociétaux de confiance et de frugalité

À mesure que les applications de l'IA progressent et irriguent de nombreux domaines sociaux-économiques, la nécessité de garantir une IA fiable, non biaisée, conforme aux normes éthiques, qui inspire la confiance des utilisateurs, est plus prégnante. En France, les questions nouvelles posées par l'IA ont été prises en compte par les pouvoirs publics en amont du lancement de la stratégie, sous l'impulsion notamment des travaux de la Commission nationale de l'informatique et des libertés (CNIL)¹⁴⁷. Ces priorités sont également identifiées au niveau européen¹⁴⁸, où la définition de l'IA de confiance retenue est celle d'une IA licite (i), « *assurant le respect des législations et réglementations applicables* », éthique (ii) « *assurant l'adhésion à des principes et valeurs éthiques* » et robuste sur le plan tant technique que social pour prévenir les préjudices involontaires (iii)¹⁴⁹.

En parallèle, l'extension d'infrastructures de calcul ou de stockage intensives en ressources énergétiques soulève la question de l'impact environnemental de l'IA, dans un contexte où le numérique émet 2,5 % des émissions de gaz à effet de serre au niveau national¹⁵⁰. Ces émissions se situent sur une trajectoire de croissance importante¹⁵¹, en dépit des objectifs

¹⁴⁵ Le CNRS a, par exemple, lancé en 2021 une stratégie européenne visant à augmenter sa participation au programme « Horizon Europe ».

¹⁴⁶ À titre d'exemple, sur la période 2014-2020, la France a bénéficié chaque année entre 7,4 % et 13,2 % des crédits consacrés à l'IA dans le programme « Horizon 2020 » (en contribution nette).

¹⁴⁷ Un débat public sur les algorithmes et l'IA, réunissant près de 3 000 participants à 45 manifestations, a notamment été mené par la CNIL en 2017, en lien avec 60 partenaires. Cette consultation a donné lieu au rapport de synthèse et de propositions *Comment permettre à l'homme de garder la main*, publié la même année.

¹⁴⁸ Cf. notamment le livre blanc de la Commission européenne COM (2020) 65 final, publié le 19 février 2020 accompagné d'un rapport sur les conséquences de l'IA, de l'internet des objets et de la robotique sur la sécurité et la responsabilité.

¹⁴⁹ Définition issue des lignes directrices en matière d'éthique pour une IA digne de confiance publiées en 2019 par le Groupe d'experts indépendants de haut niveau sur l'IA constitué par la Commission européenne.

¹⁵⁰ Source : Arcep.

¹⁵¹ Mission d'information sur l'empreinte environnementale du numérique - Sénat (senat.fr).

français et européens affichés de neutralité carbone en 2050¹⁵². Dès lors, si l'IA peut constituer une réponse aux défis de la transition écologique, en améliorant par exemple la gestion du système énergétique, des risques naturels, ou encore la productivité du système agricole, elle se doit d'intégrer les enjeux environnementaux plus largement dans sa conception, les infrastructures qu'elle nécessite et ses méthodes d'apprentissage ou d'usage.

Les notions d'IA de confiance et d'IA frugale (qui intègre pour cette dernière les problématiques environnementales) ne sont pas exclusives. Elles s'inscrivent dans le concept plus vaste d'IA « responsable » développée notamment par l'OCDE dans ses principes sur l'IA¹⁵³. Ces thématiques avaient déjà été identifiées comme prioritaires dans le cadre de la première phase de la stratégie, donnant lieu notamment au Partenariat mondial pour l'intelligence artificielle (PMIA) lancé en juin 2020 sous l'impulsion de la France et du Canada. Il s'agit en effet d'enjeux portés par la France sur les scènes européennes et internationales¹⁵⁴, pour lesquels elle ambitionne d'être à la fois pionnière et prescriptrice de référentiels. À cet égard, la Cour constate que le partage des rôles entre le coordonnateur national et le ministère des affaires étrangères gagnerait à être clarifié. Dans une étude réalisée à la demande du Premier ministre et publiée le 30 août 2022, le Conseil d'État préconise de confier de manière explicite au coordonnateur national en lien avec l'Ambassadeur pour le numérique le déploiement d'une stratégie d'influence à l'échelle européenne et internationale.

Le partenariat mondial pour l'IA (PMIA)

Le PMIA est une initiative multipartite lancée en juin 2020, à la suite de discussions menées dans le cadre du G7 sous l'impulsion notamment de la France et du Canada. Pour la France, la promotion d'un dialogue international portant sur les enjeux éthiques de l'IA était l'un des objectifs affichés de la première phase de la SNRIA et se réalise au travers notamment de la présidence française du PMIA à partir de 2021.

¹⁵² Stratégie nationale bas carbone.

¹⁵³ Adoptés en mai 2019, ils encouragent une utilisation de l'IA innovante qui soit fondée sur les valeurs de croissance inclusive, de développement durable et de bien-être (i), des valeurs centrées sur l'humain et l'équité (ii), la transparence et l'explicabilité (iii), la robustesse, la sûreté et la sécurité (iv), la responsabilité (v).

¹⁵⁴ Outre le Partenariat mondial pour l'IA, la France est active au sein de plusieurs instances internationales, comme le groupe de travail sur l'IA du Groupe de haut niveau des Nations unies sur la coopération numérique (HLP), qui cherche à favoriser l'« explicabilité » et la transparence des résultats des systèmes intelligents autonomes, ou encore le Comité du Conseil de l'Europe *ad hoc* sur l'intelligence artificielle (CAHAI), qui évalue notamment l'impact du numérique sur les droits de l'Homme. Elle participe également aux discussions portant sur le cadre juridique de l'IA au sein des organes spécialisés de l'ONU, du G7 et du G20, et de l'OCDE.

Comportant à l'origine 15 membres fondateurs, le PMIA regroupe aujourd'hui 25 membres¹⁵⁵ dont l'Union européenne et les États-Unis, engagés à mettre en œuvre les recommandations de l'OCDE pour permettre une IA responsable et durable. Deux sommets, réunissant des experts internationaux de l'IA issus de la société civile, du monde universitaire, de l'industrie et des gouvernements, ont eu lieu à Montréal et à Paris en 2020 et 2021. Ces deux villes hébergent également les deux centres d'expertise du PMIA, qui interviennent en appui scientifique technique et logistique à l'activité du partenariat. Le centre d'expertise de Paris est hébergé par Inria.

Conformément aux priorités du plan européen en intelligence artificielle défini en 2018¹⁵⁶ et mis à jour en 2021¹⁵⁷, l'IA de confiance et l'IA frugale sont deux axes majeurs transversaux de la seconde phase de la stratégie, présentés comme facteurs d'attractivité pour la France et l'Union européenne sur la scène internationale et comme leviers pour renforcer la compétitivité des entreprises. La confiance et la frugalité sont ainsi deux axes thématiques de recherche prioritaires du PEPR consacré à l'IA qui transcrite dans le domaine de la recherche ces enjeux.

La stratégie européenne en matière d'IA

À partir de 2018, une politique européenne en matière d'IA a été mise en place par la Commission européenne dont les principales étapes sont les suivantes :

- élaboration d'une stratégie européenne par la Commission, avec l'ambition d'atteindre 20 Md€ d'investissements publics et privés par an dans l'écosystème (2018) ;
- élaboration de lignes directrices pour une IA digne de confiance par un groupe d'experts indépendants de haut niveau nommés par la Commission (2019) ;
- publication d'un livre blanc accompagné d'un rapport sur les conséquences de l'IA, de l'internet des objets et de la robotique sur la sécurité et la responsabilité (2020).

¹⁵⁵ L'Allemagne, l'Australie, la Belgique, le Brésil, le Canada, la République de Corée, le Danemark, l'Espagne, les États-Unis, la France, l'Inde, l'Irlande, Israël, l'Italie, le Japon, le Mexique, la Nouvelle-Zélande, les Pays-Bas, la Pologne, la République tchèque, le Royaume-Uni, Singapour, la Slovaquie, la Suède et l'Union européenne.

¹⁵⁶ Commission européenne, Communication de la Commission au Parlement européen, au conseil européen, au conseil, au comité économique et social européen et au comité des régions : « Un plan coordonné dans le domaine de l'intelligence artificielle », COM/2018/795 final, 7 décembre 2018, Bruxelles.

¹⁵⁷ Commission européenne, Communication de la Commission au Parlement européen, au conseil, au comité économique et social européen et au comité des régions, « Favoriser une approche européenne de l'intelligence artificielle », COM/2021/205 final, 21 avril 2021, Bruxelles.

En 2021, une nouvelle phase a été lancée par la Commission avec une proposition de règlement visant à mettre en place un cadre juridique harmonisé sur l'IA, et la publication d'un nouveau plan coordonné. Ce dernier a pour but d'accélérer les investissements dans l'IA et de mettre en cohérence les stratégies et programmes nationaux afin de prévenir toute fragmentation des approches. Il se décline autour de quatre thèmes principaux : établir des conditions propices au développement et à l'adoption de l'IA dans l'UE (i) ; faire de l'UE le lieu de croissance de l'excellence, du laboratoire au marché (ii) ; veiller à ce que l'IA fonctionne pour les personnes et soit une force pour le bien dans la société (iii) ; renforcer le *leadership* stratégique dans les secteurs à fort impact (iv).

La stratégie européenne a ainsi pour but de renforcer les capacités de recherche et d'industrie sur l'IA, tout en garantissant les droits fondamentaux. La défense des principes éthiques est perçue comme un levier de compétitivité et d'attractivité par rapport notamment aux États-Unis et à la Chine qui n'occupent pas de positions dominantes sur ces aspects. Cette approche s'inscrit dans la volonté de l'UE d'adopter un cadre législatif protecteur des consommateurs et des utilisateurs de l'économie numérique plus largement, comme en témoigne l'adoption en 2022 sous présidence française de l'UE des règlements *Digital Market Act* (DMA) et *Digital Service Act* (DSA) qui entreront en vigueur en 2023.

La consultation auprès de la communauté scientifique en IA montre que ces thématiques sont encore peu prises en compte dans les recherches en IA. L'amélioration du niveau de maturité de la communauté scientifique sur ces sujets est dès lors une condition *sine qua non* de réussite dans la mise en œuvre de la seconde phase de la stratégie.

Elle se pose avec une acuité particulière s'agissant de l'IA frugale, puisqu'il peut exister une tension entre l'économie des ressources et la performance de l'IA, souvent identifiée comme axe prioritaire de son développement.

A - Une déclinaison opérationnelle de l'IA de confiance par plusieurs mesures dans la seconde phase de la stratégie

Outre les financements prévus dans le cadre du PEPR, trois mesures de la phase d'accélération sont orientées vers le renforcement de la confiance dans l'usage de l'IA. Elles témoignent d'une réorientation du volet recherche de la stratégie nationale en IA, davantage orienté vers la recherche appliquée, pour tenir compte de l'essor de l'IA industrielle¹⁵⁸.

¹⁵⁸ En parallèle de la montée en charge du volet « économie » de la stratégie : le ministère de l'économie et des finances et huit industriels français ont signé, en 2019, un manifeste pour développer une intelligence artificielle (IA) au service de l'industrie.

Des échanges réguliers ont lieu entre le coordonnateur de la stratégie et la CNIL, sans que cette dernière ait de rôle spécifiquement assigné concernant le volet éthique de la stratégie. Compte tenu de l'importance de ce volet dans la seconde phase de la stratégie, ces relations pourraient être davantage institutionnalisées et le rôle de la CNIL renforcé.

Tableau n° 25 : l'IA de confiance dans la stratégie d'accélération (hors PEPR)

Mesures	Description	Date	Montants alloués	Cheffe(s) de file
<i>Programme de recherche « Confiance.ai » dans le cadre du Grand défi « Sécuriser, certifier et fiabiliser les systèmes fondés sur l'IA ».</i>	Le programme rassemble un collectif d'acteurs académiques et industriels pour développer une plateforme d'outils logiciels permettant l'intégration d'une IA fiable et sécurisée dans les produits et services industriels innovants.	En cours	10 M€	DGE, MESR, Directeur Grand défi IA
<i>Démonstrateurs de l'IA de confiance</i>	Réalisation de quatre démonstrateurs applicatifs ambitieux dans quatre filières stratégiques : automobile, aéronautique, énergie, chaînes d'assemblages.	Premier semestre 2023	30 M€	DGE, MESR, DGEC, DGITM, Directeur Grand Défi
<i>Stratégie de normalisation de l'IA de confiance</i>		En cours	1,6 M€	DGE, MESR, DGEC, DGITM, Directeur Grand Défi

Source : Cour des comptes

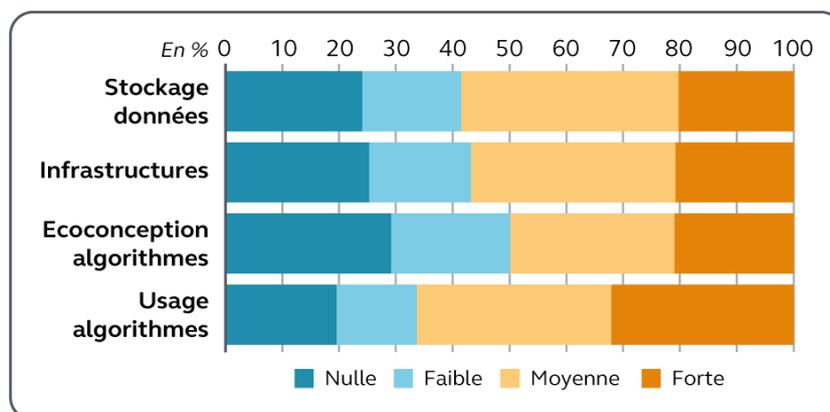
B - L'IA frugale, principe fondateur qui doit irriguer de manière concrète l'ensemble des recherches financées par la seconde phase

Les enjeux de frugalité en énergie et en données peuvent s'analyser selon deux orientations : un axe stratégique de positionnement scientifique notamment autour de l'IA hybride, c'est-à-dire une IA qui ne repose pas exclusivement sur la performance des algorithmes et sur la disponibilité de grande masse de données, et un axe de fonctionnement visant à réduire les impacts environnementaux de la recherche en IA à différentes échelles : les centres de stockage de données et ceux de calcul, l'entraînement et ensuite l'usage.

Il existe néanmoins une potentielle tension entre la performance de l'IA (mesurée au travers de la capacité de traitement de données à grande échelle, ou de la rapidité d'apprentissage et de prédiction) et la frugalité. Cela met en évidence à la fois des verrous scientifiques à lever et un enjeu d'acceptabilité sociale du développement de l'IA fondée sur une analyse du coût environnemental par rapport aux bénéfices attendus.

La prise de conscience du domaine de la recherche reste à renforcer et à systématiser au-delà de quelques initiatives prometteuses. Certains centres d'excellence développent des approches d'IA hybride et frugale. Le CNRS dispose également d'un groupe de réflexion du groupement de service Ecoinfo du CNRS qui travaille sur la problématique spécifique d'impact de l'IA sur le climat et vise en particulier à établir une première méthode simple d'estimation de la consommation d'un système d'IA, à compléter par la suite par le cycle de vie des matériels puis, à plus long terme, par les impacts indirects. La PMIA en partenariat avec le *Centre for AI and Climate* a ainsi pour objectifs de définir une stratégie mondiale de l'IA responsable destinée à guider les décideurs chargés de développer les stratégies de lutte contre le changement climatique. Dans le cas de « Jean Zay », le CNRS et GENCI ont choisi une solution de refroidissement à l'eau chaude, qui limite l'énergie nécessaire à l'alimentation électrique et au refroidissement de la machine. Les calories émises en sortie sont réutilisées pour le chauffage de plusieurs bâtiments du campus Paris-Saclay (plus de 1 000 logements sur le plateau), une première en Europe à cette échelle.

Graphique n° 15 : perception de la prise en compte de l'impact environnemental dans la recherche



Source : Cour des comptes - Consultation de la communauté scientifique en IA

Néanmoins, la consultation auprès de la communauté scientifique indique que, globalement, la prise en compte de l'impact environnemental de la recherche en IA est forte pour seulement 23,53 % de répondants et nulle pour 24,50 % révélant une maturité faible de cet écosystème sur ces enjeux. L'usage des algorithmes semble être la dimension la plus intégrée dans les travaux de recherche (cf. graphique n° 15). Cela confirme également les éléments issus des visites des centres d'excellence dans lesquelles la prise en compte et la priorisation des enjeux de frugalité et d'impact environnementaux sont hétérogènes.

Cet axe prioritaire est d'autant plus important qu'il permet de répondre au cadre de financement de la stratégie par le quatrième programme d'investissement d'avenir et France 2030. Ces crédits doivent être employés selon une doctrine inscrite dans la loi¹⁵⁹. Cette doctrine exige de contribuer à la transition écologique et d'assurer un retour financier et extra-financier des investissements, c'est-à-dire d'avoir des impacts positifs sur les enjeux sociaux et environnementaux. L'IA frugale est l'incarnation de cette problématique. Le ministère de la transition écologique et de la cohésion des territoires a certes impulsé une feuille de route pour l'IA au service de la transition écologique le 15 septembre 2021 et met l'accent sur les conditions de succès d'un impact de l'IA sur la transition écologique : qualité des données avec une logique de frugalité (comment utiliser de manière qualitative moins de données), partage et mutualisation des données pertinentes à la transition écologique, etc. Néanmoins, elle ne dresse pas un programme de travail pour le domaine de la recherche.

Afin de s'assurer la bonne atteinte des objectifs de frugalité dans le développement de l'IA lors de la deuxième phase quel que soit le secteur, il serait souhaitable d'améliorer la mesure et le suivi de l'empreinte écologique de l'IA et de construire un cadre d'actions compatible avec la maîtrise de l'empreinte environnementale. Un travail autour d'une charte et un guide de bonnes pratiques pourrait être utile à intégrer à l'ensemble des AAP de la stratégie.

¹⁵⁹ Art. 233 de la loi n° 2020-1721 du 29 décembre 2020 de finances pour 2021, modifiant le B du I de l'article 8 de la loi n° 2010-237 du 9 mars 2010 de finances rectificative pour 2010.

CONCLUSION ET RECOMMANDATIONS

La SNRIA avait comme objectif d'engager un dialogue entre « performance de l'IA et humanité ». De fait, les enjeux éthiques ont été centraux dans le déploiement de la première phase de la stratégie, et portés par la France sur la scène internationale. De manière emblématique, le PMIA, qui vise à garantir une IA responsable et durable, a été lancé en juin 2020 sous son impulsion et celle du Canada. Deux sommets mondiaux portant sur ces enjeux se sont tenus en 2020 et en 2021.

À la dernière question évaluative, la Cour est amenée à répondre positivement, tout en soulignant des points de vigilance : la SNIA a permis une plus grande prise en compte des enjeux éthiques, et notamment de frugalité et de confiance, dans le développement de l'IA. Pour autant, alors que ces thématiques sont deux axes structurants de la seconde phase, la maturité de la communauté scientifique sur ces questions reste faible. L'amélioration de la prise en compte de ces sujets par la communauté scientifique est dès lors une condition indispensable pour répondre aux ambitions que se fixe la stratégie. Cet enjeu est particulièrement prégnant s'agissant du développement de l'IA frugale, puisqu'il peut exister une tension entre la performance de l'IA et l'économie des ressources.

Les priorités de la phase d'accélération, notamment sur les aspects tenant à l'IA responsable, ont été fixées en conformité avec celles portées par le plan coordonné européen amendé en 2021. À partir de cette approche, cohérente, il serait utile de capitaliser encore davantage sur les efforts menés au niveau européen.

Bien que les formations en IA, tant initiales que continues, soient en croissance, cette dynamique n'est pas à la hauteur des ambitions de la stratégie.

De manière plus prospective, la France devrait amplifier ses efforts en matière de formation tant dans l'affirmation de l'excellence de son école de mathématiques, du positionnement de ses organismes de recherche et de ses universitaires, que dans la diffusion massive d'une offre de formation initiale et de continue de qualité à tous les niveaux. Cette double exigence répond au risque avéré de tarissement de la filière « mathématiques » qui constitue l'une des filières naturelles de formation à l'IA et, en particulier, le tarissement de vocation déjà constaté chez les femmes.

La seconde phase de la stratégie devrait permettre d'articuler efficacement formation de haut niveau et recherche d'excellence.

La Cour formule les recommandations suivantes :

5. *réaliser une cartographie harmonisée et actualisée des formations en IA à valoriser au travers d'un label commun pour les rendre visible et accompagner la massification (MESR) ;*
 6. *prévoir les besoins en enseignants du secondaire, en enseignants-chercheurs et en chercheurs formés à l'usage de l'IA et établir des plans de formation en adéquation (MESR) ;*
 7. *élaborer une charte et un catalogue de bonnes pratiques visant à définir et suivre l'impact environnemental de la recherche en IA, à favoriser le développement d'une IA responsable (SGPI, MESR).*
-

Conclusion générale

La décision au plus haut niveau de l'État et la formalisation consécutive d'une stratégie en IA ont permis d'identifier cette thématique comme une priorité stratégique de niveau national.

Le choix fondateur porté dans un premier temps sur le soutien à la recherche a permis à la France de conserver son rang dans le domaine à l'échelle mondiale (10^e) et d'éviter un décrochage.

Cette évaluation permet d'aboutir à la constatation de ce *statu quo* sur le positionnement de la France, tout en relevant la réalité d'une structuration nécessaire de l'écosystème et l'émergence à consolider de pôles d'excellence. Elle met en évidence l'enjeu de concilier massification de l'offre de formation et qualité de la recherche, ce qui incite à dépasser l'écosystème de l'IA dite « de cœur » pour embarquer, dans une dynamique plus large, les applications de l'IA dans l'ensemble des disciplines et des usages quotidiens.

La stratégie a bel et bien favorisé une première structuration d'un écosystème d'enseignement et de recherche sur cette thématique qui n'est pas une discipline en tant que telle et ne bénéficie donc pas des organes de gouvernance et de pilotage habituels.

Cependant, au regard de l'évaluation après quatre années de mise en œuvre et au moment-charnière de la seconde phase, cette gouvernance et ce pilotage ne doivent pas conduire à ralentir l'identification, la cohérence, la pertinence et la priorisation des actions à mener, non seulement pour ne pas décrocher mais également pour gagner en efficacité et atteindre le rang des pays qui comptent en IA.

L'essoufflement constaté de la gouvernance de cette politique nationale ne doit pas non plus conduire à affaiblir les efforts financiers encore à consentir et à décourager les acteurs issus de cette stratégie nationale, en premier lieu, les universités, les instituts 3IA et les organismes nationaux de recherche mais également les partenaires économiques et les collectivités locales.

Les freins identifiés résident dans l'absence de vision à moyen terme de la pérennité des centres d'excellence et de leurs modalités de financement, maillons importants de cet écosystème mais également dans l'absence de suivi budgétaire global satisfaisant.

S'agissant de la gouvernance, elle devra *a minima* être plus cohérente pour articuler la recherche et la formation et permettre un suivi plus systématique et un pilotage général plus efficace.

L'évaluation relève également que de nombreux efforts restent à mener en matière d'IA frugale et d'IA de confiance, à la fois sur le plan de la mobilisation des chercheurs, de l'affectation des moyens et de la construction indispensable d'un cadre normatif européen.

L'ancrage et l'appui de la stratégie nationale dans et sur une stratégie européenne sont à renforcer et à rechercher davantage. Cette politique publique doit trouver une place pérenne dans les politiques nationales entre un besoin d'agilité lié à l'évolution très rapide du domaine de l'IA et la nécessité d'avoir un engagement de long terme pour préparer l'avenir et utiliser les atouts français.

Plus largement, le périmètre d'évaluation de cette politique publique, notamment en termes de pertinence, de cohérence et d'efficacité, ne peut se limiter à celui de la stricte stratégie d'accélération en IA. Il est préférable d'évaluer de manière coordonnée et intégrée l'ensemble des actions prises et politiques adoptées dans tous les domaines et sur toute la chaîne de valeur pour permettre d'assurer une large diffusion de l'IA dans la société de manière éthique et frugale.

Liste des abréviations

3IA	Instituts Interdisciplinaires en Intelligence (3I) Artificielle
AAP.....	Appel à projets
ANR	Agence nationale de recherche
AI	<i>Artificial intelligence</i> en anglais
AID	Agence innovation de la défense
AISSAI.....	centre IA pour la science et sciences pour l'IA
Allistène	Alliance des sciences et technologies du numérique
AMI.....	Appel à manifestation d'intérêt
AMI-CMA	Appel à manifestation d'intérêt -compétences et métiers d'avenir
ANITI.....	<i>Artificial and natural intelligence Toulouse Institute</i>
ANR	Agence nationale de la recherche
ANRT.....	Association nationale de la recherche et de la technologie
ANSSI.....	Agence nationale de la sécurité des systèmes d'information
CAM	Centre appui métier (Cour des comptes)
CDC	Caisse des dépôts et consignations
CEA.....	Commissariat à l'énergie atomique et aux énergies alternatives
CDD	Contrat à durée déterminée
CID.....	Commission interdisciplinaire - CNRS
CNIL	Commission nationale de l'informatique et des libertés
CNRS	Centre nationale de la recherche scientifique
COP.....	Contrat d'objectifs et de performance
CPJ	Chaire professeur junior
CPU.....	Unité centrale de calcul
CR	Chargé de recherche
CRCN.....	Chargé(e) de recherche de classe normale
DATAIA	Institut convergence dédié aux sciences des données, l'intelligence artificielle et la société
DB	Direction du budget
DFKI	Centre de recherche allemande pour l'intelligence artificielle
DGAFP	Direction générale de l'administration et la fonction publique
DGE	Direction générale des entreprises

DGESCO.....	Direction générale de l'enseignement scolaire
DGRI.....	Direction générale de la recherche et de l'innovation
DUT	Diplôme universitaire technologique
DIM.....	Domaine d'intérêt majeur
DINSIC	Direction interministérielle du numérique et des systèmes d'information et de communication
DINUM	Direction interministérielle du numérique anciennement DINSIC
DMD	Direction des méthodes et des données (Cour des comptes)
DR (CE, 1 ou 2)....	Directeur de recherche (de classe exceptionnelle, de 1 ^{ère} classe ou de 2 ^e classe)
EFELIA	École Française pour l'intelligence artificielle (projet des 3 IA)
EPFL	École polytechnique fédérale de Lausanne
ERC.....	<i>European Research Council</i>
ERC <i>Starting grant</i> .	Bourse de démarrage de travaux de recherche allouée par l'ERC aux jeunes chercheurs
ETHZ	École polytechnique fédérale de Zurich
ETI	Entreprises de taille intermédiaire
FAIR	<i>Facebook Artificial Intelligence Research</i>
GAFAMI.....	Google, Amazon, Facebook, Microsoft et IBM (entreprises américaines)
GDR	Groupement de recherche
GENCI	Grand équipement national de calcul intensif
GES	Gaz à effet de serre
GPI.....	Grand plan d'investissement
GPU.....	Unité de traitement graphique
HAL	Hyper articles en ligne (archive ouverte pluridisciplinaire)
HCERES	Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur
HPC.....	<i>High performance computing</i> (calcul haute performance)
IA	Intelligence artificielle
IDRIS	Institut du développement des ressources en informatique scientifique
INPI.....	Institut national de la propriété intellectuelle
Inria	Institut nationale de recherche en sciences et technologies du numérique
Inserm	Institut national de la santé et de la recherche médicale
Insmi	Institut national des sciences mathématiques et informatiques
INS2I.....	Institut des sciences de l'information et de leurs interactions

IRT	Institut de recherche technologique
Labcom.....	Laboratoires communs
MCU	Maitre de conférences des universités
MENJ	Ministère de l'éducation nationale et de la jeunesse
MESR.....	Ministère de l'enseignement supérieur et de la recherche
MIAI	<i>Multidisciplinary Institute in artificial intelligence</i>
MITI.....	Mission pour les initiatives transverses et interdisciplinaires - CNRS
OCDE.....	Organisation de coopération et de développement économiques
ONR	Organisme national de recherche
OPECST.....	Office parlementaire d'évaluation des choix scientifiques et technologiques
OST	Observatoire des sciences et techniques
PEPR	Programmes et équipements prioritaires de recherche
PIA	Programme d'investissement d'avenir
PIB	Produit intérieur brut
PME-PMI.....	Petites et moyennes entreprises-petites et moyennes industries
PMIA.....	Partenariat mondial sur l'intelligence artificielle
PNRIA.....	Programme national de recherche en intelligence artificielle
PPTS	Dispositif de protection du potentiel scientifique et technique de la nation
PRAIRIE	<i>Paris artificial intelligence research Institute</i>
Prime	Projet de recherche interdisciplinaire multi-équipes
PSL.....	Paris Sciences et Lettres
PSSIE	Politique de sécurité des systèmes d'information de l'État-circulaire du 17 juillet 2014
PU	Professeur des universités
RIM.....	Réunion interministérielle -réunion d'arbitrage à Matignon
RNCP	Répertoire nationale des certifications professionnelles - répertoire tous les diplômes reconnus en France et délivrés par l'État
RSCH.....	Répertoire spécifique des certifications et habilitations – répertoire les formations complémentaires à l'exercice d'un métier ou des compétences transversales mobilisables dans certaines situations professionnelles
SATT.....	Société d'accélération de transfert technologique
SCAI	<i>Sorbonne center for artificial intelligence</i>
Scopus	Nom de la base de données transdisciplinaire de résumés et de citations de publications scientifiques lancée par l'éditeur scientifique Elsevier en 2004.

SGPI.....	Secrétariat général pour l'investissement
SI.....	Système d'informations
SNIA	Stratégie nationale en intelligence artificielle
SNRIA.....	Stratégie nationale de recherche en intelligence artificielle
UFTMiP	Université fédérale de Toulouse-Midi-Pyrénées
UMR	Unité mixte de recherche
Wos	<i>Web of science</i> ¹⁶⁰
ZRR.....	Zone à régime restrictif- régime relevant de la politique de protection du potentiel scientifique et technique de la nation

¹⁶⁰ Le WoS est l'une des principales bases de données utilisées en bibliométrie ; elle recense les revues scientifiques et les actes de colloques les plus influents au niveau international. Ils sont sélectionnés en fonction, à la fois, de leurs qualités éditoriales et de leur utilisation par les pairs mesurée à l'aide de citations (définition empruntée sur le site du HCERES-OST).

Annexes

Annexe n° 1 : liste des membres du comité d'accompagnement	133
Annexe n° 2 : lettre de mission Inria du 25 juin 2019.....	134
Annexe n° 3 : analyse d'ensemble des données internationales relatives à l'IA par des méthodes multicritères	136
Annexe n° 4 : approches d'apprentissage statistique supervisées et économétriques mises en œuvre pour analyser les déterminants de la performance et de l'efficience des pays en IA	144
Annexe n° 5 : analyses sémantiques et constitution de corpus de mots clés de l'IA ainsi que de ses applications.....	149
Annexe n° 6 : scientométrie à partir de sources complémentaires	158
Annexe n° 7 : exploitation de différentes sources de données sur les thèses pour suivre la dynamique et la structuration des forces scientifiques françaises en IA	166
Annexe n° 8 : analyses de réseaux pour caractériser et mesurer la structuration d'écosystèmes dans le domaine de l'IA	180
Annexe n° 9 : consultation de la communauté scientifique française dans le domaine de l'IA	200
Annexe n° 10 : évolutions des effectifs d'inscrits dans les filières d'élection préparant aux formations spécialisées en IA	212
Annexe n° 11 : cartographie de la formation continue en IA et du profil de ses apprenants avec une focale sur l'enseignement supérieur public.....	220
Annexe n° 12 : évolutions des effectifs de formateurs publics dans les secteurs de l'enseignement supérieur où l'IA est fortement représentée	223
Annexe n° 13 : bilan du financement des projets relatifs aux formations en IA dans le cadre de l'AMI CMA.....	227
Annexe n° 14 : compte-rendu de l'atelier des parties prenantes du 19 juillet 2022	228
Annexe n° 15 : synthèse des travaux de parangonnage international	233
Annexe n° 16 : monographie PRAIRIE	236
Annexe n° 17 : monographie 3IA Nice Sophia-Antipolis.....	239
Annexe n° 18 : monographie 3IA MIAI Grenoble	242
Annexe n° 19 : monographie 3IA ANITI Toulouse.....	245

Annexe n° 20 : monographie DATAIA	248
Annexe n° 21 : monographie Hi! PARIS	251
Annexe n° 22 : monographie SCAI	254

Annexe n° 1 : liste des membres du comité d'accompagnement

- Monsieur Dominique CARDON, Sciences Po, Directeur du Medialab
- Monsieur Thomas CLOZEL, Président directeur général et co-fondateur d'Owkin
- Madame Mireille FARES, Doctorante à l'ISIR, co-fondatrice et présidente de SOphIA
- Monsieur David FRABOULET, Ministère de l'enseignement supérieur et de la recherche, Directeur scientifique, directeur du département « mathématiques, physique, nano-sciences, STIC »
- Madame Martine GARNIER-RIZET, Agence nationale de recherche, Directrice de la stratégie numérique et des données
- Madame Alice GUILHON Directrice générale de Skema Business School
- Madame Isabelle HERLIN, directrice du centre d'expertise français du PMIA, coordinatrice de la SNRIA, déléguée générale d'Allistène, Inria
- Monsieur Xavier JARAVEL, Professeur d'économie à la *London School of Economics*, membre du Conseil d'analyse économique
- Madame Julia FENART, Directrice des affaires européennes, France Digitale puis M.Thomas BARREAU pour le 3^e comité
- Monsieur Bernard NORDLINGER, Membre de l'Académie nationale de médecine, Co-animateur du groupe de travail « Santé et IA » au sein de l'Académie
- Monsieur Philippe RAIMBAULT, Président de l'Université fédérale de Toulouse
- Monsieur Marc SCHOENAUER, Directeur de recherche à Inria
- Monsieur le préfet Renaud VEDEL, coordonnateur national de la stratégie nationale en intelligence artificielle puis M. Fabrice Le VOYER à partir du 3^e comité
- Madame Cécile WENDLING, Directrice de la stratégie de sécurité, de l'anticipation des menaces et de la recherche chez Axa

Annexe n° 2 : lettre de mission Inria du 25 juin 2019



*Le Ministre de l'Économie et des
Finances*

*La Ministre de l'Enseignement supérieur,
de la Recherche et de l'innovation*

*Le Secrétaire d'État chargé du
Numérique*

*Le Secrétaire général pour
l'Investissement*

Paris, le **25 JUIN 2019**

Monsieur le Président directeur général,

L'intelligence artificielle est devenue un enjeu majeur en termes de recherche, de formation et d'innovation, et le Président de la République a présenté en mars dernier une stratégie ambitieuse pour conforter et développer les atouts de notre pays dans ce domaine, notamment sur la base du rapport présenté par Monsieur le Député Cédric Villani. Comme annoncé à cette occasion, nous souhaitons confier à Inria la mission de coordination du volet recherche du programme national en IA (« programme national de recherche en IA »).

Cette mission de coordination comprendra notamment les tâches suivantes :

- Proposer et mettre en place une gouvernance adaptée pour le programme national de recherche en IA, avec un comité stratégique et un comité scientifique international.
- Proposer et coordonner la mise en œuvre par les opérateurs adaptés, de programmes de soutien tels que :
 - le développement d'instituts interdisciplinaires en intelligence artificielle (3IA) ;
 - un dispositif de chaires internationales et/ou industrielles ;
 - un dispositif de développement des doctorats en IA et plus généralement de développement des formations dans ce domaine ;
 - le financement de projets de recherche collaboratifs ;
 - un dispositif de soutien à des scientifiques et/ou des équipes de recherche, en complément des instituts interdisciplinaires en intelligence artificielle, pour permettre l'inclusion de l'ensemble des meilleurs chercheurs du territoire dans un ensemble cohérent ;
 - l'organisation de défis ambitieux en IA mobilisant les acteurs de la recherche publique et les entreprises ;
 - le déploiement de moyens de calcul dédiés en s'appuyant notamment sur le Grand Equipement National de Calcul Intensif, et particulièrement l'organisation d'un accès adapté aux spécificités de l'IA.

Monsieur Bruno SPORTISSE
Président Directeur général de l'INRIA
Domaine de Voluceau-Rocquencourt
BP10
78153 LE CHESNAY CEDEX

21 rue Descartes – 75231 Paris cedex 05 Tél : 01 55 55 90 90

.../...

- l'organisation de l'accès à des jeux de données pour la recherche ou le test ;
- le renforcement des synergies entre la recherche publique et la recherche privée française en IA en incitant davantage de chercheurs à utiliser plus largement la possibilité qui leur est offerte de consacrer une partie de leur temps à une entité privée.
- Assurer en tant que de besoin un rôle d'expertise scientifique et technologique au bénéfice des différentes actions de la stratégie IA nationale.
- Animer et coordonner le réseau des « 3IA » et plus généralement animer scientifiquement la communauté scientifique nationale dans le domaine.
- Participer au développement d'actions de coopérations bilatérales, européennes et internationales selon les lignes définies par le gouvernement dans ce domaine
- Faciliter le développement de l'activité économique nationale dans le domaine de l'IA à partir du transfert de résultats de recherche ; vous porterez une attention particulière à la création de start-ups dans ce cadre.

Vous voudrez bien présenter régulièrement au comité stratégique une vision consolidée et cohérente des mécanismes mis en place et des actions menées au titre de la présente mission. Au titre de celle-ci vous participerez au comité de pilotage interministériel qui suit la mise en œuvre de la stratégie nationale définie par le Président de la République.

INRIA étant lui-même opérateur de recherche, il conviendra que vous mettiez en place une direction de programme spécifique et adaptée pour l'exercice de cette mission, qui s'effectue au bénéfice de l'ensemble du système de recherche français. Inria veillera notamment à éviter toute situation de conflit d'intérêt.

INRIA recevra pour effectuer cette mission une dotation complémentaire de 500k€ annuels à compter de septembre 2018 et pour une durée de quatre ans, dont vous présenterez l'utilisation faite au comité stratégique chaque année.

Convaincus de votre engagement sans faille au service de ces objectifs, nous vous prions d'agréer, Monsieur le Président directeur général, l'expression de notre meilleure considération.



Bruno Le MAIRE



Cédric O



Frédérique VIDAL



Guillaume BOUDY

Annexe n° 3 : analyse d'ensemble des données internationales relatives à l'IA par des méthodes multicritères

I. Données internationales prises en compte

Afin de suivre le positionnement de la France aux échelons européen et mondial (sections II à VI ci-après), la Cour a croisé et analysé de façon conjointe les données brutes issues de l'Observatoire des politiques IA de l'OCDE et du IA Index Report de Stanford, qui compilent diverses sources (voir schéma ci-après). Un total de 333 variables est ainsi pris comme référence, dont 321 variables numériques et 12 variables textuelles ou assimilées.

Par ailleurs, pour évaluer la robustesse des résultats précédents tirés des données de l'OCDE et de Stanford, la Cour a mobilisé l'outil *InCites* de *Web of Science* pour suivre, sur la base de données indépendantes, l'évolution du classement international de la France en IA (section VII ci-après).

Schéma n° 4 : données internationales mobilisées

Recherche	Formation	Talents	Valorisation
<ul style="list-style-type: none"> • Microsoft Academic Graph (MAG) • Scopus • arXiv 	<ul style="list-style-type: none"> • Studyportal 	<ul style="list-style-type: none"> • LinkedIn Economic Graph • Adzuna 	<ul style="list-style-type: none"> • Github • MAG • NetBase Quid • Preqin

Source : Cour des comptes d'après l'Observatoire des politiques IA de l'OCDE et du IA Index Report de Stanford

II. Variables mobilisées pour les analyses

321 variables numériques directement issues de l'Observatoire des politiques IA et de l'IA Index Report ont été combinées. La Cour a ajouté à cet ensemble :

- 321 variables d'efficacité obtenues en normalisant les variables numériques précédentes par le PIB annuel de chaque pays. Les

variables normalisées sont ensuite multipliées par le PIB moyen pour l'ensemble des pays et années¹⁶¹ ;

- une variable qui décrit l'appartenance d'un pays à l'Union européenne au 1^{er} janvier 2022 ;
- une variable qui décrit l'appartenance d'un pays à l'OCDE au 1^{er} janvier 2022.

Au total, 644 variables numériques sont ainsi prises en compte dans les analyses quantitatives.

III. Période étudiée

La Cour s'est focalisée sur la période 2014-2021 pour plusieurs raisons :

- cette période permet de couvrir deux sous-périodes : une sous-période de quatre années (2014-2017) avant l'adoption de la stratégie nationale d'IA française (SNIA) ; une sous-période de quatre années qui suit l'adoption de la SNIA (2018-2021) ;
- le taux de renseignement des données est généralement bon sur cette période, à l'exception de l'année 2021 pour laquelle de nombreuses variables sont non renseignées. Cette année a toutefois été conservée pour être en capacité de fonder l'évaluation sur les dernières données disponibles et pour focaliser les analyses sur des indicateurs qui seront plus aisés à consolider en temps réel par les acteurs.

IV. Filtrages, imputations et corrélations

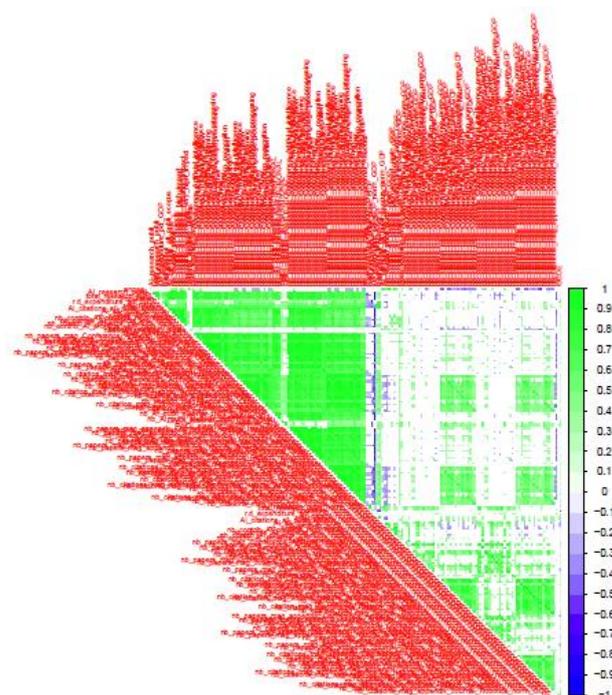
Sur la période 2014-2021, 52 % des données sont manquantes. Un ensemble d'opérations de filtrage et d'imputation a donc été mis en œuvre :

- les variables avec plus de 25 % de valeurs manquantes sont écartées. Ce filtrage revient à retenir 169 variables, avec un taux de données manquantes de 15 % ;
- les variables sans variance sont ensuite écartées. Cette opération revient à retenir 168 variables après filtrage, avec un taux de données manquantes de 15 % ;
- l'imputation des 15 % de valeurs manquantes est réalisée par la méthode de la médiane.

¹⁶¹ Cette opération, qui revient à multiplier les indicateurs normalisés par une constante, permet d'éviter les artefacts numériques liés aux très petits nombres.

Afin d'identifier les redondances et complémentarités entre les 168 variables, des calculs de corrélations ont été réalisés pour l'ensemble des paires de variables (cf. graphique n° 16).

Graphique n° 16 : corrélations entre les variables



Source : Cour des comptes. Le code couleur pour l'interprétation des corrélations est le suivant. En vert : les corrélations positives significatives. En bleu : les corrélations négatives significatives. En blanc : les corrélations non-significatives. Le seuil retenu pour la coloration est de 5 % divisé par le nombre de paires de variables considérées pour limiter le risque de faux positifs (tests multiples). La corrélation ici calculée est celle des rangs de Spearman

Appliquée à l'ensemble des données rassemblées relatives à l'IA, cette analyse indique qu'au global, les variables non-normalisées sont très fortement associées, ce qui traduit des redondances élevées. En effet, les variables non-normalisées sont situées dans le triangle vert situé sur la gauche du graphique, avec des niveaux de corrélations très élevés. L'ajout de variables normalisées, situées sur la droite du graphique, induit davantage de complémentarité dans les données, matérialisée par les bandes blanches.

V. Identification d'axes thématiques structurants par une analyse en composantes principales (ACP)

Formalisée par le statisticien et économiste Harold Hotelling dans les années 1940¹⁶², l'ACP est une méthode statistique d'apprentissage non-supervisé¹⁶³ qui permet de résumer l'information contenue dans un grand nombre de variables, souvent corrélées entre elles et partiellement redondantes, en quelques variables synthétiques, décorréées les unes des autres. Ces variables synthétiques sont appelées « composantes principales ».

Les composantes principales vont ici correspondre à des axes thématiques structurants, c'est-à-dire à de grands ensembles d'indicateurs permettant de cerner les principaux facteurs qui expliquent les différences observées entre les pays en matière d'IA.

Sur la base d'une analyse d'ensemble des indicateurs ainsi combinés, deux principaux axes thématiques discriminants¹⁶⁴ ont pu être identifiés (cf. le graphique ci-dessous), dans le sens où chacun de ces axes explique plus de 10 % de la variance observée entre les pays. Ces deux axes, qui expliquent à eux seuls près de 50 % des différences observées entre les pays, sont, par ordre décroissant :

- la production de la recherche en IA, pour près de 36 % des variations observées. Cet effet, qui s'assimile à une mesure de performance globale, se mesure par des indicateurs de volumes tels que le nombre d'articles, de conférences ou de citations associées (cf. figure ci-après). En pratique, ces différents indicateurs sont très fortement corrélés entre eux et à des variables d'échelle telle que le PIB (cf. matrice de corrélation ci-avant) ;
- l'efficacité de la recherche en IA, pour environ 14 % des variations observées. Cet axe se mesure par des indicateurs analogues à ceux du premier axe mais normalisés par le PIB des pays (cf. figure ci-après). En clair, certains pays « réussissent » même avec moins de moyens.

Si le poids des deux premiers axes est prépondérant, les poids des autres dimensions ne sont pas négligeables car il faut remonter au 11^e axe pour expliquer plus de 80 % de la variance observée. Cette situation traduit

¹⁶² Harold Hotelling "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 1933.

¹⁶³ Une forme d'« intelligence artificielle » qui ne nécessite pas d'étiquettes sur les observations, en l'occurrence les pays pour la période 2014-2021.

¹⁶⁴ Les thématiques sont discriminantes dans le sens où elles permettent de distinguer aisément les différents pays au sens des indicateurs pris en compte. En pratique, plus un axe est discriminant, plus son influence sera élevée dans les classements des pays.

La structure des deux premiers axes implique que la recherche joue un rôle prépondérant dans le positionnement des différents pays, loin devant la valorisation et probablement la formation¹⁶⁵.

VI. Sélection d'indicateurs pour la production de classements thématiques

En lien avec les résultats des ACP, des calculs de corrélations pour l'ensemble des paires de variables permettent d'identifier des indicateurs stratégiques et complémentaires les uns des autres. Ces indicateurs permettent par exemple de produire des classements thématiques ou sectoriels des pays pour chaque indicateur. Ces classements ciblés apportent une information plus intéressante qu'un classement général car ils permettent d'établir un diagnostic critique des forces et les faiblesses des pays, comme le représente le tableau ci-après. En outre, quand ils sont employés de façon combinée, ces indicateurs permettent d'établir un classement général.

Pour la production des classements, la Cour a retenu quatre indicateurs sur la base de l'ACP et des indicateurs qui apparaissent comme les plus aisés à calculer et interpréter :

- deux indicateurs associés au 1^{er} axe : le nombre d'articles scientifiques en IA (« AI_research_publi ») ainsi que le nombre de citations de ces articles (« AI_citations_all_publi_scopus ») ;
- deux indicateurs associés au 2^e axe : les deux indicateurs précités normalisés par le PIB (« AI_research_publi_norm_GDP » et « AI_citations_all_publi_scopus_norm_GDP » respectivement).

Les résultats relatifs aux rangs de la France (cf. corps de texte) confirment que les classements induits par les indicateurs au sein d'un même axe sont globalement semblables alors qu'ils sont différents entre indicateurs d'axes distincts.

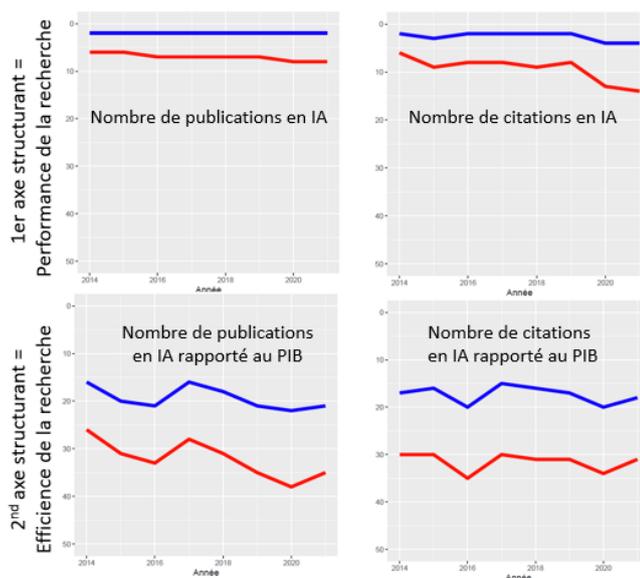
¹⁶⁵ Il n'est pas possible de conclure dans le cas de la formation, puisque les variables associées sont peu renseignées et ont donc été préalablement filtrées. Une analyse portant sur des années mieux renseignées en termes de données (période 2018-2020) suggère que la formation jouerait également un rôle limité en comparaison de la recherche.

VII. Comparaison des résultats avec ceux obtenus à partir de Web of Science

L'analyse de l'ensemble des éléments indexés dans *Web of Science* en IA (voir graphiques n° 18 et 19 pour le schéma principal « *Web of Science* »¹⁶⁶, au titre de la catégorie « *Computer Science, Artificial Intelligence* ») conduit à des résultats analogues à ceux rapportés dans le corps du texte (section I.A).

Deux autres catégorisations proches du cœur de l'IA ont été employés¹⁶⁷ et conduisent à des résultats convergents avec ceux rapportés dans le corps de texte (voir archive en ligne).

Graphique n° 18 : évolution des rangs de la France dans le monde (en rouge) et en UE à 27 (en bleu) suivant des critères complémentaires et différenciant sur la scène internationale (données *Web of Science*)



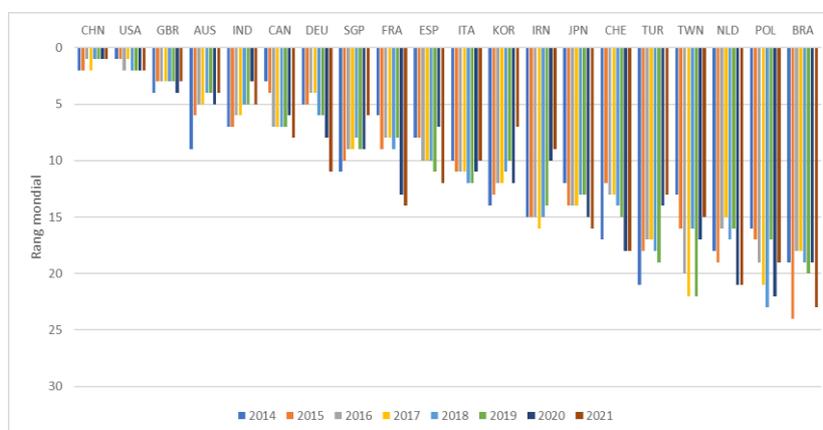
Source : Cour des comptes d'après l'outil InCites interrogé en septembre 2022. Le périmètre retenu correspond à l'ensemble des documents indexés par Web of Science au 31 juillet 2022 et appartenant

¹⁶⁶ Voir <https://incites.help.clarivate.com/Content/Research-Areas/wos-research-areas.htm>.

¹⁶⁷ Le schéma « *Citation topics* » a également été mobilisé au titre des deux catégorisations suivantes : la catégorie « *4.61 Artificial Intelligence & Machine Learning* » de l'échelle « *Meso* » ainsi que la catégorie « *4.17.128 Deep Learning* » de l'échelle « *Micro* ». Voir <https://incites.help.clarivate.com/Content/Research-Areas/citation-topics.htm>.

au domaine de recherche « Computer Science, Artificial Intelligence ». Les rangs sont suivis sur la période allant de 2014 à 2021. S'agissant des indicateurs relatifs à l'efficacité de la recherche, il est à noter que la crise de la covid 19 a pu impacter les PIB des pays à compter de 2020. Le suivi des rangs des pays permet de limiter les biais associés à la normalisation par les PIB annuels, comme le montre la relative stabilité des deux séries temporelles concernées sur la période 2019 à 2021

Graphique n° 19 : évolution du rang relatif des vingt premiers pays en IA au sens des citations de 2014 à 2021



Source : Cour des comptes d'après l'outil InCites interrogé en septembre 2022. Le périmètre retenu correspond à l'ensemble des documents indexés par Web of Science au 31 juillet 2022 et appartenant au domaine de recherche « Computer Science, Artificial Intelligence ». Les rangs sont suivis sur la période allant de 2014 à 2021

Annexe n° 4 : approches d'apprentissage statistique supervisées et économétriques mises en œuvre pour analyser les déterminants de la performance et de l'efficacité des pays en IA

I. Exposé des principaux enjeux

Dans la mesure où les données internationales mobilisées contiennent une myriade d'informations associées, il est possible de « reconstruire »¹⁶⁸ les relations entre ces différentes données par une méthode statistique bien choisie.

À titre d'illustration, s'il est observé que les pays qui publient le plus en IA sont les pays les plus riches au sens du PIB, il est possible de supposer que le PIB joue un rôle important pour rendre compte du volume de publication. La question de l'effet de l'adoption d'une stratégie en IA sur le nombre de publications dans ce domaine peut être posée de manière analogue.

La méthode décrite ci-dessous met en œuvre ce principe de façon automatisée dans le cas où un très grand nombre de critères potentiels sont examinés de façon concomitante pour l'ensemble des pays.

II. Données prises en compte

Les données considérées sont identiques à celles de l'étape IV de l'annexe 3.

Seules les variables corrélées à moins de 99 % sont retenues pour la suite des analyses supervisées¹⁶⁹, soit un ensemble de 120 variables (suppression de 48 variables redondantes).

III. Description de l'algorithme d'apprentissage supervisé employé

Il est possible de « modéliser » les relations entre les différentes variables par une méthode statistique qui va tirer parti des associations entre les variables clés dont on cherche à rendre compte, d'une part, et de leurs déterminants potentiels, d'autre part.

¹⁶⁸ Le verbe adéquat est « inférer ».

¹⁶⁹ Les variables redondantes sont identifiées avec la fonction findCorrelation du package « Caret » R, avec les paramètres cutoff = 0.99 et exact = T.

Sur un plan mathématique, nous cherchons à construire un modèle appelé régresseur de la forme $[y = F(x_1, \dots, x_i, \dots, x_n)]$ permettant de déterminer si les prédictors $x = (x_1, \dots, x_i, \dots, x_n)$ associés aux pays jouent un rôle pour rendre compte d'une variable d'intérêt associée à l'IA telle que le nombre de publications dans ce domaine (y , dite « étiquette »).

Ce modèle (F), peut être obtenu par un algorithme d'apprentissage supervisé à partir des étiquettes (y) et des prédictors (x). Ici le modèle F est donné par une approche de régression non-paramétrique appelée « forêt aléatoire »¹⁷⁰ à partir du package « ranger » de R¹⁷¹.

L'ensemble d'apprentissage est obtenu en tirant au hasard deux tiers des observations formés par les couples des pays et années. Concernant l'hyper-paramétrage de la forêt, les valeurs usuelles sont employées avec 1 000 arbres par forêt et un nombre de variables testées à chaque division fixée au nombre de prédictors divisé par trois.

L'erreur du modèle F entraîné est calculée à partir d'un ensemble test comportant le tiers des observations qui n'a pas été utilisé lors de la phase d'entraînement (données indépendantes). Deux mesures de mesure de l'erreur sont employées : l'erreur absolue moyenne et l'erreur quadratique. Le recours à des tests de permutation non-paramétrique avec 1 000 répétitions permet de s'assurer que le modèle entraîné est plus performant que le hasard au sens de ces deux critères d'erreur¹⁷². En outre, un contrôle graphique de la précision est opéré en comparant les valeurs prédites et les valeurs réelles en regard de la première bissectrice.

Une fois la forêt entraînée, les déterminants potentiels des étiquettes « y » sont obtenus en mesurant l'importance des variables quantifiées par l'accroissement du pourcentage d'erreur moyenne quadratique après permutation¹⁷³ de chaque prédictor dans les arbres composants F. Afin de déterminer la robustesse des importances ainsi calculées, une p-valeur est calculée pour chaque variable suivant la méthode de Altmann¹⁷⁴.

¹⁷⁰ Breiman, « Random forests », *Machine learning*, 2001.

¹⁷¹ « Ranger, A Fast Implementation of Random Forests », <https://cran.r-project.org/web/packages/ranger/ranger.pdf>

¹⁷² P-valeur calculée à partir de 1 000 prédictors à information nulle donnés par 1 000 tirages bootstraps dans l'ensemble d'apprentissage ; tous les modèles ici rapportés sont significativement plus précis que le hasard tant pour ce qui concerne l'erreur absolue moyenne que l'erreur quadratique, avec $p < 0.1$ %.

¹⁷³ Une permutation correspond à un changement au hasard de l'ordre de succession des valeurs du déterminant potentiel sous revue.

¹⁷⁴ Cf. Altmann et al (2010). “ *Permutation importance : a corrected feature importance measure* ” *Bioinformatics*. La méthode d'Altmann est implémentée dans le package « Ranger ».

IV. Résultats de l'apprentissage supervisé

La méthode de régression ainsi décrite est mise en œuvre pour identifier les déterminants potentiels du nombre de publication en IA et du nombre de publication en IA rapporté au PIB en regard des 119 autres variables prises en compte.

Sans surprise, le nombre de publication en IA est principalement associé à des variables d'échelle telles que les publications totales et le PIB et le nombre de publication en IA rapporté au PIB est principalement associé à des variables normalisées, tout particulièrement les publications totales rapportées au PIB.

Le fait d'avoir adopté une stratégie en IA joue un rôle marginal pour ces deux types d'indicateurs.

V. Analyse économétrique

Une analyse économétrique permet de mesurer l'effet de la stratégie sur les leviers institutionnels de productions scientifiques et de brevet. Un modèle de régression linéaire a été utilisé pour expliquer le nombre d'articles scientifiques et le nombre de brevets produits par un pays par :

- le niveau de production scientifique total du pays ;
- le niveau annuel moyen de production mondiale (effet fixe année) ;
- les caractéristiques spécifiques de chaque pays constantes au cours du temps en termes de production scientifique (effet fixe pays).

Tableau n° 26 : mesure de l'impact d'une stratégie sur la production scientifique et de brevets en IA de 2014 à 2021 par analyse de différence en différence

<i>Variables</i>	NB publications IA (1)	NB publications IA/PIB (2)	NB Brevets IA (3)	NB Brevets IA/PIB (4)
<i>Total Publication</i>	0.18*** (0.01)			
<i>Total publication x stratégie</i>	0.01*** (0.00)			
<i>Total publication/PIB</i>		0.23*** (0.02)		
<i>Total publication/PIB x stratégie</i>		0.01*** (0.00)		
<i>Total Brevet</i>			0.18*** (0.01)	
<i>Total Brevet x stratégie</i>			0.04*** (0.00)	
<i>Total Brevet / PIB</i>				0.09*** (0.02)
<i>Total Brevet / PIB x stratégie</i>				-0.01 (0.01)
<i>PIB</i>	0.00 (0.00)	0.00*** (0.00)	0.00 (0.00)	-0.00 (0.00)
<i>Population</i>	0.00*** (0.00)	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
<i>Stratégie</i>	-0.00** (0.00)	-459.97** (174.19)	-24.65 (32.80)	0.00 (0.00)
Observations	376	376	376	376
Nb Groupe	47	47	47	47
R-squared	0.99	1.00	1.00	0.78

Sources : Analyse de la Cour des comptes à partir des données de l'OCDE. Le modèle économétrique explique le nombre de publications ou de brevets (1 et 3), corrigé du PIB (2 et 4) en IA par la production scientifique (ou brevet) totale, le PIB, la population, et les effets fixes années et pays. Le coefficient mesure la contribution de chaque variable explicative à celle expliquée : par exemple, dans la colonne 1, l'augmentation du nombre de publications scientifiques de 100 est corrélée de manière significative à une augmentation de 18 publications en IA. Le modèle économétrique est un modèle à effet fixe pays et année pour prendre en compte l'hétérogénéité temporelle et spatiale. La corrélation entre données est prise en compte pour le calcul des écart-types (technique des clusters). La significativité des résultats (effet de la variable significativement différent de zéro) est indiquée par les étoiles : * <5 %, ** <1 % et *** <0,1 %.

En utilisant des données depuis 2014, il est possible de mesurer un effet moyen relatif à la formalisation d'une stratégie en comparant les niveaux de production avant et après l'adoption de celle-ci. Étant donné que le fait qu'adopter une stratégie est endogène à des caractéristiques de chaque pays, les effets estimés ne peuvent pas être considérés comme causaux mais donnent une idée de la corrélation entre avoir une stratégie et la production scientifique.

Cette étude est également reproduite en rapportant la production scientifique au PIB. Le tableau ci-dessous présente les résultats des variables principales. Les astérisques donnent la significativité de la corrélation. Afin de prendre en compte les liens entre observations d'un même pays (le PIB de l'année N est corrélé à celui de l'année N-1), la significativité des résultats est corrigée par la méthode des clusters.

En tenant compte de l'hétérogénéité structurelle entre pays, l'analyse économétrique permet de montrer que la mise en place d'une stratégie nationale est positivement corrélée à la productivité d'un pays en termes de publications et de brevets (respectivement 5,60 % et 22,22 %). Corrigé du PIB, cet effet diminue à 4,35 % pour la production scientifique et n'est plus significatif pour les brevets. Cela converge vers la conclusion d'une forte corrélation entre pays les plus développés et mise en place d'une stratégie.

Annexe n° 5 : analyses sémantiques et constitution de corpus de mots clés de l'IA ainsi que de ses applications

Enjeux et champs d'application des analyses sémantiques

L'intelligence artificielle (IA) regroupe l'ensemble des outils, techniques et technologiques fondés sur les algorithmes permettant d'accomplir une tâche, de formuler une proposition, de prendre des décisions, dans un système multiparamétrique et dans une dynamique d'apprentissage.

Dans le cadre de l'évaluation de politique publique, certaines observations s'appuient sur des analyses sémantiques rendues possibles grâce à des filtrages par mots clés¹⁷⁵ afin d'isoler les éléments qui relèvent spécifiquement de l'IA ainsi que de ses applications.

Pour produire des indicateurs adaptés aux enjeux de l'évaluation, ces analyses sémantiques sont appliquées à des éléments de nature diversifiée : publications scientifiques (HAL, *Web of Science*), thèses (CIFRE, theses.fr), formations et appels à projets nationaux et européens (ANR et E-Corda).

Principes guidant la constitution de corpus pour les analyses

Deux corpus principaux ont été constitués : le corpus lié au champ lexical du « cœur » scientifique et technologique de l'IA ainsi que le corpus des applications de l'IA. Les filtres sémantiques peuvent être appliqués à l'un, à l'autre ou à l'intersection de ces deux corpus.

Le champ lexical du cœur de l'IA, qui suscite de vifs débats entre les experts du domaine¹⁷⁶, a fait l'objet d'une attention particulière pour limiter les risques de biais et été obtenu par un process collaboratif détaillé ci-dessous.

¹⁷⁵ Le terme « mot clé » est employé pour faciliter la lecture. Dans le cas général, ce sont en réalité des expressions clés qui sont employées pour les analyses sémantiques. Par exemple, les termes « intelligence » et « artificielle », employés de façon isolée, ne sont pas propres à l'IA. C'est l'expression « intelligence artificielle » qui est donc employée pour les analyses sémantiques. Les groupes de mots ainsi obtenus ne sont pas séparés pour limiter les faux positifs en lien avec le cœur scientifique et technologique de l'IA.

¹⁷⁶ Notamment entre les partisans des écoles « données / apprentissage » (dont la communauté des « connexionnistes » qui s'intéresse plus particulièrement aux réseaux de neurones profonds) de l'IA, d'une part, et l'école « symbolique » de l'IA (école historique qui a précédé l'IA par apprentissage), d'autre part, avec les partisans de l'IA « hybride » qui cherchent à combiner ces deux grandes approches du domaine.

Le champ lexical des applications est obtenu à titre principal à partir des sites web des quatre instituts 3IA et de trois instituts spécialisés en IA comparables aux 3IA¹⁷⁷. Les versions bilingues des mots sont considérées et le cas échéant obtenues par traduction. Seules les racines des mots sont gardées pour disposer d'une vision large des applications.

Afin de compléter ces deux corpus, les mots-clés employés par les sociétés savantes françaises et européennes actives dans le domaine de l'IA ont été mobilisés¹⁷⁸.

Une approche bilingue française et anglaise a été mise en œuvre avec une traduction systématique des mots dans les deux langues lorsque des traductions existent. En effet, la langue préférentiellement employée dans les données analysées est soit l'anglais (cas des données scientométriques et des appels à projets européens), soit le français (cas des formations et des appels à projets nationaux), avec parfois une combinaison des deux langues retrouvées au sein d'un jeu de donnée particulier.

Une démarche collaborative pour identifier les mots clés au cœur de l'IA

Pour retenir un ensemble de mots clés visant à caractériser l'IA de façon précise et opérationnelle dans l'optique d'étayer certaines observations quantitatives de l'évaluation, une démarche collaborative et itérative a été mise en œuvre. Cette approche s'apparente à une méthode d'élicitation d'experts, une méthode employée dans le domaine des statistiques mobilisées en appui à la décision publique¹⁷⁹.

Les conseils de l'OCDE ainsi que des membres du comité d'accompagnement ont permis de retenir un ensemble large de sources qui traitent de l'IA dans ses différentes composantes sur la période 2014-2021. Les sources sélectionnées sont les suivantes :

- l'ensemble des sous-catégories et pages bilingues rattachées aux catégories « apprentissage automatique » (en français) et « *machine learning* » (en anglais) de Wikipedia ;
- un échantillon des titres des sessions¹⁸⁰ et des titres des interventions des conférenciers invités¹⁸¹ dans cinq des plus grandes conférences internationales en IA ainsi que dans trois conférences internationales

¹⁷⁷ DATAIA, SCAI et Hi !Paris.

¹⁷⁸ Liste des 25 ensembles de mots-clés liés à de l'IA employée par la Société française d'IA (SFIA) ; échantillon de mots clés associés aux conférences sur l'apprentissage automatique soutenues par la Société savante francophone d'apprentissage machine (SSFAM) ; mots apparaissant sur les pages des association CLAIRE (*Confederation of artificial Intelligence research laboratories in Europe*) et du réseau d'excellence ELLIS (*European laboratory for learning and intelligent systems*).

¹⁷⁹ Granger Morgan, « Use (and abuse) of expert elicitation in support of decision making for public policy », PNAS, 2014.

¹⁸⁰ Lorsque les sessions n'avaient pas de noms, les titres de certains exposés ont été retenus.

¹⁸¹ « *keynote speakers* ».

de référence dans des domaines connexes (voir tableau ci-dessous et sa légende qui décrit la logique du choix des sources à partir des propositions des acteurs) ;

- un échantillon des mots-clés employés par les sociétés savantes françaises et européennes actives dans le domaine de l'IA ;
- un échantillon des mots clés issus des pages web des instituts 3IA et d'instituts spécialisés en IA comparables aux 3IA.

Tableau n° 27 : liste des conférences internationales retenues par la Cour pour identifier des mots clés au cœur de l'IA

<i>Domaines de l'IA</i>		
<i>Nom de la conférence internationale</i>	<i>Domaine principal (/ sous domaines principaux)</i>	<i>Lien web</i>
<i>International Joint Conference on Artificial Intelligence (IJCAI)</i>	IA, y compris IA symbolique	ijcai.org
<i>AAAI Conference on Artificial Intelligence (AAAI)</i>	IA	aaai.org
<i>International Conference on Machine Learning (ICML)</i>	IA / apprentissage automatique	icml.cc
<i>International Conference on Learning Representations (ICLR)</i>	IA / apprentissage automatique / apprentissage profond	iclr.cc
<i>Neural Information Processing Systems (NeurIPS)</i>	IA / apprentissage automatique / apprentissage profond	nips.cc
<i>Domaines connexes à l'IA</i>		
<i>Nom de la conférence internationale</i>	<i>Domaine principal (/ sous domaines principaux)</i>	<i>Lien web</i>
<i>International conference on Knowledge Discovery and Data mining (KDD)</i>	Fouille de données, dont par IA	kdd.org
<i>Conference on Computer Vision and Pattern Recognition (CVPR)</i>	Vision par ordinateur (avec de nombreux travaux en IA, notamment en apprentissage profond)	cvpr2021.thecvf.com (pour l'année 2021)
<i>Meeting of the Association for Computational Linguistics (ACL)</i>	Traitement automatique du langage, avec de nombreux travaux en IA, notamment en apprentissage profond	2021.aclweb.org (pour l'année 2021)

Source : Cour d'après les recommandations des acteurs rencontrés. Parmi les ensembles de conférences pointées par les acteurs en IA, l'équipe a privilégié les conférences qui se tiennent annuellement à l'échelle mondiale, qui sont classées A* au classement CORE¹⁸² et classées dans le top 10 du classement Google Scholar pour l'IA¹⁸³, et qui permettent de couvrir un ensemble de sous-domaines de l'IA, ainsi que des conférences internationales de référence de domaines connexes à l'IA.

¹⁸² <http://portal.core.edu.au/conf-ranks/>

¹⁸³ https://scholar.google.fr/citations?view_op=top_venues&hl=fr&vq=eng_artificialintelligence

Au total, un ensemble de plus de 700 mots clés employés en IA a pu être rassemblé. Certains mots clés n'étant pas spécifiques à l'IA, ils doivent être écartés pour limiter le risque de faux positifs. Cette méthode privilégie la précision ; elle ne prétend pas couvrir l'IA de façon exhaustive.

Le corpus de l'IA a ensuite été divisé en sous-corpus :

- IA de type « apprentissage automatique » ou « *machine learning* » en anglais ; en identifiant de façon spécifique les termes propres à l'apprentissage profond ou « *deep learning* » qui constitue le sous-ensemble de l'apprentissage machine actuellement le plus prévalant dans les sources scientifiques en IA. L'apprentissage automatique se fonde sur l'exploitation de données dans le cadre d'une démarche inférentielle ;
- IA de type « symbolique », un domaine de l'IA qui se fonde sur la manipulation de représentations symboliques ainsi que de règles logiques dans le cadre d'une démarche hypothético-déductive ;
- IA hybride, qui combine des approches des deux types d'IA précédents ;
- certains mots clés sont communs à toutes les formes d'IA ;
- enfin, quelques mots clés sont issus de domaines fortement connexes à l'IA tels que la fouille de données, la vision par ordinateur et le traitement automatique du langage naturel.

Un sous-ensemble de 122 mots clés en anglais et français a été synthétisé en tentant de respecter un équilibre entre les différentes approches de l'IA, notamment entre IA symbolique (16 mots ; 13 %), IA hybride (12 mots ; 10 %) et IA par apprentissage (74 mots soit 61 %, dont 22 spécifiquement, soit 18 %, au titre de l'apprentissage profond qui constitue le sous-domaine le plus représenté dans les sources consultées). Les autres mots clés sont partagés par toutes les formes d'IA (10 mots ; 8 %) ou issus de domaines connexes (10 mots ; 8 %).

Les membres du comité ont ensuite été appelés à s'exprimer sur la pertinence de chacun des 122 mots clés à travers une démarche d'élicitation quantitative. En pratique, le recueil des votes a été opéré par voie électronique avec une interface permettant de contrôler les biais de présentation (présentation des mots clés dans un ordre aléatoire) et de conformité sociale (les votants ne connaissent pas les votes des autres votants). Il a été demandé aux votants d'indiquer pour chaque mot clé s'il se situe au cœur scientifique et technique de l'IA. Les votants ont pu choisir entre cinq modalités de réponses qui ont ensuite été interprétées par la Cour en probabilités que le mot soit au cœur de l'IA, par le biais d'une conversion à travers une échelle numérique (cf. tableau n° 28). Pour chaque

mot, la moyenne des votes exprimés et convertis en nombres a été prise, ce qui conduit à retenir une probabilité pour chaque mot clé où 0 représente un rejet unanime et 1 une adoption unanime, avec un continuum de valeurs possibles entre ces deux extrêmes.

Tableau n° 28 : choix possibles pour indiquer si un mot clé est au cœur scientifique et technique de l’IA et probabilités associées.

<i>Choix possible pour indiquer si le mot est au cœur de l’IA (modalité de réponse au vote proposée par la Cour aux membres du comité)</i>	<i>Probabilité correspondante que le mot clé soit au cœur de l’IA (interprétation numérique donnée par la Cour pour la suite des analyses)</i>
<i>Oui, tout à fait</i>	1
<i>Oui, plutôt</i>	2/3
<i>Non, plutôt pas</i>	1/3
<i>Non, pas du tout</i>	0
<i>Je ne connais pas ce terme</i>	Nombre manquant (NA)

Source : Cour des comptes.

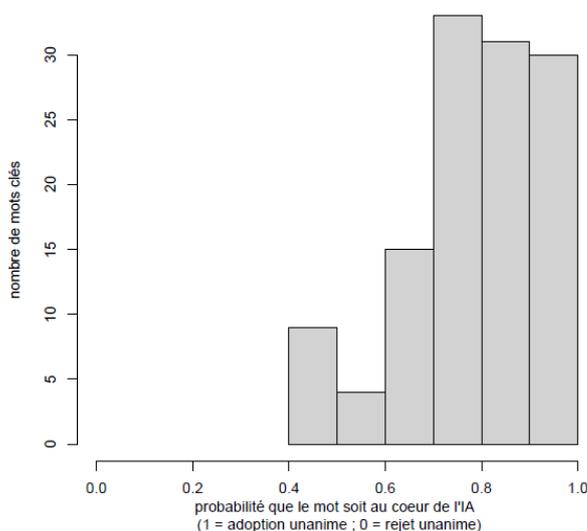
Suivant la recommandation d’un membre du comité d’accompagnement, un mot clé récent de l’apprentissage profond a été ajouté postérieurement à la procédure de vote. Par souci d’équité, ce mot a été pris en compte dans la suite des analyses sous la forme d’un « prior non-informatif », c’est-à-dire en lui appliquant un poids de 0,5. Un sondage international de grande ampleur réalisé en octobre 2022 est venu confirmer l’importance croissante, sur la période 2019 à 2022, du mot clé ainsi ajouté¹⁸⁴. Au total, 123 mots clés ont ainsi été rassemblés et 122 ont fait l’objet d’un vote.

Résultats des votes

Dans l’ensemble, les votes du comité d’accompagnement ont permis de conforter les mots clés proposés par la Cour avec une probabilité moyenne de 80 %. Une part importante de mots clés bénéficie d’un haut niveau d’adhésion (cf. figure). En outre, aucun mot clé n’a été unanimement rejeté, c’est-à-dire que chaque mot a reçu au moins la mention « Oui, plutôt » à la question de savoir s’il est au cœur de l’IA. La Cour a donc mobilisé l’ensemble des mots clés pour la plupart des analyses sémantiques.

¹⁸⁴ Voir <https://www.kaggle.com/kaggle-survey-2022>, page 20 : “Transformer architectures are becoming more popular for deep learning models (both image and text data)”

Graphique n° 20 : répartition des votes moyens obtenus pour les mots clés de l'IA.



Source : Cour des comptes

Les votes n'ont pas drastiquement modifié la répartition des mots clés proposés l'échelle des principaux sous-domaines de l'IA (cf. tableau ci-après).

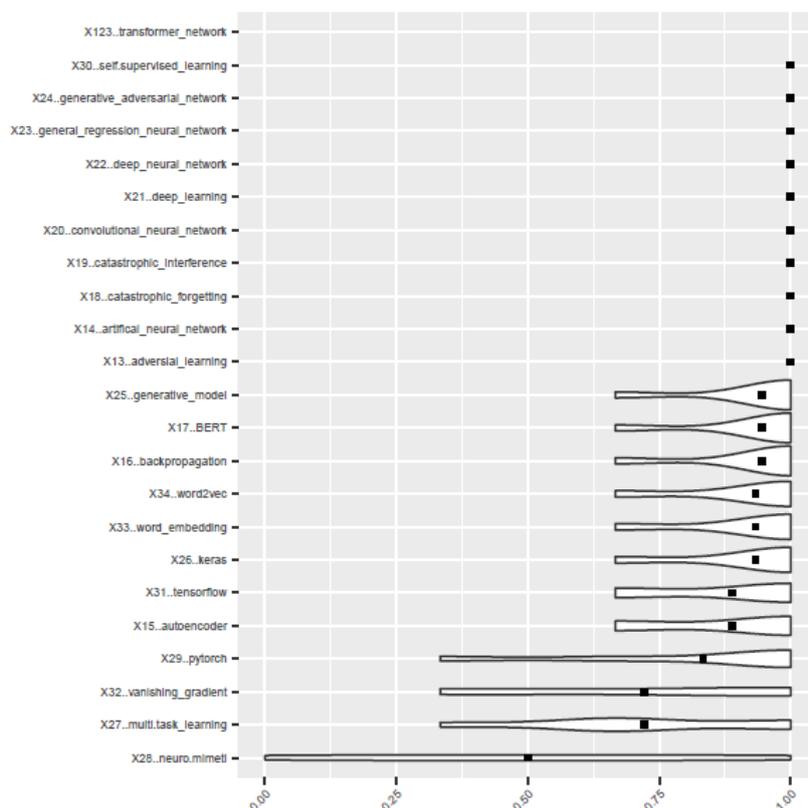
Tableau n° 29 : résultats des votes par sous-corpus de l'IA

Sous-corpus de l'IA	Moyenne des votes obtenus (en %)
Apprentissage profond (spécifiquement)	92 %
Apprentissage automatique	85 %
IA hybride	78 %
Approches partagées par toutes les formes d'IA	71 %
IA symbolique	68 %
Domaines connexes à l'IA	56 %

Source : Cour des comptes.

Les corpus de l'IA par apprentissage ont généralement été privilégiés aux autres catégories de mots clés, notamment de l'IA symbolique. Ce résultat est cohérent avec le fait que l'apprentissage (apprentissage

Graphique n° 22 : distributions de probabilités pour les mots clés de l'apprentissage profond



Source : Cour des comptes. L'axe horizontal représente la probabilité associée au mot (0 à gauche, 1 à droite). L'axe vertical représente les mots clés examinés. À l'exception du mot clé « transformer network » qui n'a pas fait l'objet d'un vote, les mots clés sont classés des plus plébiscités en haut au moins plébiscité en bas. Le carré noir représente la moyenne des votes exprimés. La courbe représente la distribution des votes avec une largeur proportionnelle aux votes obtenus, ce qui permet de visualiser les dissensus éventuels. Une absence de courbe signifie un consensus parfait

Filtrages sémantiques

La recherche de correspondance dans les différentes données analysées est effectuée avec la fonction « grepl » de R en ignorant la casse des chaînes de caractères. Cette méthode déterministe est aisée à mettre en œuvre et à expliquer. À la différence de méthodes probabilistes, elle a l'avantage de permettre une reproductibilité et une explicabilité complète des résultats.

Analyses de sensibilité

Pour analyser la portée des observations les plus importantes, tout particulièrement la performance des instituts et centres français en IA en lien avec les principaux acteurs universitaires des sites associés, des analyses de sensibilité ont été mises en œuvre afin de mesurer la robustesse des résultats suivant les corpus et mots employés. Deux méthodes principales ont été mobilisées :

- l'étude de la variation des résultats lorsque les mots-clés de l'IA sont tirés au hasard proportionnellement aux poids donnés par les experts. Ces procédures ont été répétées un grand nombre de fois de façon indépendante ;
- l'analyse de la variation des résultats suivant la nature du corpus employé, en prenant l'ensemble des mots clés d'une sous-catégorie donnée.

Annexe n° 6 : scientométrie à partir de sources complémentaires

I. Préambule

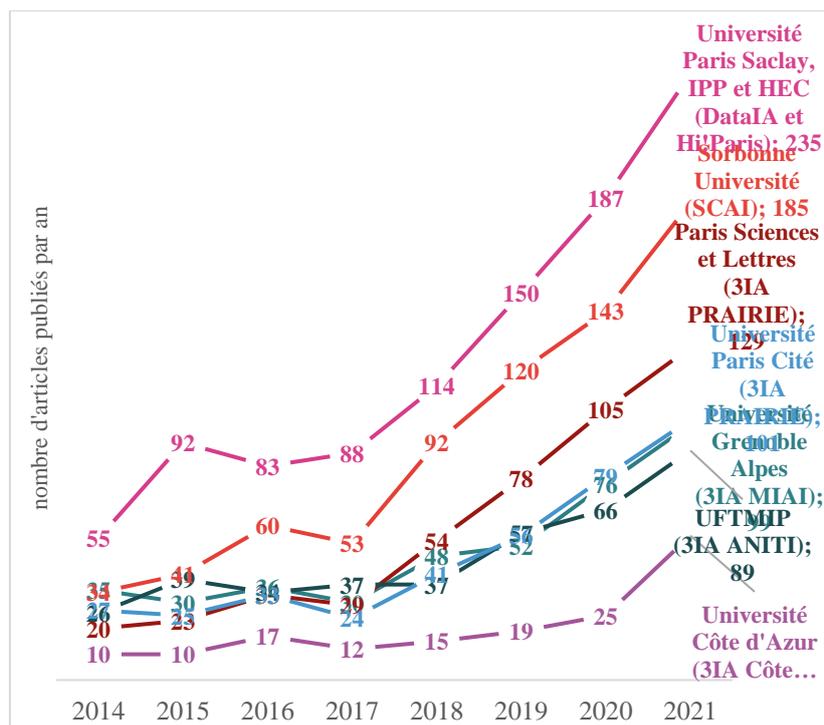
Deux types de données scientométriques complémentaires sont mobilisées : l'archive HAL (voir II) et les ressources de *Web of Science* (WoS ; voir III). Si l'archive HAL n'est pas exhaustive au niveau national, le dépôt est obligatoire pour les 3IA, avec des archives spécifiques pour chacun des quatre instituts. Quant à la base WoS, si elle a l'avantage sur HAL d'être exhaustive, y compris au niveau international, la production des 3IA n'y est pas retracée directement. En pratique, les indicateurs issus des bases HAL et WoS sont très corrélés (voir IV).

II. Exploitation des données ouvertes de HAL pour suivre la production scientifique associée aux 3IA en lien avec les sites associés

Résultats

Les résultats des analyses sémantiques sont ici rapportés. Les données mobilisées et les méthodes mises en œuvre pour l'analyse de la recherche en IA sont détaillées ci-après.

Graphique n° 23 : production scientifique des principaux sites universitaires associés aux 3IA et aux centres d'excellence en IA (données issues de HAL avec analyse sémantique de la Cour)



Source : Cour des comptes à partir des données et méthodes détaillées ci-dessous. * À la différence des universités et regroupements d'établissements pour lesquels un filtre sémantique relatif à l'IA est appliqué (cœur scientifique et technique), les nombres d'articles calculés pour les 3IA correspondent aux totaux de leurs archives HAL respectives, sans application de filtre sémantique. Les données associées à cette figure et les indicateurs associés sont disponibles sur l'archive électronique en ligne

Corpus HAL constitués ou mobilisés pour les analyses

Pour l'analyse de la production scientifique sur la période 2014-2021 (années civiles), un corpus d'articles internationaux a été constitué à partir de l'archive ouverte HAL.

Les archives HAL des instituts 3IA, listées sur leurs sites web respectifs, ont été extraites de façon intégrale à compter d'avril 2022 à

travers des requêtes paramétrées par la Cour¹⁸⁵. Les champs choisis pour les requêtes sont tirés du dictionnaire de HAL¹⁸⁶.

Dans la mesure où les chercheurs ne déposent pas nécessairement leurs articles dans HAL au moment de leur production, et afin d'éviter un biais de comparaison liés à des requêtes sur HAL opérées à des dates distinctes, un filtrage des articles déposés jusqu'à la fin mars 2022 inclus a été appliqué¹⁸⁷.

Les partenaires des 3IA ont été identifiés et classés par ordre de fréquence décroissante de mentions dans les articles. Les archives HAL associées ont ensuite été récupérées pour le principal partenaire universitaire du site¹⁸⁸ et/ou le porteur institutionnel du 3IA. Les identifiants HAL des acteurs concernés ont été retracés à partir des données de référence de HAL à travers le référentiel des structures¹⁸⁹, avec un

¹⁸⁵ Les fichiers sont obtenus par la requête suivante tapée depuis un navigateur web : « [https://api.archives-ouvertes.fr/search/3IA-COTEDAZUR/?q=*&fq=publicationDateY_i:\[2014+TO+2021\]&fq=docType_s:ART&fq=peerReviewing_s:1&classification_t:!NR&fq=language_s:en&rows=10000&wt=csv&fl=docid,publicationDateY_i,submittedDateY_i,submittedDateM_i,submittedDate_s,label_s,keyword_s,domain_s,title_s,anrProjectReference_s,authStructId_i,structAcronym_s](https://api.archives-ouvertes.fr/search/3IA-COTEDAZUR/?q=*&fq=publicationDateY_i:[2014+TO+2021]&fq=docType_s:ART&fq=peerReviewing_s:1&classification_t:!NR&fq=language_s:en&rows=10000&wt=csv&fl=docid,publicationDateY_i,submittedDateY_i,submittedDateM_i,submittedDate_s,label_s,keyword_s,domain_s,title_s,anrProjectReference_s,authStructId_i,structAcronym_s) ». Les répertoires HAL des 3IA sont respectivement donnés par les codes « 3IA-COTEDAZUR » (exemple présenté), « ANITI », « MIAI » et « PRAIRIE-IA ». La taille du fichier à exporter, qui ne peut excéder 10 000 références par export (« rows=10000 »), est préalablement vérifiée en remplaçant le paramètre « wt=csv » par le paramètre « wt=xml ». Dans le cas où le nombre de références à exporter serait supérieur à 10 000, la taille des fichiers est ajustée à travers le paramètre « publicationDateY_i » et les fichiers individuels portant sur des fenêtres temporelles consécutives sont ensuite concaténés localement.

¹⁸⁶ <https://api.archives-ouvertes.fr/docs/search/?schema=fields>

¹⁸⁷ À travers les champs « submittedDateY_i » et « submittedDateM_i » de HAL.

¹⁸⁸ Par exemple, pour le 3IA Côte d'Azur, le principal partenaire universitaire est l'Université Côte d'Azur (UCA) depuis le 1^{er} janvier 2020 qui a été précédée de la COMUE UCA entre le 1^{er} mars 2015 et le 31 décembre 2019. L'ancienne université de Nice Sophia-Antipolis, qui a existé jusqu'au 31 décembre 2019, peut également être adjointe à ce périmètre pour couvrir la période sous-revue. Au total, l'archive HAL associée à cet ensemble universitaire est donnée par la requête : « [https://api.archives-ouvertes.fr/search/?q=structId_i:\(1039632+OR+523042+OR+117617\)&fq=publicationDateY_i:\[2014+TO+2021\]&fq=docType_s:ART&fq=peerReviewing_s:1&classification_t:!NR&fq=language_s:en&rows=10000&wt=csv&fl=docid,publicationDateY_i,submittedDateY_i,submittedDateM_i,submittedDate_s,label_s,keyword_s,domain_s,title_s,anrProjectReference_s,authStructId_i,structAcronym_s](https://api.archives-ouvertes.fr/search/?q=structId_i:(1039632+OR+523042+OR+117617)&fq=publicationDateY_i:[2014+TO+2021]&fq=docType_s:ART&fq=peerReviewing_s:1&classification_t:!NR&fq=language_s:en&rows=10000&wt=csv&fl=docid,publicationDateY_i,submittedDateY_i,submittedDateM_i,submittedDate_s,label_s,keyword_s,domain_s,title_s,anrProjectReference_s,authStructId_i,structAcronym_s) ».

¹⁸⁹ <https://aurehal.archives-ouvertes.fr/structure/index>.

contrôle dans la mesure du possible des biais associées aux regroupements universitaires lors de la période sous revue¹⁹⁰.

Dans le cadre de l'API proposée par HAL, la taille limite des exports dans le format csv apparaît limitée à 10 000 publications. Un export complet des articles internationaux a donc été réalisé à travers un ensemble de requêtes portant sur des entités et années civiles distinctes satisfaisant cette contrainte. Les fichiers ainsi obtenus ont ensuite été regroupés en un fichier unique regroupant l'ensemble des articles sur la période sous revue.

L'analyse sémantique des titres et mots clés des articles en lien avec le champ lexical de l'IA et/ou de ses applications a été réalisée sur la base des fichiers ainsi obtenus.

III. Exploitation des données de WoS pour suivre la production scientifique des sites et établissements associés aux 3IA

Résultats

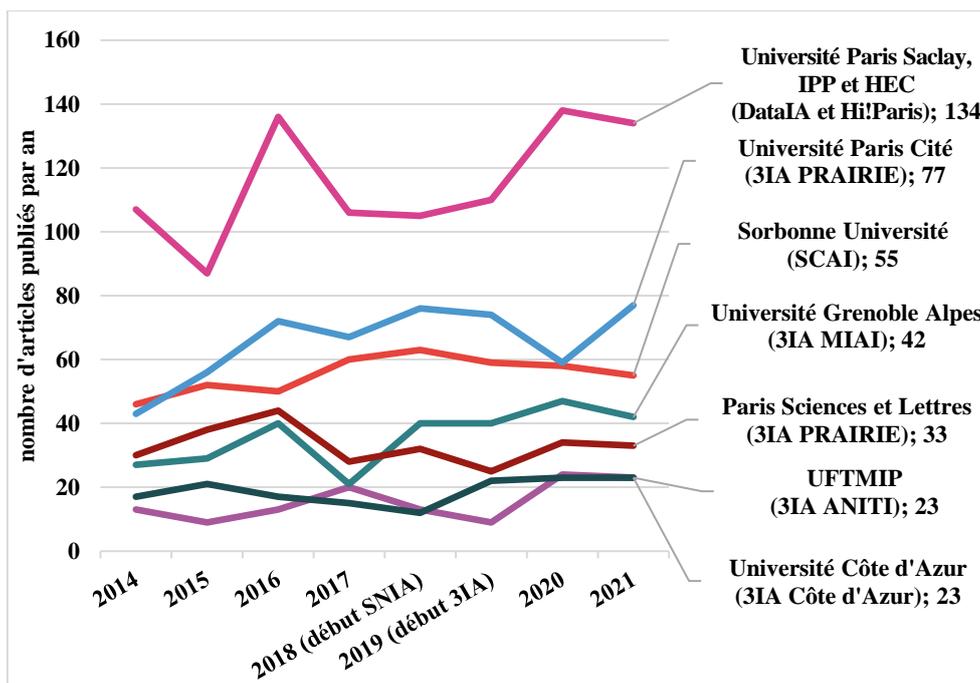
Sont ici rapportés les résultats issus de *Web of Science* (outil *InCites* ; schéma « *Web of Science* »¹⁹¹ dans la présente annexe et « *Citation topics* »¹⁹² sur l'archive électronique en ligne). Les paramétrages employés pour la production des indicateurs sont détaillés dans les sources des figures.

¹⁹⁰ Cf. note de bas de page n° 22 pour un exemple dans le cas de l'Université Côte d'Azur (UCA) dont la création date du 1^{er} janvier 2020.

¹⁹¹ Voir <https://incites.help.clarivate.com/Content/Research-Areas/wos-research-areas.htm> (catégorie « Computer science, artificial intelligence »).

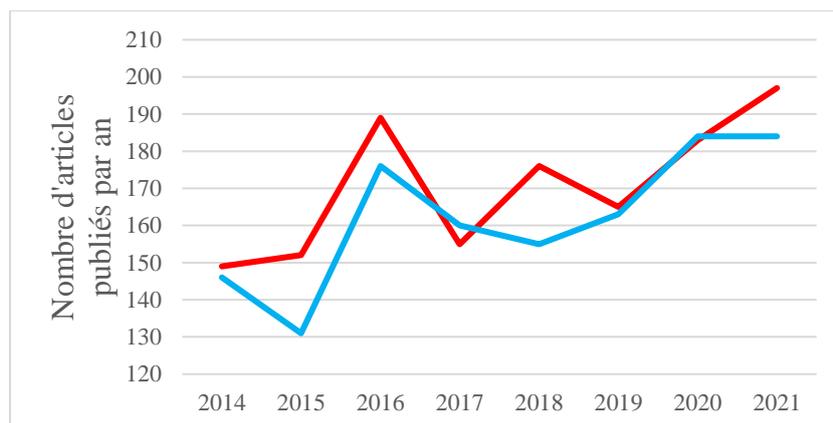
¹⁹² Voir : <https://incites.help.clarivate.com/Content/Research-Areas/citation-topics.htm> (catégorie « 4.61 Artificial Intelligence & Machine Learning » de l'échelle « Meso »).

Graphique n° 24 : production scientifique des principaux sites universitaires associés aux 3IA et aux centres d'excellence en IA (données issues de WoS à travers l'outil InCites ; schéma WoS)



Source : Cour des comptes d'après l'outil « InCites ». Le périmètre retenu correspond à l'ensemble des articles indexés par « Web of Science » au 30 septembre 2022 et appartenant au domaine de recherche « Computer Science, Artificial Intelligence » (articles en IA) du schéma « Web of Science » ainsi qu'à l'ensemble des articles (articles totaux). Les nombres d'articles calculés pour les 3IA correspondent aux totaux de leurs archives HAL respectives, sans application de filtre sémantique. Les données associées à cette figure et les indicateurs associés sont disponibles sur l'archive électronique en ligne. S'agissant des articles de la catégorie « 4.61 Artificial Intelligence & Machine Learning » de l'échelle « Meso » (articles en IA) du schéma « Citation topics », ils sont disponibles sur l'archive électronique en ligne

Graphique n° 25 : production scientifique comparée des principaux sites universitaires associés aux 3IA (en rouge) et aux centres d'excellence (en bleu)

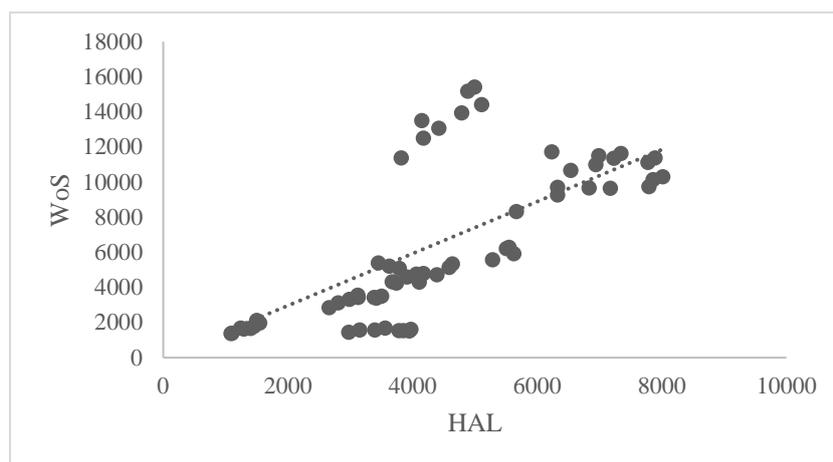


Source : Cour des comptes d'après l'outil « InCites ». Le périmètre retenu correspond à l'ensemble des articles indexés par Web of Science au 30 septembre 2022 et appartenant au domaine de recherche « Computer Science, Artificial Intelligence » (articles en IA) du schéma « Web of Science ». S'agissant des articles de la catégorie « 4.61 Artificial Intelligence & Machine Learning » de l'échelle « Meso » (articles en IA) du schéma « Citation topics », ils sont disponibles sur l'archive électronique en ligne

IV. Comparaison des données issues de HAL et WoS

Au global, la couverture de HAL, tous articles confondus, apparaît réduite par rapport à WoS mais demeure néanmoins très satisfaisante à l'échelle de l'ensemble des sites universitaires examinés (voir figure ci-dessous).

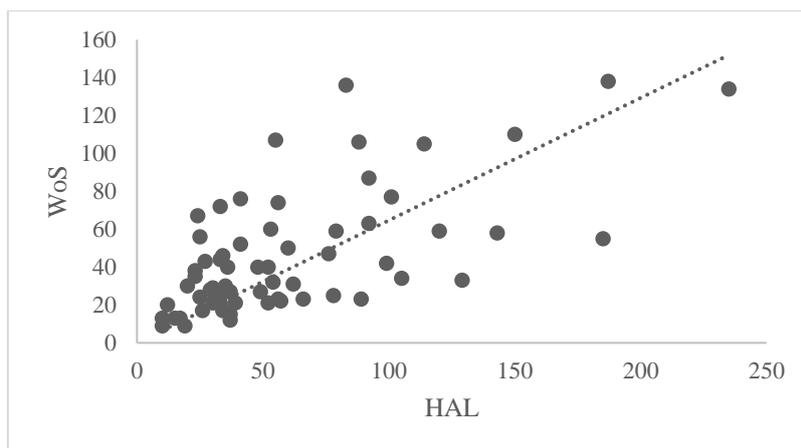
Graphique n° 26 : comparaison entre WoS et HAL du nombre d'articles publiés pour les sites universitaires examinés



Source : Cour des comptes. Huit sites universitaires sont comparés pour huit années, soit 64 points de comparaison au total. L'interpolation linéaire est donnée par la relation $[WoS = 1,5 HAL]$ ($R^2 = 0,51$). Le coefficient de corrélation de Pearson est de + 0,72 avec une p-valeur négligeable. Les écarts moyens constatés entre les deux jeux de données sont stables sur la période 2014 à 2021. La couverture de HAL n'est toutefois pas parfaitement homogène suivant les sites, ce qui explique les « nuages » observés

S'agissant plus spécifiquement des articles en IA, les deux sources conduisent à des résultats corrélés positivement de façon significative sur le plan statistique (voir figures et analyses ci-dessous). Si les écarts moyens constatés entre les deux jeux de données sont stables sur la période 2014 à 2018, ils tendent à s'accroître sur la période 2019 à 2021, en particulier pour le schéma « *Web of Science* » (voir tableau ci-dessous). Comme ce phénomène n'est pas observé pour les articles tous domaines confondus, l'effet d'accroissement des écarts observé traduit probablement des différences de méthodes et de définitions pour classer les articles comme étant dans le domaine de l'IA.

Graphique n° 27 : comparaison entre WoS et HAL du nombre d'articles publiés en IA sur la période 2014 à 2021 pour les sites universitaires examinés



Source : Cour des comptes. Huit sites universitaires sont comparés pour huit années, soit 64 points de comparaison au total. Par souci de concision, seuls les résultats issus du schéma Web of Science sont ici représentés (se reporter à l'archive électronique en ligne pour les résultats détaillés issus du schéma Citation topics). Les interpolations linéaires sont données par la relation $[WoS = 0,6 HAL]$ (avec $R^2 = 0,3$) pour le schéma Web of science ainsi que par la relation $[WoS = 0,2 HAL]$ (avec $R^2 = 0,6$) pour le schéma Citation topics. Les coefficients de corrélation de Pearson associés sont de $+ 0,65$ (schéma WoS) et de $+ 0,75$ (schéma Citation topics) avec des p-values négligeables

Tableau n° 30 : analyse des écarts observés entre HAL et WoS (schémas WoS et Citation topics) sur la période 2014 à 2021

Année	Articles en IA schéma WoS	Articles en IA schéma Citation topics	Articles tous domaines confondus
2014	39 %	107 %	38 %
2015	33 %	97 %	38 %
2016	35 %	130 %	39 %
2017	43 %	114 %	42 %
2018	38 %	139 %	36 %
2019	59 %	150 %	37 %
2020	60 %	148 %	42 %
2021	82 %	148 %	47 %
Moyenne générale	49 %	129 %	40 %

Source : Cour des comptes

Annexe n° 7 : exploitation de différentes sources de données sur les thèses pour suivre la dynamique et la structuration des forces scientifiques françaises en IA

I. Préambule

Les formations doctorales en IA constituent un double marqueur :

- un marqueur de la recherche en IA puisque les équipes de recherche s'appuient sur les travaux de leurs doctorants pour développer leurs projets scientifiques ;
- un marqueur de la formation des futurs talents de l'IA, docteurs en ce domaine, qui ont vocation à faire carrière tant dans le monde académique que dans les autres domaines de l'économie et de la société, notamment en lien avec les enjeux de la transition numérique.

La SNRIA ayant été mise en place il y a quatre ans (2018), les effets causaux de celle-ci sur la production scientifique peuvent être difficiles à identifier de façon certaine. Les formations doctorales constituent un signal facilement mesurable *via* les financements des projets de thèses et les inscriptions au diplôme de doctorat que chaque étudiant réalise chaque année (suivi des moyens y compris sur les temps courts), ainsi qu'à travers les publications des thèses qui permettent de suivre conjointement les travaux en IA et les effectifs de docteurs formés dans ce domaine (suivi des résultats en matière de recherche et de formation sur des temps plus longs).

À ce titre, le suivi des thèses en IA a fait l'objet d'une attention particulière de la part de la Cour, qui a pu collecter trois sources d'informations complémentaires :

- l'exploitation de la base de données publique « thèses.fr » maintenue par l'agence bibliographique de l'enseignement supérieur¹⁹³ entre 1989 et 2019, qui permet utilement de suivre les évolutions des thèses en IA sur des séries longues à partir de données presque exhaustives (voir II ci-après) ;
- les financements des CIFRE auprès de l'ANRT qui catégorise elle-même les thèses en IA dans son système d'information depuis 2014. Cette base distingue les 50 CIFRE financés dans le cadre de la SNRIA

¹⁹³ L'agence bibliographique de l'enseignement supérieur (ABES) constitue un établissement public administratif placé sous la tutelle du ministère chargé de l'enseignement supérieur et de la recherche.

et a également l'avantage de permettre d'analyser les travaux en IA sur une période plus récente que les thèses publiées (voir III ci-dessous) ;

- le recensement des doctorants par deux organismes nationaux de recherche : CNRS et Inria ainsi que par la collecte auprès des Instituts 3IA et centres d'excellence.

Les sources peuvent également être croisées pour analyser la logique d'octroi des CIFRE en IA et la structuration de la communauté scientifique dans ce domaine (voir IV ci-dessous).

II. Exploitation des données ouvertes de thèses.fr pour détecter les travaux scientifiques dans le domaine de l'IA sur des séries longues

Sources mobilisées et périmètre des analyses

Afin de rendre compte des évolutions des travaux en IA sur des séries longues, la base de données ouverte « theses.fr », a été mobilisée par la Cour. Ces données peuvent être considérées comme exhaustives pour les thèses soutenues entre 1989 et 2019¹⁹⁴.

Les thèses dans le domaine de l'IA ont été détectées par la méthode sémantique décrite en annexe n° 6 appliquée aux titres, aux disciplines et aux mots clés associés aux thèses.

Les établissements de soutenance ont été reconstruits en croisant la base theses.fr avec la liste des principaux établissements du supérieur maintenue à jour par le MESR¹⁹⁵. Le croisement a été opéré à partir de l'identifiant « idref » ou, à défaut, à partir des dénominations actuelles des

¹⁹⁴ Pour la dernière décennie, la Cour a comparé les données de theses.fr compilées par année avec les statistiques du MESR (https://publication.enseignementsup-recherche.gouv.fr/eestr/FR/T744/le_doctorat_et_les_docteurs/ ; voir données relatives à l'évolution du nombre de doctorats délivrés sur la période 2009 à 2020) afin de retenir les années où le taux de couverture est élevé et homogène dans le temps. Ce résultat est cohérent avec le guide de réutilisation des données de theses.fr mis à jour en mai 2020 (ABES) qui indique que « les données des deux dernières années sont non consolidées et doivent être considérées comme non exhaustives. Les délais de traitements des thèses, propres à chaque établissement, influent sur la complétude de ces données. Pour l'année 2021, le délai de traitement moyen entre la date de soutenance de la thèse et sa publication effective dans theses.fr était de 250 jours ». Pour les données plus anciennes, les années antérieures à 1989 ont été écartées car elles rapportaient des effectifs bien en deçà des effectifs rapportés pour les années 1989 et suivantes.

¹⁹⁵ Cf. <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-principaux-etablissements-enseignement-superieur/>

établissements¹⁹⁶. Cette approche permet de corriger les biais liés aux regroupements d'établissements intervenus sur la période. En outre, les données géographiques relatives aux établissements sont incluses dans la liste du MESR, ce qui permet la réalisation d'analyses spatialisées.

Principaux résultats

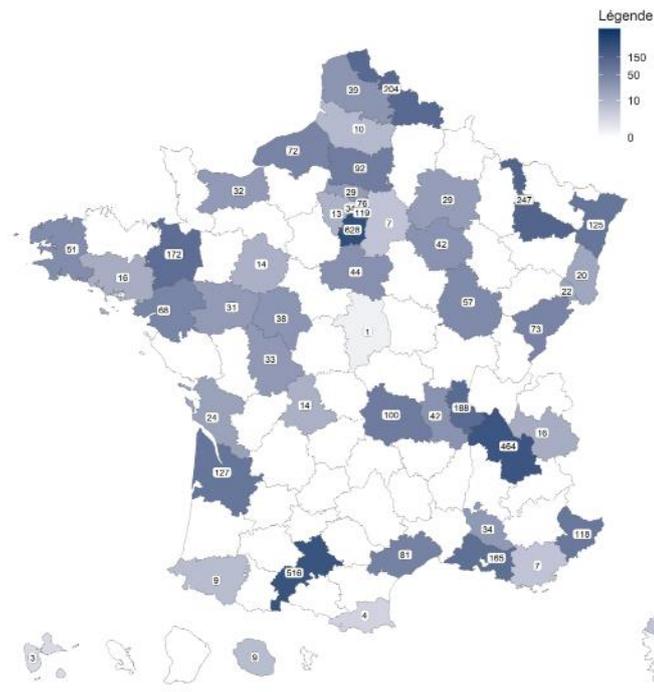
Tableau n° 31 : des thèses en IA marquées par un fort accroissement tant en absolu qu'en relatif

<i>Année de soutenance</i>	Nombre de thèses en IA (T_IA)	Nombre de thèses (T)	Part des thèses en IA (R = T_IA / T) en %
1989	78	10 835	0,7
2019	556	12 829	4,3
<i>Moyenne sur la période 1989 à 2019</i>	184	11 940	1,5
<i>Total sur la période 1989 à 2019</i>	5705	37 0135	1,5
<i>Évolution de 1989 à 2019 (en % et en points)</i>	613 %	18 %	+ 3,6 points

Source : Cour des comptes à partir des données de theses.fr. Les données annuelles détaillées sur la période 1989 à 2019 sont proposées au format électronique sur le site web de la Cour. Les thèses en IA sont principalement réalisées en sciences « dures », tout particulièrement en informatique (voir l'archive électronique pour les résultats détaillés)

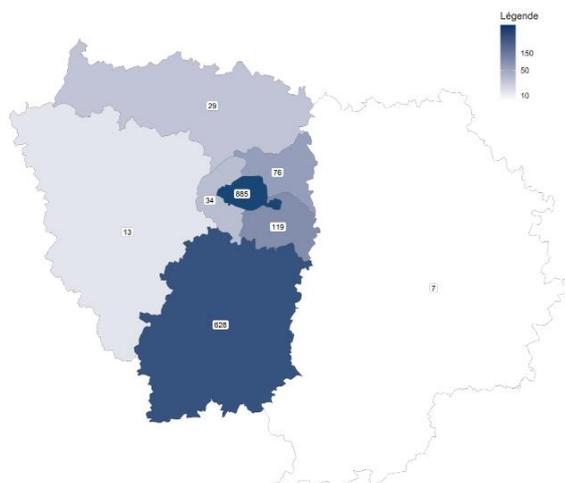
¹⁹⁶ Une correspondance a été établie par la Cour entre le champ « établissements_soutenance.0.nom » de theses.fr et le champ « Libellé » des principaux établissements du MESR en rattachant les anciens établissements aux établissements aujourd'hui regroupés.

Carte n° 5 : répartition départementale des thèses en IA soutenues entre 1989 et 2019 dans les établissements français d'enseignement supérieur



Source : Cour des comptes à partir des données ouvertes de theses.fr (ABES). Les thèses sont données par année de soutenance. La détection des thèses en IA est réalisée suivant la méthode sémantique développée par la Cour

Carte n° 6 : répartition départementale des thèses en IA soutenues entre 1989 et 2019 dans les établissements franciliens d'enseignement supérieur



Source : Cour des comptes à partir des données ouvertes de theses.fr (ABES). Les thèses sont données par année de soutenance. La détection des thèses en IA est réalisée suivant la méthode sémantique développée par la Cour

I. Exploitation des données relatives aux thèses CIFRE pour suivre les tendances récentes relatives aux travaux de recherche partenariale en IA

Sources mobilisées et périmètre des analyses

L'Association nationale recherche technologie (ANRT) opère le dispositif CIFRE pour le compte du ministère de l'enseignement supérieur et de la recherche. Les conventions industrielles de formation par la recherche (CIFRE) regroupe une entreprise, un laboratoire et un doctorant autour d'un projet doctoral depuis 1981. L'entreprise reçoit une subvention sur trois ans de 14 000€ en contrepartie de l'emploi du doctorant.

L'ANRT a fourni à la Cour une base de données comportant l'ensemble des dossiers retenues dans le cadre d'une CIFRE depuis 2014. Cette base identifie le doctorant et sa civilité, le laboratoire, l'entreprise, le sujet de thèse ainsi que la discipline de la thèse. Dans le cadre des 50 bourses CIFRE financé par le plan IA, l'aire géographique et la date de naissance ont été fournies.

Les 50 CIFRE IA ont été identifiées sur la cohorte de CIFRE déposées en 2019 et acceptées à partir d'une liste de mots clés se rapportant à l'IA élaborée conjointement par l'ANRT, le MESR et l'ANR.

C'est l'ANRT lors de l'analyse des dossiers qui a déterminé si le projet de thèse est dans le domaine de l'Intelligence artificielle sur la base de mots clés.

Principaux résultats

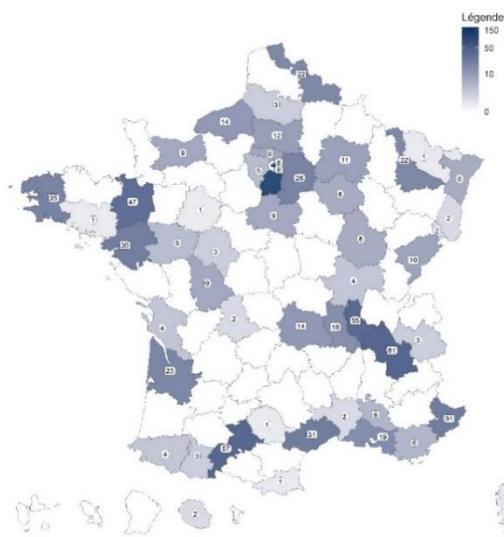
Entre 2014 et 2021, le nombre de thèse CIFRE en IA a cru de 130 % passant de 97 thèses à 233 avec un pic en 2019 à hauteur de 283 thèses, alimenté par le financement spécifique de la SNRIA (cf. tableau n° 32). Dans le même temps, les thèses CIFRE en général n'ont augmenté que de 19 %. Aujourd'hui, les thèses IA représentent environ 15 % des thèses CIFRE.

Tableau n° 32 : évolution de la part des thèses CIFRE en IA débutées sur la période 2014-2021

<i>Année de début de thèse</i>	2014	2015	2016	2017	2018	2019	2020	2021	Total général	Moyenne	Évolution 2014 à 2021
<i>Nombre de thèses CIFRE en IA (estimation ANRT)</i>	97	123	161	185	250	283	235	223	1 557	188	130 %
<i>Nombre total de thèses CIFRE</i>	1 364	1 364	1 449	1 583	1 682	1 696	1 511	1 621	12 270	1 534	19 %
<i>Part des thèses CIFRE en IA</i>	7 %	9 %	11 %	12 %	15 %	17 %	16 %	14 %	13 %	12 %	7 %

Source : Cour des comptes d'après les données ANRT. L'année de référence est l'année de dépôt du dossier de thèse ; celle-ci peut débiter l'année suivante. La détection des thèses dans le domaine de l'IA a dans le cas présent été réalisée par l'ANRT. Il est à noter que les 50 thèses CIFRE financées par la SNRIA ont été déposés en 2019

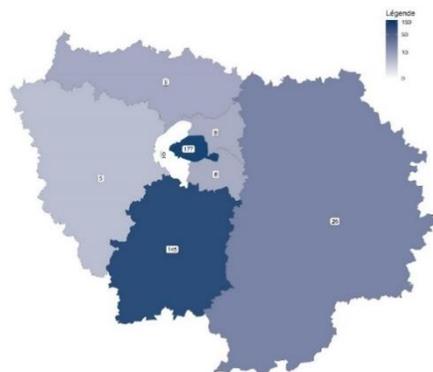
Carte n° 7 : répartition départementale des thèses CIFRE en IA entre 2018 et 2021 dans les établissements français d'enseignement supérieur



Source : Données ANRT, représentation Cour des comptes

La représentation spatiale permet de montrer la congruence entre les sites d'excellences, en particulier les instituts 3IA.

Carte n° 8 : répartition départementale en Île-de-France des thèses CIFRE en IA entre 2018 et 2021 dans les établissements français d'enseignement supérieur



Source : Données ANRT, représentation Cour des comptes

La représentation des femmes parmi les bourses CIFRE est assez stable depuis 2014 oscillant entre 36,07 et 39,67 % en général. Il existe néanmoins une progression pour les CIFRE IA de 16,50 % en 2014 à 28,94 % en 2020 mais ce taux redescend à 20,63 % en 2021. Il est à noter qu'il y a 15 bourses CIFRE IA financées par la SNRIA détenues par des étudiants d'origine étrangère et 12 par des femmes. Concernant les disciplines (voir l'archive électronique en ligne), il n'existe pas de différences significatives entre les 50 CIFRE IA de la SNRIA et celles financées sur la période. Les quatre disciplines majeures sont Informatique et applications (37,24 %) suivie de Mathématiques et leurs interactions (13,42 %), Sciences de l'information et de la communication (11,81 %) et Traitement du signal et des images (10,09 %).

Il est à remarquer que 41,68 % des thèses CIFRE en IA sont réalisées dans des établissements partenaires d'un institut d'excellence depuis 2018 (cf. tableau n° 33).

Tableau n° 33 : nombre de thèses CIFRE IA dans la sphère géographique des instituts d'excellence entre 2018 et 2021

<i>Centres d'excellence</i>	Nombre	%
<i>3IA COTE D'AZUR</i>	31	3,13
<i>ANITI</i>	55	5,55
<i>MIAI</i>	60	6,05
<i>PRAIRIE</i>	56	5,65
<i>DATAIA</i>	95	9,59
<i>HI ! PARIS</i>	59	5,95
<i>SCAI</i>	57	5,75
<i>Total (sur 991 CIFRE IA)</i>	441	41,68

Source : Données ANRT, retraitement Cour des comptes. À noter qu'il n'est pas possible d'identifier spécifiquement les doctorants appartenant aux instituts d'excellence. Ce tableau regroupe l'ensemble des doctorants inscrits dans un établissement partenaire de l'institut d'excellence

Pour aller au-delà des analyses descriptives, la Cour a mené une analyse économétrique pour répondre à trois questions :

- la Stratégie a-t-elle eu un effet sur le nombre de thèse CIFRE en IA ?
- les centres d'excellence ont-ils eu un effet additionnel ?
- la stratégie a-t-elle permis une meilleure diversité de genre ?

Il est mené une analyse économétrique pour expliquer le nombre de thèses CIFRE en IA en introduisant des effets fixes pour chaque institution et en analysant l'interaction avant/après la mise en œuvre de la stratégie. N'étant pas dans un contexte d'expérience quasi naturelle, la stratégie d'identification présente des biais reconnus par la Cour. Néanmoins, ce traitement prend en compte l'hétérogénéité entre institutions (tous les laboratoires n'ont pas la même capacité à accueillir des thèses CIFRE et plus particulièrement en IA), et la période pré et post stratégie (2014-2017 et 2018-2021).

Tableau n° 34 : analyse du nombre de thèses CIFRE IA

<i>Variables</i>	Nb CIFRE IA (1)	Nb CIFRE IA (2)
<i>Nb de CIFRE</i>	- 0.09***	- 0.00
	(0.03)	(0.03)
<i>NB CIFRE x post</i>	0.02***	0.03***
	(0.01)	(0.01)
<i>2015</i>	0.20	0.20*
	(0.13)	(0.11)
<i>2016</i>	0.47***	0.46***
	(0.17)	(0.14)
<i>2017</i>	0.67***	0.57***
	(0.16)	(0.14)
<i>2018</i>	0.82***	0.35**
	(0.19)	(0.15)
<i>2019</i>	1.10***	0.61***
	(0.24)	(0.18)
<i>2020</i>	0.64***	0.25*
	(0.16)	(0.15)
<i>2021</i>	0.61***	0.22
	(0.19)	(0.15)
<i>Côte d'Azur</i>		1.67***
		(0.60)
<i>ANITI</i>		3.77***
		(1.21)
<i>DATAIA</i>		1.94***
		(0.55)

Variables	Nb CIFRE IA (1)	Nb CIFRE IA (2)
<i>Hi!PARIS</i>		2.05***
		(0.56)
<i>MIAI</i>		2.07*
		(1.16)
<i>PRAIRIE</i>		1.98***
		(0.60)
<i>SCAI</i>		3.68**
		(1.59)
<i>Côte d'Azur x post</i>		0.56
		(0.57)
<i>ANITI x post</i>		- 0.75
		(0.61)
<i>DATAIA x post</i>		1.45*
		(0.78)
<i>Hi! PARIS x post</i>		0.89
		(0.54)
<i>MIAI x post</i>		1.80**
		(0.84)
<i>PRAIRIE x post</i>		0.89
		(0.73)
<i>SCAI x post</i>		2.49*
		(1.44)
<i>Constante</i>	1.61***	0.38
	(0.31)	(0.32)
<i>Observations</i>	1,134	1,134
<i>Nombre d'institutions</i>	143	143
<i>R2</i>	0.387	0.540
<i>F</i>	10.11	10.95
<i>P</i>	0	0

Source : Données ANRT, Traitement Cour des comptes. Le modèle économétrique explique le nombre de thèse CIFRE en IA le nombre de thèse CIFRE, les effets fixes années et institutions ainsi que les effets instituts au regard de la stratégie. Le modèle économétrique est un modèle à effet fixe institutions et année pour prendre en compte l'hétérogénéité temporelle et spatiale. La corrélation entre données est prise en compte pour le calcul des écart-types (technique des clusters). La significativité des résultats (effet de la variable significativement différent de zéro) est indiquée par les étoiles : * <5 %, **<1 % et ***<0,1 %

L'analyse montre un effet annuel persistant conforme à l'augmentation du nombre de thèse CIFRE en IA. L'analyse 1 montre qu'il y a un effet de seuil concernant le nombre de thèse CIFRE en IA qui a été diminué par la mise en place de la stratégie. En introduisant le centre d'excellence, cet effet de seuil disparaît et montre un effet positif de la stratégie. L'ensemble des centres d'excellence sont attractifs pour des thèses CIFRE en IA sans que la stratégie modifie cette attractivité de manière significative à l'exception de MIAI.

Concernant la diversité de genre, le nombre de thèses CIFRE en IA ne dépend pas de la féminisation d'un organisme sauf dans certains sites d'excellence comme DATAIA, PRAIRIE et Hi! PARIS post stratégie.

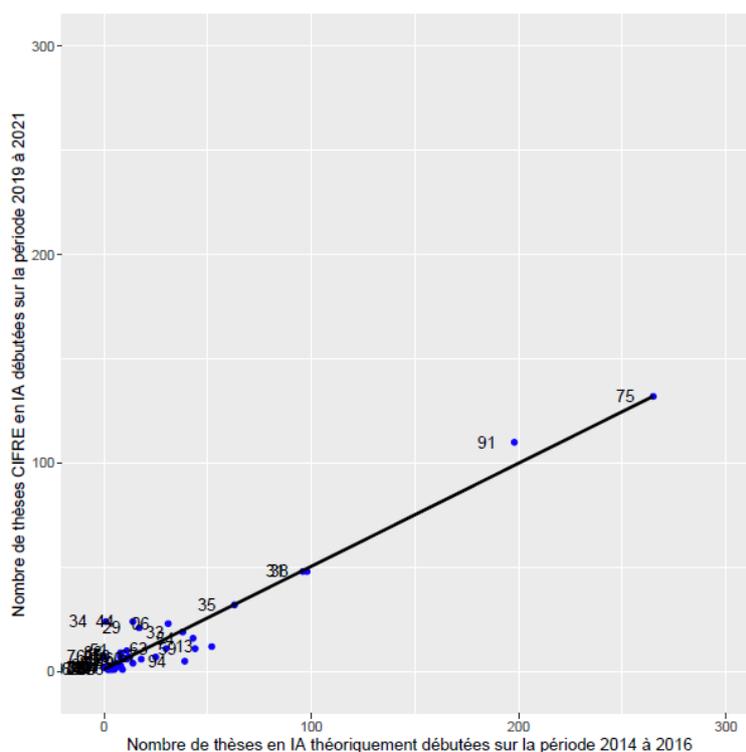
Tableau n° 35 : nombre de femmes en thèses CIFRE IA

VARIABLES	Nb CIFRE IA (1)	Nb CIFRE IA (2)
<i>Femme</i>	- 0.06***	- 0.06***
	(0.02)	(0.02)
<i>Nb de CIFRE</i>	0.07**	0.07**
	(0.03)	(0.03)
<i>NB CIFRE x post</i>	0.01**	0.01*
	(0.00)	(0.00)
<i>2015</i>	0.09*	0.10*
	(0.05)	(0.06)
<i>2016</i>	0.09*	0.10**
	(0.04)	(0.05)
<i>2017</i>	0.14**	0.17**
	(0.06)	(0.07)
<i>2018</i>	0.16***	0.20***
	(0.06)	(0.07)
<i>2019</i>	0.18***	0.23***
	(0.06)	(0.07)
<i>2020</i>	0.21***	0.25***
	(0.07)	(0.08)
<i>2021</i>	0.08	0.11
	(0.07)	(0.08)
<i>Constante</i>	0.29***	0.36***
	(0.05)	(0.06)
<i>Observations</i>	1,414	1,134
<i>R-squared</i>	0.32	0.29
<i>N_Group</i>	255	143
<i>F</i>	4.145	4.414
<i>p</i>	2.55e-05	2.09e-05

Source : Données ANRT, Traitement Cour des comptes. Le modèle économétrique explique le nombre de thèse CIFRE en IA le nombre de thèse CIFRE, le genre, les effets fixes années et institutions ainsi que les effets instituts au regard de la stratégie. Le modèle économétrique est un modèle à effet fixe institutions et année pour prendre en compte l'hétérogénéité temporelle et spatiale. La corrélation entre données est prise en compte pour le calcul des écart-types (technique des clusters). La significativité des résultats (effet de la variable significativement différent de zéro) est indiquée par les étoiles : * <5 %, **<1 % et ***<0,1 %

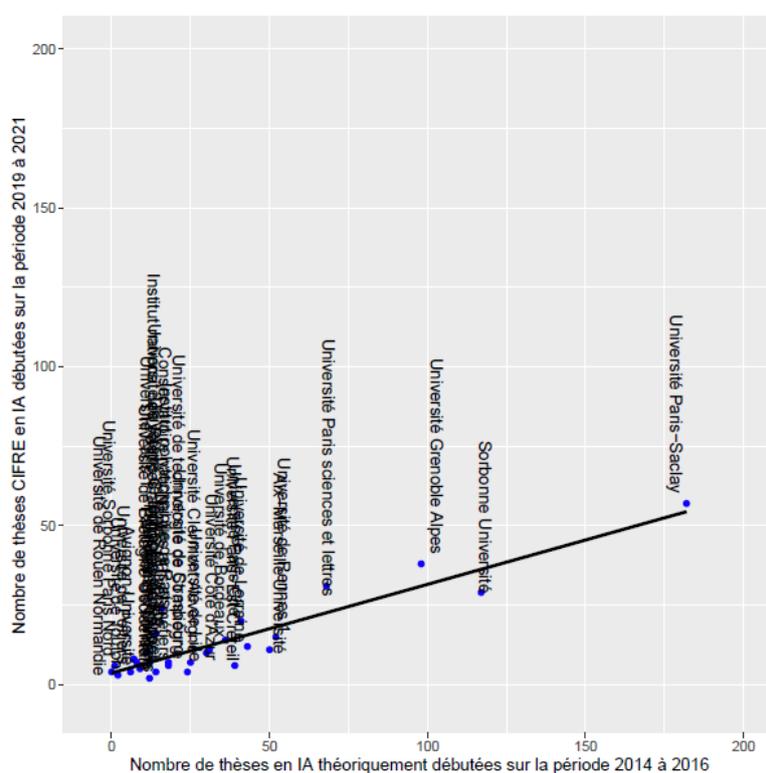
II. Analyse comparée des deux types de données pour analyser la logique d'octroi des CIFRE en IA et la structuration de la communauté scientifique dans ce domaine

Graphique n° 28 : les thèses CIFRE en IA octroyées depuis 2019 ont été octroyées aux zones géographiques déjà actives dans ce domaine, de façon proportionnelle à leur productivité en matière de thèses soutenues en IA telle qu'observée sur des bases historiques



Source : Cour des comptes à partir des données ouvertes de theses.fr (ABES) ainsi que des données de l'ANRT relatives aux thèses CIFRE. Chaque point représente un département français, ici indiqué par son numéro. Au grain des départements, le nombre de thèses CIFRE en IA sur la période 2019 à 2021 et le nombre de thèses en IA sur la période 2014 à 2016 sont associées linéairement (corrélation de Pearson est de + 0,97 avec une p-valeur négligeable)

Graphique n° 29 : les thèses CIFRE en IA octroyées depuis 2019 ont été octroyées aux établissements d'enseignement déjà actifs dans ce domaine, de façon proportionnelle à leur productivité en matière de thèses soutenues en IA telle qu'observée sur des bases historiques



Source : Cour des comptes à partir des données ouvertes de *theses.fr* (ABES) ainsi que des données de l'ANRT relatives aux thèses CIFRE. Chaque point représente un établissement d'enseignement supérieur français. Au grain des établissements, le nombre de thèses CIFRE en IA sur la période 2019 à 2021 et le nombre de thèses en IA sur la période 2014 à 2016 sont associées linéairement (corrélation de Pearson est de + 0,90 avec une p-valeur négligeable)

Annexe n° 8 : analyses de réseaux pour caractériser et mesurer la structuration d'écosystèmes dans le domaine de l'IA

I. Préambule

L'emploi de réseaux permet de cartographier les écosystèmes de l'IA sous différents angles tels que les collaborations scientifiques ou les relations d'influence sur le web et de donner des mesures de l'importance des acteurs et groupes qui les constituent sur des fondements aussi objectifs que possible.

Après une présentation des indicateurs employés pour mesurer la centralité d'acteurs ou de catégories d'acteurs au sein d'écosystèmes représentés sous la forme de réseaux (II), les modalités détaillées de construction de ces réseaux sont décrites tant pour le réseau de collaboration scientifique des instituts 3IA (III) que pour les réseaux des relations institutionnels sur le web dans le domaine de l'IA, tant en France qu'en Allemagne (IV).

II. Indicateurs pour mesurer la centralité des acteurs, catégories d'acteurs et communautés des réseaux

Au sein des réseaux analysés (cf. III et IV ci-après pour les réseaux scientifiques et institutionnels respectivement), certains acteurs ou groupes d'acteurs peuvent apparaître plus centraux ou importants que d'autres. Des indicateurs de centralité peuvent être employés pour confirmer ces intuitions et comparer les différentes catégories d'acteurs tels que les instituts 3IA, les centres d'excellence français en IA en dehors des 3IA ou encore les « homologues » allemands des instituts 3IA.

À l'échelle des acteurs représentés par les nœuds d'un réseau donné, deux grands types d'indicateurs de centralité sont ici employés :

- un indicateur « local » de centralité calculé à travers le nombre de liens d'un acteur représentant un nœud réseau. Dans le cas d'un réseau non-dirigé tel que les collaborations scientifiques, le nombre total de liens non-dirigés est employé, en tenant compte de la pondération des liens (un article scientifique compte au global toujours pour une unité quel que soit le nombre de co-auteurs associés de manière à éviter les biais). Dans le cas d'un réseau dirigé tel que les citations de sites web institutionnels, le nombre total de liens entrants est employé afin de refléter l'influence de l'acteur au sein de l'écosystème. Ces deux types d'indicateurs peuvent être normalisés par le nombre total de liens du

réseau afin de mesurer l'influence relative d'un acteur au sein d'un écosystème donné. Les mesures absolues sont également utiles lorsqu'il s'agit de comparer la performance d'acteurs analogues de deux écosystèmes à l'instar des instituts 3IA français et de leurs homologues allemands ;

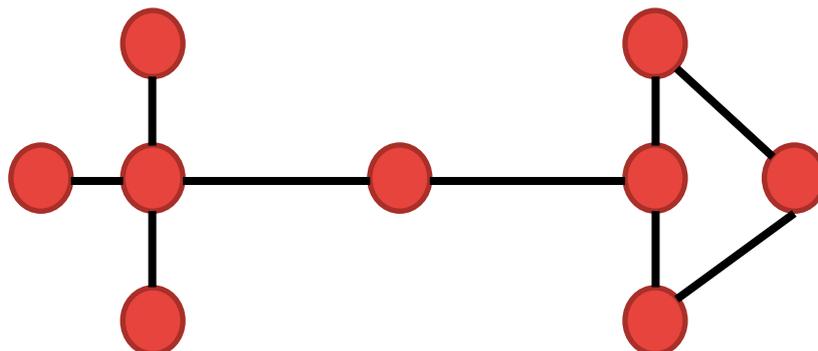
- un indicateur « global » de centralité mesuré par la perte relative de la taille de la plus grande composante connexe du réseau considéré lorsque l'acteur fait l'objet d'une procédure d'ablation. Il s'agit d'une mesure dérivée de la théorie physique de la percolation¹⁹⁷ qui reflète l'importance systématique d'un acteur au sein d'un réseau, par exemple dans le cas de la prévention des épidémies¹⁹⁸. Dans le cas des réseaux dirigés, c'est la plus grande composante fortement connexe qui est employée. Intuitivement, cette mesure reflète le fait que « supprimer » un nœud stratégique pour un réseau va conduire à un effondrement de sa structure d'ensemble.

Les deux indicateurs sont complémentaires dans la mesure où les deux types de mesures ne conduisent pas nécessairement aux mêmes résultats. Par exemple, un acteur peut être central au sens du critère de centralité « globale » mais pas « locale ». Dans le schéma ci-dessous, si l'acteur A est le plus central au sens « local », l'acteur B est plus important au sens « global » dans la mesure où il vient connecter des communautés du réseau qui ne serait pas « atteignables » par un autre « passage ».

¹⁹⁷ Albert R, Jeong H, Barabasi AL. 2000 Error and attack tolerance of complex networks. doi:10.1038/35019019.

¹⁹⁸ Meyers LA, Newman MEJ, Pourbohloul B. 2006 Predicting epidemics on directed contact networks. *J. Theor. Biol.* doi:10.1016/j.jtbi.2005.10.004.

Schéma n° 5 : illustration des grandes catégories de mesures de centralité employées pour mesurer l'importance des acteurs et groupes d'acteurs dans les réseaux analysés par la Cour



Source : Cour des comptes. Dans cet exemple schématique, l'acteur A est le plus central au sens « local » (quatre liens, contre liens pour l'acteur C et deux pour l'acteur B). Toutefois, l'acteur B est le plus central au sens « global » car sa disparition entrainerait la déstabilisation de l'ensemble de la structure du réseau, en supprimant les « passages » existants entre les nœuds A et C et donc aux autres acteurs associés

Ces mesures sont également employées de façon plus large à l'échelle d'un ensemble de nœuds d'un réseau pour mesurer l'importance de groupes d'acteurs ou de communautés. Dans ce cas de figure, quand des comparaisons sont opérées, les indicateurs peuvent être normalisés de façon à prendre en compte le fait que le nombre de nœuds dans les différents groupes comparés ne sont pas nécessairement identiques.

III. Réseau des collaborations scientifiques des instituts 3IA à partir des archives HAL

Les archives HAL précisent les affiliations institutionnelles des co-auteurs des articles scientifiques. Il est donc possible de construire le réseau des collaborations scientifiques à l'œuvre entre institutions. La Cour a procédé à cette opération sur la base du regroupement des archives HAL des quatre instituts 3IA en considérant chaque article scientifique publié exactement une fois. Afin d'éviter les biais, il a été considéré qu'un article vaut systématiquement une unité quel que soit le nombre d'institutions co-auteurs de cet article. Cette démarche revient à pondérer les liens et nœuds du réseau de manière à assurer la « conservation » du nombre total d'articles publiés par les instituts 3IA.

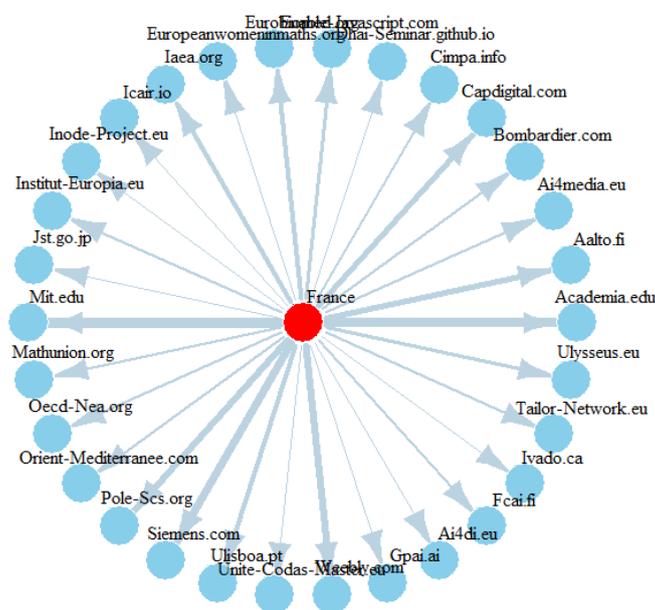
IV. Réseaux des citations institutionnels, notamment des instituts 3IA, à partir des données du web

a. Construction du réseau

En partant d'une liste de site Web de référence (cf. liste référence), un *scraping* a été réalisé d'une façon automatisée *via* un script R afin de récupérer tous les sites Web découverts en suivant les liens externes. Les liens pointant vers les réseaux sociaux : Twitter, LinkedIn, Facebook, etc. sont exclus de cette analyse car leur apparition est due à leur présence sur le site de l'entité qu'ils représentent et non pas à une interconnexion entre deux entités différentes.

Les résultats de ce *scraping* permettent, d'une part, d'identifier des nouveaux acteurs de formation, de recherche et de valorisation participant à la structuration de la communauté nationale IA, et, d'autre part, d'identifier les liens internationaux. Les liens internationaux repérés sont représentés dans le graphe ci-dessous.

Graphique n° 30 : représentation des liens internationaux sortants des sites de référence français



Source : Cour des comptes

Les sites internet nationaux sont classifiés selon les catégories d'acteur suivantes :

- administrations nationales (ministères, secrétariat général pour l'investissement et autres services du Premier ministre, etc.) ;
- autres opérateurs de l'État et structures publics, en dehors des établissements d'enseignement et de recherche : IGN, CDC, BDT, Bpifrance, etc. ;
- autorités indépendantes et représentations nationales (Senat, CNIL, etc.) ;
- collectivités locales ;
- universités françaises ;
- grandes écoles françaises ;
- structures et organismes de recherche (CNRS, Inria, CEA, Inserm, INRAE, Institut Pasteur, etc.) : secteur public ;
- structures partenariales : statut mixte entre partenaires privés et publics ;
- entreprises : secteur privé ;
- associations professionnelles dans le domaine de l'IA (ELLIS ; CLAIR ; AFIA ; SSFAM ; etc.) ;
- associations professionnelles en dehors de l'IA ;
- Instituts 3IA : MIAI, ANITI, PRAIRIE, 3IACôte d'Azur ;
- sites d'excellence : DATAIA, SCAI, HI-Paris ;
- média : Actu IA, etc. ;
- chaires IA.

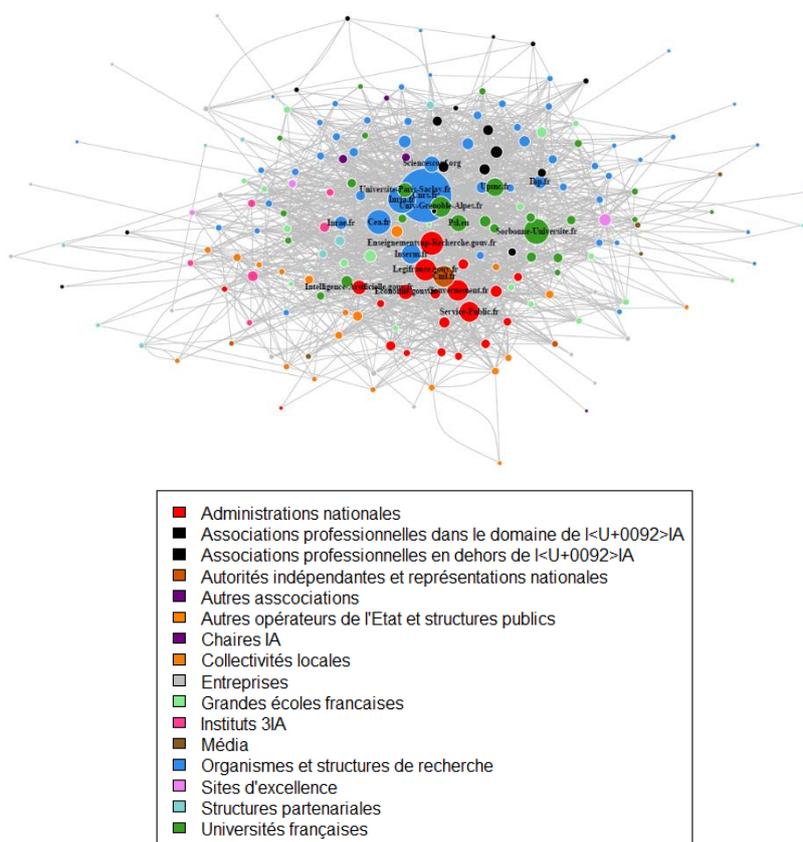
Dans le but d'avoir l'ensemble des liens mutuels entre ces sites, un *Web crawling* a été réalisé par l'aide de l'outil *Hyphe* du Media Lab-Sciences Po Paris : un outil de fouille systématique des pages internet. Le lien vers l'interface Hyphe dédié à cet usage est le suivant : <https://hyphe.medialab.sciences-po.fr/cour-des-comptes/>.

Les pages appartenant au même site Internet sont considérées comme une seule entité et sont représentées par un seul nœud. Le réseau résultant est dirigé mais la pondération des liens, qui désigne le nombre de citations, n'est pas prise en compte. Le réseau est représenté sous trois formes :

- une représentation graphique 2D (cf. Réseau_francais, legend_france) ;
- une représentation graphique 3D disponible sur le site de la Cour des comptes ;

- une forme xlsx sous deux tableaux, l'un représentant tous les nœuds du réseau et leur catégorie (nodes_france) et un tableau représentant les liens entre les nœuds (edges_france).

Graphique n° 31 : représentation 2D du réseau français



Source : Cour des comptes d'après le web et l'outil Hyphe de Sciences Po Paris

b. Clustering

Afin de visualiser et analyser les effets de groupe, un algorithme de *clustering* par méthode de Louvain¹⁹⁹ a été exécuté sur le réseau *via* la

¹⁹⁹ Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics : Theory and Experiment*, 2008.

fonction *cluster_louvain* du *package igraph* sur R. L'algorithme de Louvain permet de détecter des communautés au sein d'un réseau de façon automatisée. Intuitivement, une communauté constitue un sous-ensemble de nœuds du réseau densément connectés. Les nœuds d'une communauté sont généralement davantage connectés entre eux qu'avec les nœuds appartenant aux autres communautés. Les communautés repérées par l'algorithme sont au nombre de sept et réparties selon le tableau suivant :

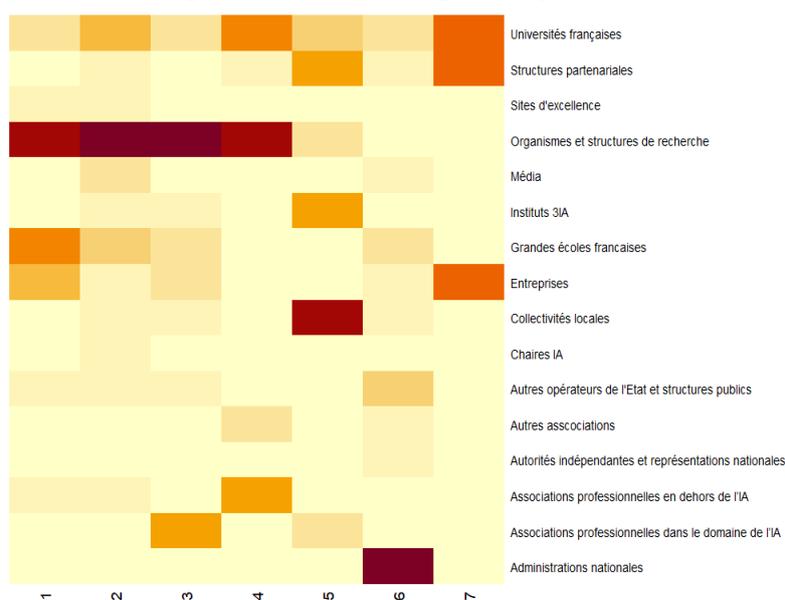
Tableau n° 36 : taille des communautés découvertes par la méthode de Louvain

<i>Nom de groupe</i>	1	2	3	4	5	6	7
<i>Pourcentage des institutions</i>	17,8 %	20,9 %	18,8 %	10,2 %	8,0 %	20,9 %	3,0 %

Source : Cour des comptes

Le dendrogramme ci-dessous représente le croisement entre la classification par communautés et par catégories :

Graphique n° 32 : croisement entre la classification par communautés et par catégories

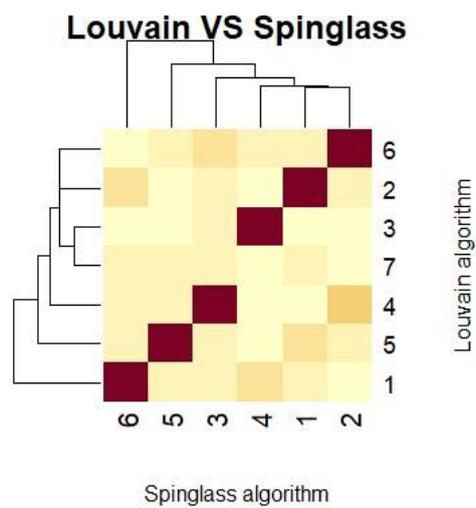


Source : Cour des comptes

Afin de valider l'adéquation de la méthode de *clustering* choisie à la structure du réseau, et en particulier de détecter les nœuds frontaliers qui séparent implicitement les *clusters*, ces résultats sont confrontés à ceux obtenus par un autre algorithme, Spinglass²⁰⁰, qui se base sur un critère différent pour la classification inspiré de la physique quantique. L'implémentation est réalisée à partir de la fonction *cluster_spinglass* du *package igraph* sur R. La figure ci-dessous confronte la classification par les deux méthodes.

²⁰⁰ J. Reichardt and S. Bornholdt, "Statistical Mechanics of Community Detection", *Physical Review E*, 2006.

Graphique n° 33 : croisement entre la classification par les méthodes de Louvain et Spinglass



Source : Cour des comptes

c. Indicateurs

Tableau n° 37 : indicateurs de centralité des catégories d'acteurs du réseau français

<i>Catégorie d'acteurs</i>	Importance « globale » (part « d'effondrement » de la structure du réseau lorsque la catégorie est « retirée »)	Importance « locale » (part des liens hypertextes entrants que représente la catégorie)
<i>Organismes et structures de recherche</i>	33 %	32 %
<i>Universités françaises</i>	18 %	16 %
<i>Administrations nationales</i>	11 %	23 %
<i>Grandes écoles françaises</i>	10 %	6 %
<i>Autres opérateurs de l'État et structures publics</i>	7 %	5 %
<i>Collectivités locales</i>	7 %	3 %
<i>Instituts 3IA</i>	6 %	2 %
<i>Associations professionnelles dans le domaine de l'IA</i>	5 %	1 %
<i>Associations professionnelles en dehors de l'IA</i>	4 %	2 %
<i>Structures partenariales</i>	4 %	2 %
<i>Sites d'excellence</i>	3 %	1 %
<i>Entreprises</i>	3 %	1 %
<i>Autorités indépendantes et représentations nationales</i>	2 %	4 %
<i>Média</i>	2 %	1 %
<i>Autres associations</i>	1 %	0 %
<i>Chaires IA</i>	1 %	0 %

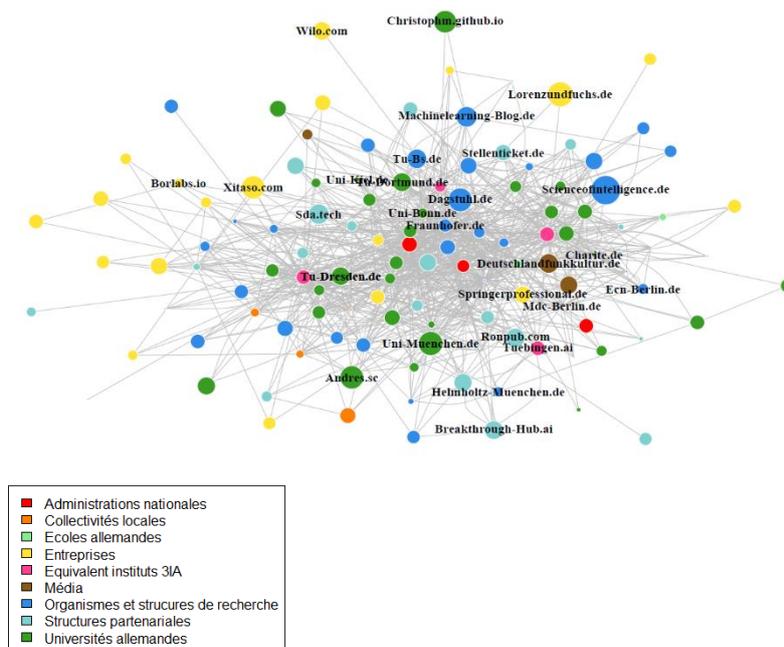
Source : Cour des comptes. Les définitions et modalités de calcul des indicateurs employés pour quantifier les importances globale et locale sont précisés ci-dessus

d. Réseau allemand et indicateurs associés par catégories d'acteurs :

La même méthodologie a été suivie pour établir le réseau allemand en partant des cinq centres équivalents aux instituts 3IA français :

- le *Berlin institute for the foundations of learning and data* ;
- *Center for Scalable Data Analytics and Artificial Intelligence* ;
- *ML2R – The Competence Center Machine Learning Rhine-Ruhr* ;
- *The Munich Center for Machine Learning* ;
- *Tübingen AI Center*.

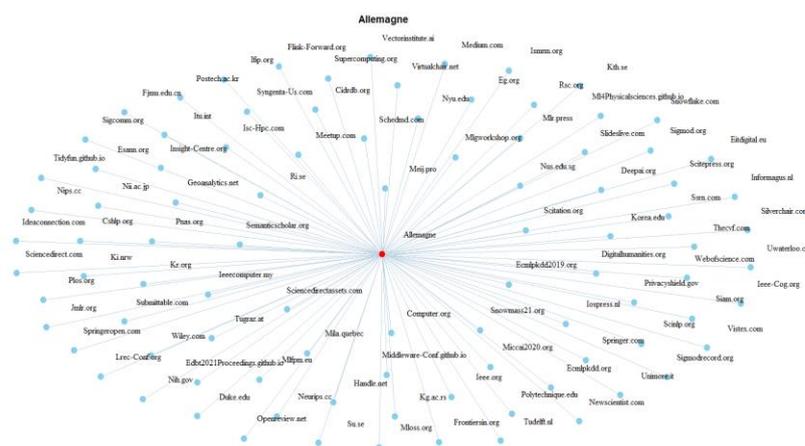
Graphique n° 34 : représentation du réseau allemand résultant des instituts équivalents aux 3IA



Source : Cour des comptes d'après le web et l'outil Hype de Sciences Po Paris

Les liens internationaux identifiés à partir de ces centres sont représentés dans le graphe ci-dessous :

Graphique n° 35 : représentation des liens internationaux sortants des centres allemands équivalents aux instituts 3IA français



Source : Cour des comptes d'après le web et l'outil Hyphe de Sciences Po Paris.

Les sites Web allemands sont classifiés selon les catégories d'acteur suivantes :

- administrations nationales (BMBF : ministère fédéral de l'éducation et de la recherche, etc.) ;
- collectivités locales (Dresde, Leipzig, Bayern, Berlin, etc.) ;
- universités allemandes ;
- écoles supérieures allemandes ;
- structures et organismes de recherche (DFG, DFKI, Fraunhofer, *Max-Planck institutes*, etc.) : secteur public ;
- structures partenariales : statut mixte entre privé et public ;
- entreprises : secteur privé ;
- équivalents des instituts 3IA : BIFOLD, ML2R, SCADS, Tübingen ;
- média.

Le réseau est également représenté sous trois formes :

- une représentation graphique 2D (cf. graphique n° 35) ;
- une représentation graphique 3D disponible sur le site de la Cour des comptes.

Tableau n° 38 : indicateurs de centralité des catégories d'acteurs du réseau allemand.

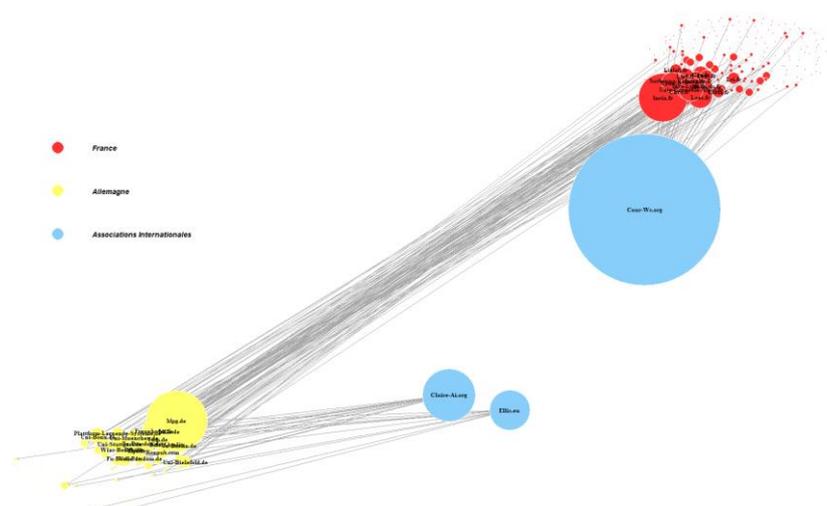
<i>Catégorie d'acteurs</i>	Importance « globale » (part « d'effondrement » de la structure du réseau lorsque la catégorie est « retirée »)	Importance « locale » (part des liens hypertextes entrants que représente la catégorie)
<i>Universités allemandes</i>	36 %	37 %
<i>Organismes et structures de recherche</i>	31 %	25 %
<i>Équivalent instituts 3IA</i>	19 %	5 %
<i>Structures partenariales</i>	18 %	10 %
<i>Entreprises</i>	9 %	7 %
<i>Collectivités locales</i>	8 %	4 %
<i>Administrations nationales</i>	5 %	7 %
<i>Média</i>	5 %	4 %
<i>Écoles allemandes</i>	2 %	1 %

Source : Cour des comptes. Les définitions et modalités de calcul des indicateurs employés pour quantifier les importances globale et locale sont précisées ci-dessus

e. Réseau franco-allemand :

Les nœuds du réseau franco-allemand sont l'union des nœuds des deux réseaux allemands et français et les liens représentés sont les liens entre les nœuds de chaque réseau. Les nœuds isolées ne sont pas représentés sur le graphe mais sont présents dans les tableaux `xlsx` : `nodes_franco_allemand` et `edges_franco_allemand`, pour pouvoir étudier de point de vue métrique la réciprocité et l'assortativité des liens entre les réseaux en entier.

Graphique n° 36 : représentation des liens franco-allemand entre les sites présents dans les deux réseaux



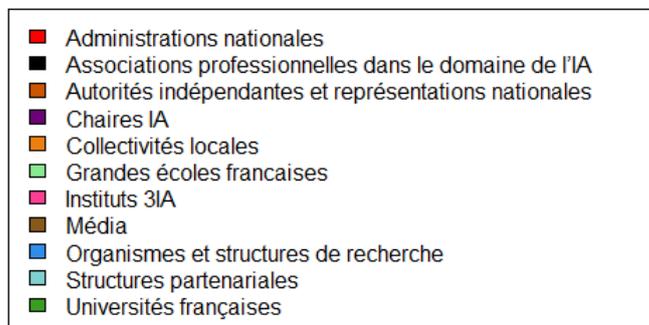
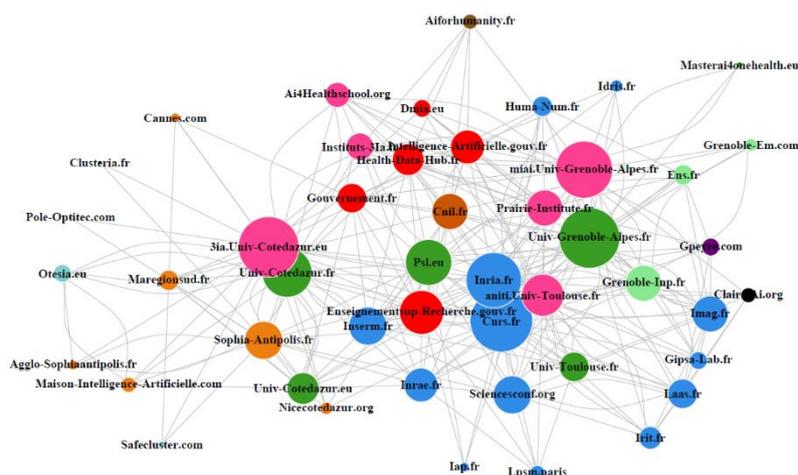
Source : Cour des comptes d'après le web et l'outil Hyphe de Sciences Po Paris

f. Instituts 3IA

Sous réseau français et indicateurs associés par catégories d'acteurs :

Afin de comparer de point de vue métrique les deux réseaux allemands et français, et pour contourner un biais potentiel lié aux modalités d'exploration du réseau (tailles de la « graine » de sites web différentes dans les deux cas : uniquement les 3IA allemand pour l'Allemagne ; nombreux autres sites pour la France), un autre réseau est établi en partant seulement des instituts 3IA.

Graphique n° 37 : représentation du sous-réseau français résultant des instituts 3IA



Source : Cour des comptes d'après le web et l'outil Hyphe de Sciences Po Paris

Tableau n° 40 : indicateurs de centralité des catégories d'acteurs du sous-réseau français (au périmètre des 3IA)

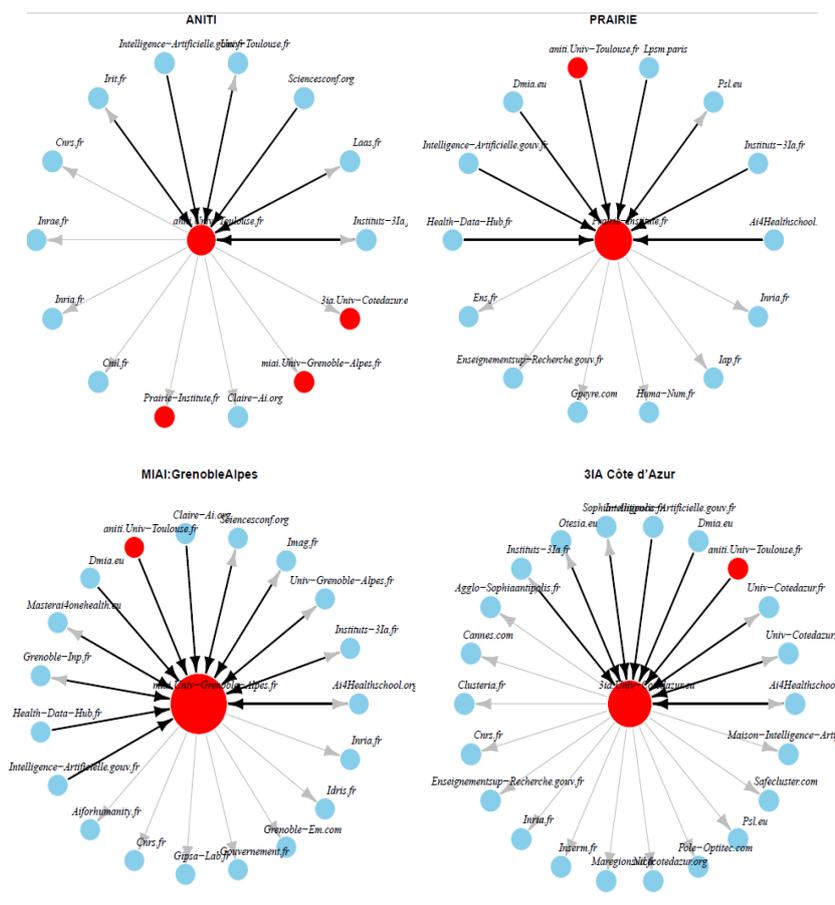
<i>Catégorie d'acteurs</i>	Importance « globale » (part « d'effondrement » de la structure du réseau lorsque la catégorie est « retirée »)	Importance « locale » (part des liens hypertextes entrants que représente la catégorie)
<i>Organismes et structures de recherche</i>	43 %	35 %
<i>Instituts 3IA</i>	26 %	15 %
<i>Universités françaises</i>	17 %	15 %
<i>Administrations nationales</i>	17 %	7 %
<i>Collectivités locales</i>	12 %	5 %
<i>Grandes écoles françaises</i>	7 %	11 %
<i>Autorités indépendantes et représentations nationales</i>	7 %	6 %
<i>Structures partenariales</i>	7 %	3 %
<i>Média</i>	2 %	2 %
<i>Associations professionnelles dans le domaine de l'IA</i>	2 %	1 %
<i>Chaires IA</i>	2 %	0 %

Source : Cour des comptes. Les définitions et modalités de calcul des indicateurs employés pour quantifier les importances globale et locale sont précisés ci-dessus

Réseaux centraux français :

L'ensemble des liens sortants et entrant par institut 3IA est représenté par des réseaux centraux. À ce stade de l'analyse, les liens entrants n'étaient pas exhaustifs et ne représentent que les liens avec les entités présentes dans le corpus.

Graphique n° 38 : représentation des liens sortants et entrants par instituts 3IA



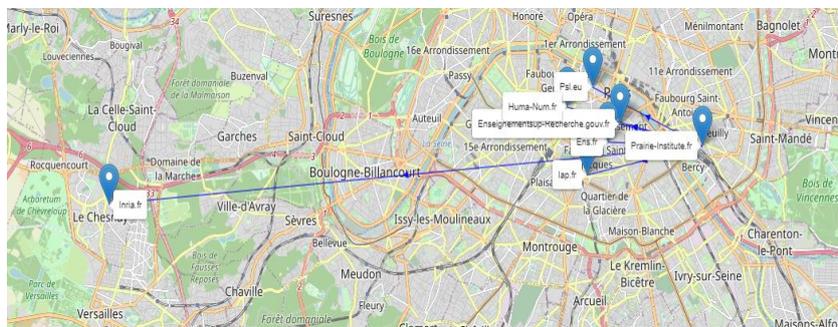
Source : Cour des comptes

Cartographie :

Dans le but de modéliser l'impact territorial de chacun des instituts 3IA, des représentations géographiques des liens sortants par centre ont été réalisées. C'est la représentation cartographique des feuilles et des liens présents dans les réseaux centraux. Dans un premier temps, les adresses postales de tous les sites en lien avec l'un des instituts 3IA ont été récupérées en cherchant l'url sur Google. Ensuite, pour chaque adresse postale des coordonnées géographiques (latitude/longitude) ont été affectées par un script automatisé sur R en utilisant la fonction geocode_OSM du package Tmap. Les quatre versions interactives des

cartes sont réalisées également sur R par les deux packages : Leaflet pour la représentation du fond de la carte et les points et Leaflet.minicharts pour ajouter les liens entre les points, et sont disponibles en format html en passant par le package HtmlTools (cf. prairie, miai, azur, aniti).

Graphique n° 39 : représentation cartographique des liens sortants de l'institut 3IA : Prairie



Source : Cour des comptes

Liste des sites de référence pour la construction du réseau français :

- CEA : <https://www.cea.fr/Pages/liens-utiles/sites-internet-du-cea-et-autres-liens-utiles.aspx?Type=Chapitre&numero :3> ;
- SNPIA : <https://www.intelligence-artificielle.gouv.fr/fr/partenaires> ;
- I3IA : <https://instituts-3ia.fr/> ;
- MIAI_pjt : <https://miai.univ-grenoble-alpes.fr/reseaux-internationaux/projets-internationaux/> ;
- MIAI_masters_int : <https://miai.univ-grenoble-alpes.fr/reseaux-internationaux/masters-internationaux-ecoles-internationales-/> ;
- MIAI_partenaires : <https://miai.univ-grenoble-alpes.fr/institut-miai/ecosysteme-d-innovation-/> ;
- I3IA_CA_ecosys : <https://3ia.univ-cotedazur.eu/about/ecosystem-1> ;
- I3IA_CA_inter : <https://3ia.univ-cotedazur.eu/about/international> ;
- PRAIRIE : <https://prairie-institute.fr/> ;
- PRAIRIE_dhai : <https://dhai-seminar.github.io/> ;
- PRAIRIE_cogai : <https://cogai-seminar.github.io/> ;
- ANITI : <https://aniti.univ-toulouse.fr/linstitut/reseau-des-3ia/> ;
- HIPARIS_symp : <https://aisymposium.hi-paris.fr/list-of-organizers/>, #mêmes partenaires que HI Paris ;
- SCAI_partners : <https://scai.sorbonne-universite.fr/public/contactus> ;
- SCAI_education : <https://scai.sorbonne-universite.fr/public/education> ;
- SCAI_research : <https://scai.sorbonne-universite.fr/public/research> ;
- DATAIA_inter : <https://DATAIA.eu/partenaires/cooperations-international> ;
- DATAIA_ecosys : <https://DATAIA.eu/partenaires/ecosysteme-du-site> ;
- DATAIA_ellis : <https://DATAIA.eu/partenaires/ellis-paris> ;
- ACTUIA : <https://www.actuia.com/actualite/etudes-en-intelligence-artificielle-le-grand-panel-des-formations-dossier/> ;
- ACTUIA_ES : <https://www.actuia.com/universites-enseignement-superieur/> ;
- BDT : <https://www.banquedesterritoires.fr/nos-filiales-et-partenaires> ;

- Health Data Hub : <https://www.health-data-hub.fr/nos-partenaires> ;
- ELLIS_Paris : <https://ellis-paris.github.io/events/> ;
- CNRS : <https://www.cnrs.fr/fr/> ;
- Inria : <https://www.inria.fr/fr> ;
- Inserm : <https://www.inserm.fr/> ;
- MESRI : <https://www.enseignementsup-recherche.gouv.fr/fr/> ;
- SGPI : <https://www.gouvernement.fr/secretariat-general-pour-l-investissement-sgpi> ;
- GPAI : <https://gpai.ai/> ;
- ANR : <https://anr.fr/fr> ;
- Bpifrance : <https://www.bpifrance.fr/> ;
- ADEME : <https://www.ademe.fr/> ;
- INSMI= <https://www.insmi.cnrs.fr/> ;
- INSMI_GAL=<https://www.insmi.cnrs.fr/fr/insmi> ;
- INSMI_REC=<https://www.insmi.cnrs.fr/fr/recherche> ;
- INSMI_INT=<https://www.insmi.cnrs.fr/fr/interactions> ;
- INS2I = <https://www.ins2i.cnrs.fr/> ;
- INS2I_REC = <https://www.ins2i.cnrs.fr/fr/recherche> ;
- INS2I_INN = <https://www.ins2i.cnrs.fr/fr/innovation> ;
- INS2I_TAL = <https://www.ins2i.cnrs.fr/fr/talent/index> ;
- INS2I_GAL = <https://www.ins2i.cnrs.fr/fr/ins2i>.

Annexe n° 9 : consultation de la communauté scientifique française dans le domaine de l'IA

Le cadre de la consultation

Dans le cadre de l'évaluation, un exercice de sondage a été mené sous la forme d'une consultation. Elle vise à analyser la mise en œuvre de la première phase de cette stratégie (2018-2022) et à éclairer celle de la seconde phase (stratégie d'accélération 2022-2025).

Cette consultation est adressée aux chercheurs, enseignants-chercheurs, ingénieurs de recherche, doctorants, post-doctorants, et à toutes autres personnes identifiées comme travaillant dans le domaine de l'IA par le CNRS, Inria, le CEA et les universités (76 % d'entre elles ont répondu à la sollicitation de la Cour). L'Inserm n'a pas identifié en son sein de chercheur spécialisé en IA. La consultation a pour principal objectif d'évaluer les conditions de travail des personnels d'enseignement et de recherche en IA, de l'attractivité du métier en France et à l'étranger ainsi que les orientations stratégiques en IA. La consultation a été administrée au travers de l'outil Sphinx de manière anonyme.

Les thèmes de la consultation par le biais de 44 questions :

- mieux cerner votre profil et votre parcours professionnel ;
- votre perception de la stratégie nationale de recherche en IA ;
- votre avis sur l'attractivité et la rétention des talents ;
- votre avis sur la valorisation et la collaboration avec le secteur privé ;
- les perspectives d'amélioration qui pourraient être suggérées.

Les éléments quantitatifs et qualitatifs

La consultation a été envoyée par mail à 2 125 destinataires à partir des adresses mails fournis par le CNRS, Inria, le CEA et les universités. 7 % des mails n'ont pas été délivrés et donc 1 978 destinataires ont été contactés.

Entre le 22 juin et le 29 août 2022, il y a eu 772 réponses complètes et exploitables. Soit un taux de réponses de 39 % (dont les trois quarts de réponses complètes).

Les enseignants-chercheurs et chercheurs sondés ayant répondu sont rattachés à 53 universités, ainsi qu'à trois organismes de recherche (CEA, CNRS, Inria) avec un taux de réponse similaire, ce qui assure une certaine représentativité des réponses à partir de la base de consultation

(cf. tableau n° 40). Néanmoins, étant donné les biais d'identification de la population cible, les résultats ne peuvent être interprétés comme représentatifs et seront interprétés avec précaution.

Tableau n° 41 : taux de réponses des personnels à la consultation

<i>Organismes</i>	<i>Université</i>	<i>Inria</i>	<i>CNRS</i>	<i>CEA</i>
<i>Part au sein de la consultation</i>	1 857	487	203	66
<i>Taux de réponse</i>	31,02 %	25,05 %	21,67 %	33,33 %

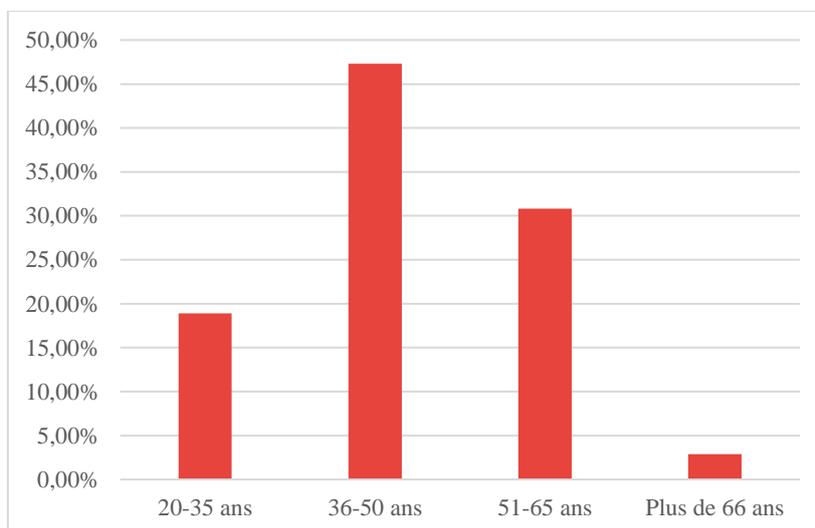
Source : Cour des comptes

Les principaux résultats et enseignements

Le profil des répondants

Les chercheurs sont à 92,2 % de nationalité française. Pour les autres, 28,6 % sont italiens, 10,7 % allemands et 8,9 % belges. La consultation étant administrée en français, cela a pu occasionner une exclusion des personnels non-francophones. Plus de 35 % ont entre 41 et 50 ans et 62 % entre 36 et 55 ans, et près de 80 % sont des hommes.

Graphique n° 40 : âge des répondants à la consultation



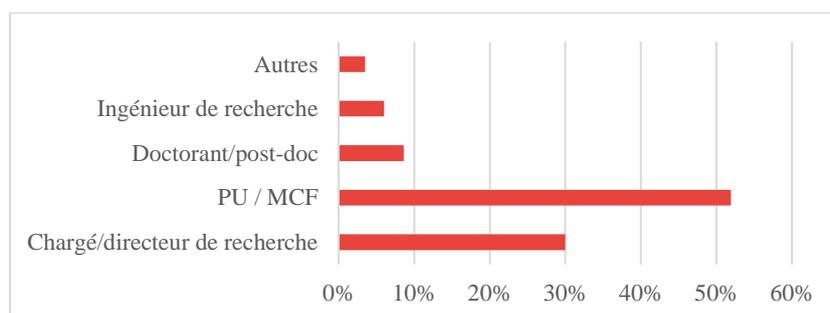
Source : Cour des comptes

Ils ont effectué leur formation initiale en université (61,3 %) ou dans une grande école (50,9 %). Il est à noter que 19,4 % ont effectué leur formation initiale dans un établissement d'enseignement supérieur à l'étranger. Près de 63 % ont une habilitation à diriger des recherches (HDR) et 27,1 % sont d'un niveau doctorat. Ce résultat suggère que les répondants sont plutôt des chercheurs séniors.

Près de 54 % ont passé plus de deux ans hors de France dans le cadre de leurs études supérieures ou activités professionnelles.

Plus de 50 % exercent leur activité en université (30,4 % professeurs d'université et 21,5 % maîtres de conférences) et 30 % sont directeurs de recherche (16,7 %) ou chargés de recherche (13,3 %).

Graphique n° 41 : activité professionnelle principale



Source : Cour des comptes. Représentation de la profession des répondants à la consultation. PU professeur des universités et MCF maître de conférences

Environ 48 % sont affiliés à un pôle d'excellence du type « 3IA » ou hors « 3IA » (DATA IA pour 13,4 %, ANITI pour 8,8 %, SCAI pour 7,3 % et PRAIRIE pour 7 %). Une majorité des scientifiques n'est affiliée à aucune association ou réseau professionnel en IA (58,6 %). L'association la plus citée est l'AFIA²⁰¹ (13,7 %).

²⁰¹ AFIA : association française pour l'intelligence artificielle ayant pour but la promotion, le développement de l'IA, sous ses différentes formes, de regrouper et de faire croître la communauté française en IA, et d'en assurer la visibilité.

Tableau n° 42 : affiliation à un centre d'excellence en IA

<i>Centre d'excellence</i>	Taux d'affiliation
<i>Prairie</i>	6,83 %
<i>MIAI</i>	2,23 %
<i>3IA Côte d'Azur</i>	4,32 %
<i>ANITI</i>	8,65 %
<i>SCAI</i>	7,11 %
<i>DATAIA</i>	13,11 %
<i>Hi! PARIS</i>	1,53 %
<i>Autres chaires 3IA (hors Instituts 3IA)</i>	5,58 %
<i>Autres</i>	50,63 %

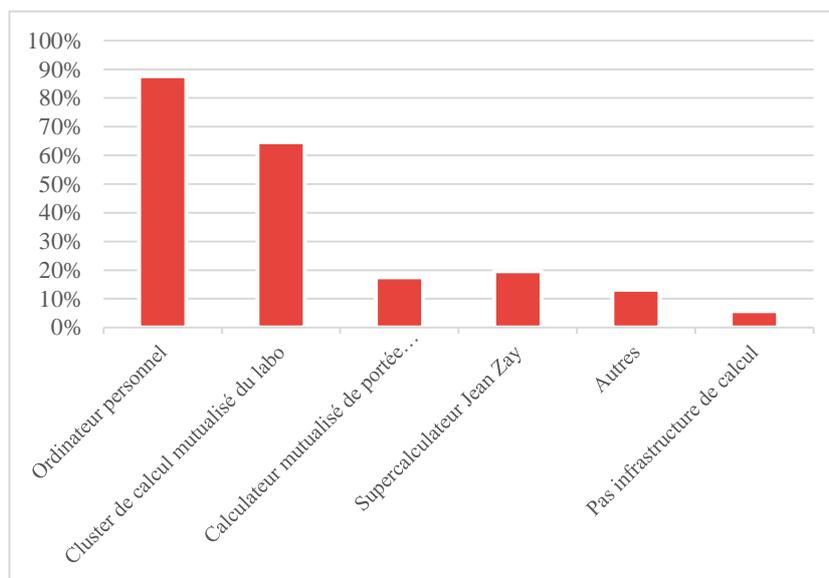
Source : Cour des comptes

En matière de rattachement à une section CNU, 46,5% correspondent à la section informatique, 18,5 % aux mathématiques appliqués et 15,3 % au génie informatique, automatique et traitement du signal.

Le chercheur développe principalement des méthodes nouvelles dans le domaine de l'IA (73,7 %), mais peut également mobiliser des méthodes existantes (46,8 %). Pour 86,1 % d'entre eux, ce sont des méthodes statistiques par apprentissage (à partir de données). Des méthodes mathématiques hypothético-déductives, dont l'IA symbolique à partir de connaissance préalable, sont utilisées pour 29,2 %.

C'est l'intérêt de la thématique (83,6 %) qui a induit le choix de l'IA comme domaine professionnel. Son émulation scientifique est évoquée pour 53,2 % des répondants. Les valeurs importantes qui sous-tendent la motivation et l'engagement des répondants sont la découverte scientifique pour 72,6 % des chercheurs ayant répondu, 42 % la vie familiale et 37,5 % l'impact sociétal positif de son travail de recherche.

Les infrastructures de calcul utilisées sont en premier lieu l'ordinateur personnel (87,6 %), puis le *cluster* de calcul ou moyen de calcul mutualisés du laboratoire (64,6 %). Le supercalculateur Jean Zay est utilisé pour près de 20 % des chercheurs. Ces résultats montrent des usages par un même répondants de plusieurs infrastructures de calcul et que Jean Zay n'est mobilisé que par une petite partie de la communauté.

Graphique n° 42 : infrastructures de calcul utilisées

Source : Cour des comptes

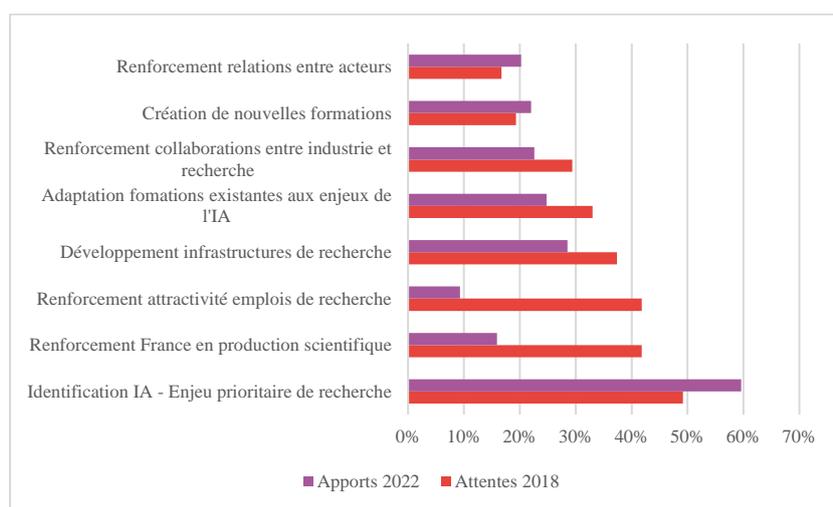
Perception de la stratégie nationale de recherche en IA

Les chercheurs connaissent à 82 % l'existence de de la stratégie nationale de recherche en IA (SNRIA). Mais seulement 18,1 % en connaissent bien son contenu.

En 2018, au moment de son lancement, les attentes principales étaient une identification de l'IA comme un enjeu prioritaire de recherche (49,2 %), le renforcement de la place de la France en termes de production scientifique (41,8 %), le renforcement de l'attractivité des emplois de recherche (41,8 %), le renforcement des infrastructures de recherche (37,4 %) et le renforcement de collaborations entre industrie et recherche (29,4 %).

En 2022, après quatre années de mise en œuvre de la SNRIA, un certain nombre d'apports correspondent aux attentes. Une identification de l'IA comme enjeu prioritaire de recherche pour 59,6 % des chercheurs. Le développement des infrastructures de recherche pour 28,5 % et le renforcement des collaborations entre industrie et recherche (22,6 %) Un autre aspect valorisé est l'adaptation des formations existantes aux nouveaux enjeux de l'IA (24,6 %), qui était une attente forte également en 2018 (33 %).

Graphique n° 43 : attentes et apports de la SNRIA



Source : Cour des comptes

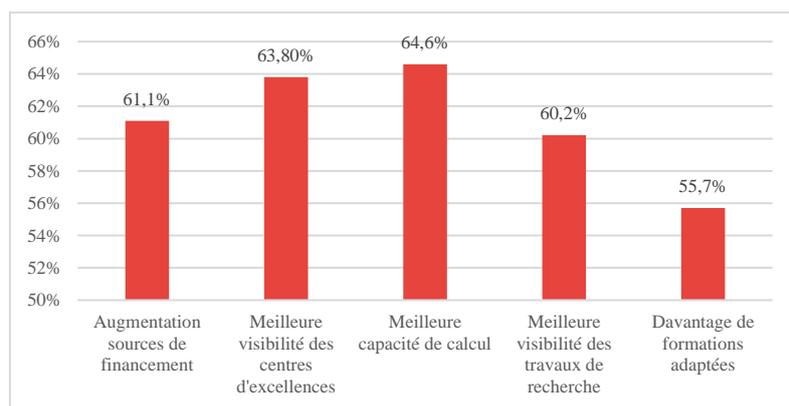
Les dispositifs mis en œuvre les plus connus sont les appels à projets « IA » de l'ANR à 71,7 %, le supercalculateur « Jean Zay » à 68,6 %, les instituts « 3IA » à 65,7 %, ainsi que les chaires académiques en IA à 64,5 %.

Près de 70 % des chercheurs ont été lauréats ou bénéficiaires d'un ou plusieurs appels à projets en dehors de la SNIA, principalement de l'ANR pour 39,3 % ou d'une entreprise pour 19,2 %. Près de la moitié (44 %) considère que la SNRIA n'a pas eu d'impact sur les réussites aux appels à projets.

Depuis la mise en place de la stratégie nationale, les principaux changements dans l'exercice du métier concernent l'augmentation des sources de financements (61,1 %), la meilleure capacité de calcul (64,4 %), une meilleure visibilité des travaux de recherche (60,2 %) et davantage de formations adaptées (55,7 %).

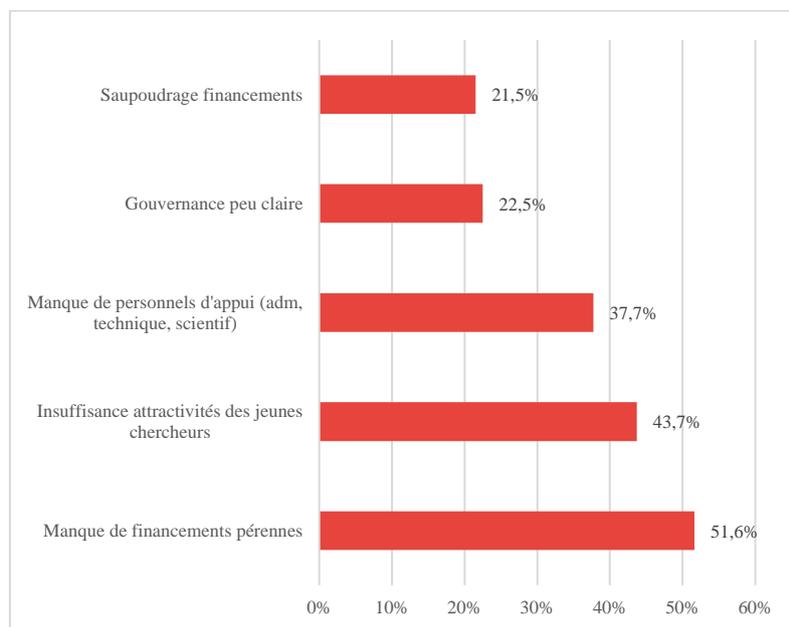
Les principaux freins à la SNRIA concernent le manque de financements pérennes (51,6 %), une insuffisance d'attractivité des jeunes chercheurs (43,7 %) et le manque de personnel technique, administratif et scientifique d'accompagnement des chercheurs (37,7 %).

Graphique n° 44 : principaux changements dans l'exercice du métier de chercheur depuis la SNRIA



Source : Cour des comptes

Graphique n° 45 : les principaux freins à la SNRIA



Source : Cour des comptes

Attractivité et la rétention des talents

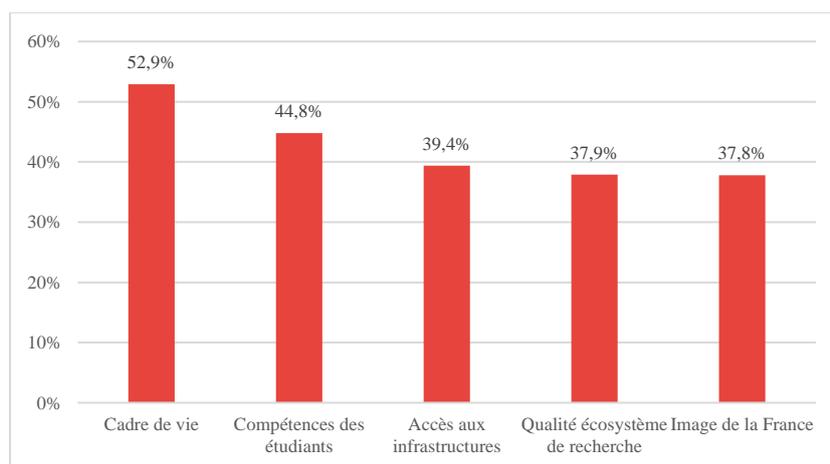
Le salaire annuel minimum en début de carrière d'un jeune chercheur en IA qui ne soit pas un élément dissuasif se situe dans une fourchette entre 30 000 € et 50 000 € pour 80 % des sondés. La moyenne étant d'environ 40 000 €. Quant au « package » financier (fonds pour le recrutement de doctorants, post-doctorants et achat de matériel) hors rémunération du chercheur, celui-ci est estimé majoritairement dans une fourchette entre 31 000 € et 151 000 €. La moyenne étant d'environ 90 000 €.

Les chercheurs considèrent, par ailleurs, que la SNIA n'a pas réellement permis d'améliorer la qualité des recrutements. L'effet le plus positif concerne les étudiants en masters (pour 28,8 % des réponses).

Plusieurs profils sont considérés comme en tension (à 69,5 % pour les jeunes chercheurs et 58 % pour les doctorants hors CIFRE). Ce qui est moins vrai pour les étudiants en masters (32,5 %) et les doctorants CIFRE (39 %). Globalement, ces profils sont considérés en tension pour 49,8 % des sondés, des profils disponibles pour 33 % et surabondants pour 4,8 %.

L'attractivité de la recherche en France est contrastée. Les points positifs de la recherche en France sont le cadre de vie (53 %), les compétences des étudiants (44,8 %), l'accès aux infrastructures (39,4 %), la qualité de l'écosystème de recherche (37,9 %) et l'image de la France (37,8 %).

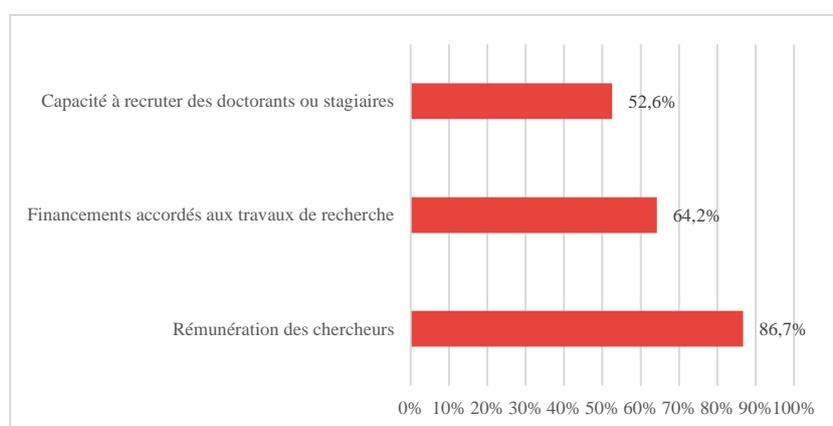
Graphique n° 46 : attractivité de la recherche en France : aspects positifs



Source : Cour des comptes

Les points négatifs de la situation de la recherche en France sont la faible rémunération des chercheurs (86,7 %), les financements accordés aux travaux de recherche (64,2 %) et la capacité à recruter des doctorants ou des stagiaires (52,6 %)

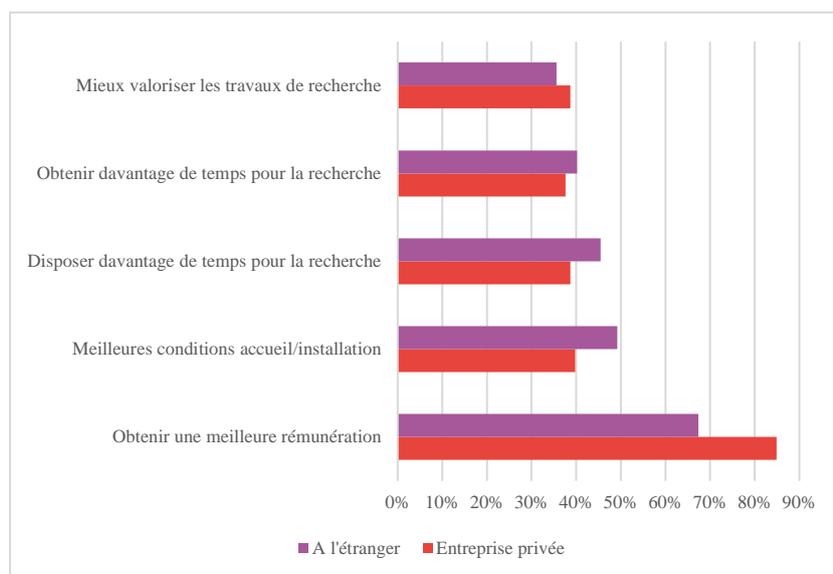
Graphique n° 47 : attractivité de la recherche en France : aspects négatifs



Source : Cour des comptes

Pour les années à venir, 86,4 % des sondés souhaitent continuer leurs travaux de recherche en France, et 20 % à l'étranger. Environ 30 % souhaitent prendre des responsabilités dans le management de la recherche. Ceux qui projettent d'exercer dans une entreprise privée le font pour obtenir une meilleure rémunération pour 84,9 %. Mais d'autres motivations, de moindre mesure, sont exprimées (obtenir de meilleures conditions d'accueil et d'installation, obtenir davantage de moyens pour leurs travaux, disposer davantage de temps pour la recherche, mieux valoriser les travaux de recherche). Pour ceux qui envisagent de travailler à l'étranger, les motivations sont les mêmes que travailler dans une entreprise privée. La principale motivation est ainsi d'obtenir une meilleure rémunération (67,4 %), obtenir de meilleures conditions d'installation (49,2 %), disposer de davantage de temps pour la recherche (45,5 %) et obtenir davantage de temps pour la recherche (40,2 %).

Graphique n° 48 : motivations pour exercer en entreprise et/ou à l'étranger



Source : Cour des comptes

Les pays considérés par la communauté scientifique comme étant les plus dynamiques et attractifs en IA sont les États-Unis (95,8 %), l'Allemagne (58,6 %), le Royaume-Uni (52,4 %), le Canada (48 %) et la Chine (44,1 %). La France arrivant en sixième place (42,1 %).

La valorisation et la collaboration avec le secteur privé

Près de 70 % des répondants disent être incités à valoriser leurs travaux de recherche. Ceux qui ne le font pas arguent du manque de temps (48 %), de la complexité des procédures (37,8 %) ou par l'insuffisance du soutien par les organismes en charge de la valorisation (SATT ou autres structures de valorisation) pour 34,7 %.

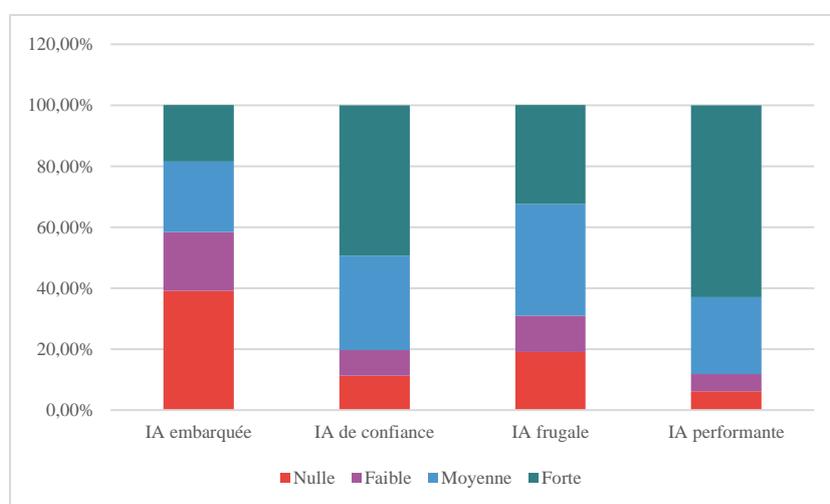
Les perspectives d'amélioration qui pourraient être suggérées

Les chercheurs sont plutôt partagés sur le fait de savoir si la communauté éducative est ou pas suffisamment coordonnée sur les thématiques de formation en IA pour accroître les effectifs étudiants (55,5 % non et 44,5 % oui).

La prise en compte, par les chercheurs, des enjeux de l'IA est différente selon sa nature. L'IA embarquée est pas ou peu intégrée dans les

recherches pour près de 60 % des chercheurs. Par contre, l'IA performante est considérée comme un enjeu moyen ou fort pour près de 90 % d'entre eux. L'IA de confiance est également un enjeu bien intégré (80,3 %). L'IA frugale est un enjeu moyen ou fort pour près de 70 % des chercheurs ayant répondu.

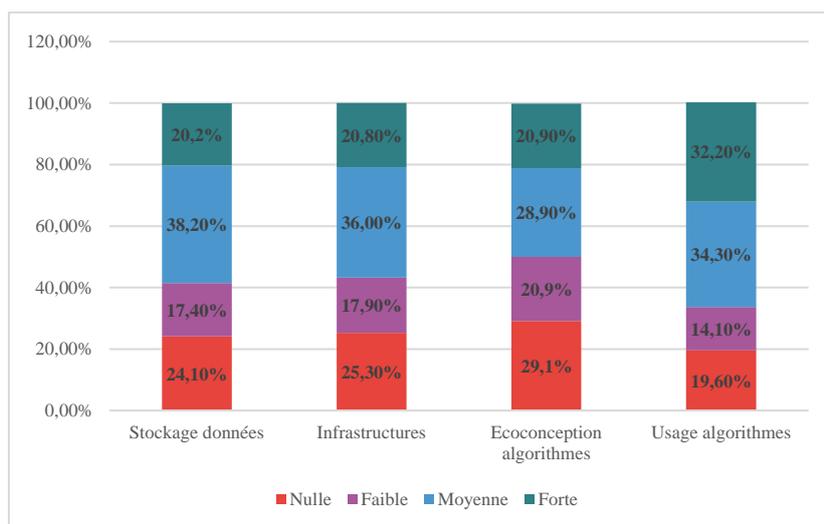
Graphique n° 49 : niveau de prise en compte dans les recherches



Source : Cour des comptes

Globalement, la prise en compte de l'impact environnemental de la recherche en IA est forte dans seulement 23,53 % de répondants et nulle dans 24,5 %. Avec des situations légèrement différentes selon les éléments suivants : stockage des données, infrastructures, écoconception des algorithmes et leur usage.

Graphique n° 50 : prise en compte de l'impact environnemental dans la recherche



Source : Cour des comptes

Le niveau considéré comme la plus pertinente pour concevoir et porter la stratégie en IA est l'échelle européenne (66,8 %), puis nationale (55,3 %).

Annexe n° 10 : évolutions des effectifs d'inscrits dans les filières d'élection préparant aux formations spécialisées en IA

I. Filières de formation initiale de l'enseignement scolaire

A- Cartographie des filières

Schéma n° 6 : cartographie et dynamique d'adoption des principales matières de l'enseignement scolaire introduisant l'IA ou préparant aux filières scientifiques telles que l'IA

Seconde générale et technologique	Terminale générale (visée de poursuite d'études dans le supérieur pour une durée de trois à cinq ans au moins)
<p>571 604 élèves formés en 2021 (+1,1 % depuis 2020) dont 53,9 % de femmes (part de femmes identique en 2020)</p>	<p>375 326 élèves formés en 2021 (-0,1 % depuis 2020) dont 56,2 % de femmes (+0,1 point depuis 2020)</p>
<p>Deux enseignements communs obligatoires préparant à des filières scientifiques telles que l'IA :</p> <ul style="list-style-type: none"> ➤ Mathématiques (4h / semaine) ➤ Sciences numériques et technologies (1h30 / semaine) 	<p>Un enseignement commun obligatoire appelé « enseignement scientifique » (2h / semaine) où les principes et enjeux de l'IA sont introduits</p> <p>Des spécialisations préparant à des filières scientifiques telles que l'IA :</p> <ul style="list-style-type: none"> ➤ Spécialité mathématiques (6h / semaine) 140 649 élèves formés en 2021 (-3,9 % depuis 2020) dont 39,8 % de femmes (-2,1 points depuis 2020) ➤ Option « mathématiques complémentaires » (3h / semaine) proposée aux élèves qui ne choisissent pas la spécialité mathématique 64 336 élèves formés en 2021 (-0,8 % depuis 2020) dont 62,6 % de femmes (-0,4 points depuis 2020) ➤ Option « mathématiques expertes » réservée aux élèves qui choisissent la spécialité mathématiques et qui visent des filières où les mathématiques sont prépondérantes (+ 3h / semaine, soit 9h / semaine au total) 48 491 élèves formés en 2021 (-6,7 % depuis 2020) dont 31,4 % de femmes (-1,6 points depuis 2020) ➤ Spécialité numérique et sciences informatiques (6h / semaine) 16 184 élèves formés en 2021 (+ 16,4 % depuis 2020) dont 13,7 % de femmes (+ 0,6 points depuis 2020)

Sources : Cour des comptes d'après les chiffres de la DEPP du MENJS ainsi que les ressources documentaires du MENJS et de l'Onisep qui décrivent les programmes

Tableau n° 43 : évolution de la répartition des élèves en seconde générale et technologique

2020		2021		Évolution 2020 à 2021 (en % et en points)	
Nombre d'élèves	Part des femmes (%)	Nombre d'élèves	Part des femmes (%)	Nombre d'élèves	Part des femmes
565 476	53,9	571 604	53,9	+ 1,1 %	0,0

Source : évolutions calculées par la Cour des comptes d'après les chiffres de la DEPP du MENJS (section 4.10 du RERS 2021 et 2022 de la DEPP pour les chiffres 2020 et 2021 respectivement, au périmètre des établissements publics et privés sous contrat)

Tableau n° 44 : évolution de la répartition des élèves de terminale générale selon l'enseignement de spécialité suivi de la rentrée 2020 à la rentrée 2021.

Enseignement de spécialité	2020			2021			Évolution 2020 à 2021 (en % et en points)		
	Nb élèves	Part des élèves de terminale générale (%)	Part des femmes (%)	Nb élèves	Part des élèves de terminale générale (%)	Part des femmes (%)	Nb élèves	Part des élèves de terminale générale	Part des femmes
Mathématiques	154 444	41,2	41,9	140 648	37,5	39,8	- 8,9%	- 3,7	- 2,1
Physique-chimie	126 671	33,7	47,5	117 136	31,2	47,4	- 7,5%	- 2,5	- 0,1
Sciences de la vie et de la Terre (SVT)	101 022	26,9	63,2	96 679	25,8	63,3	- 4,3%	- 1,1	+ 0,1
Numérique et sciences informatiques (NSI)	13 907	3,7	13,1	16 184	4,3	13,7	+ 16,4 %	+ 0,6	+ 0,6
Ensemble des élèves de terminale générale	375 826	100,0	56,1	375 326	100,0	56,2	- 0,1%	0,0	+ 0,1

Sources : chiffres tirés de la note d'information n° 21.41 de décembre 2021 et de la note d'information n° 20.38 de novembre 2020 de la DEPP. S'agissant des données relatives à l'année 2022, la DGESCO a indiqué à la Cour que les données de 2022 de la DEPP ne seront pas disponibles avant l'automne prochain.

Tableau n° 45 : évolution de la répartition des élèves de terminale générale selon l'enseignement optionnel facultatif suivi de la rentrée 2020 à la rentrée 2021

Option	2020			2021			Évolution 2020 à 2021 (en % et en points)		
	Nombre d'élèves	Part des élèves de terminale générale (%)	Part des femmes (%)	Nombre d'élèves	Part des élèves de terminale générale (%)	Part des femmes (%)	Nombre d'élèves	Part des élèves de terminale générale	Part des femmes
Mathématiques expertes	51 969	13,8	33,0	48 491	13,0	31,4	- 6,7 %	- 0,8	- 1,6
Mathématiques complémentaires	64 866	17,3	63,0	64 336	17,2	62,6	- 0,8 %	- 0,1	- 0,4

Source : évolutions calculées par la Cour des comptes d'après les chiffres de la DEPP du MENJS (sections 4.09 et 4.16 du RERS 2021 et 2022 de la DEPP pour les chiffres 2020 et 2021 respectivement)

B- Actions annoncées par la DGESCO pour renforcer l'attrait pour les enseignements préparant à des domaines tels que l'IA et réduire les biais de genre

Actions pour augmenter la diffusion de la spécialité NSI

Dans sa réponse à la Cour sur les actions envisagées pour augmenter la diffusion de la spécialité NSI, la DGESCO indique que « le ministère chargé de l'éducation nationale a apporté son appui à un projet porté par l'association « Talents du numérique ». Ce projet a abouti à la mise en place, pour la première fois en 2022, des « Trophées NSI », qui récompensent des projets remarquables réalisés au lycée au cours de l'année scolaire dans le cadre de l'enseignement de spécialité NSI. La mobilisation des classes autour de ce concours vise à promouvoir la richesse de cet enseignement et à inciter les élèves, et particulièrement les filles, à s'y intéresser lors de leurs choix de spécialités.

À ce titre, les projets, proposés par les enseignants et leurs élèves en 2022, ont démontré la grande variété des thèmes abordés par cette spécialité : en ingénierie ; dans les jeux vidéo ; dans l'écologie (covoiturage, panneau voltaïque, etc.), la citoyenneté (création de ville, conséquences des modalités de votes, etc.) et dans l'art (redimensionnement d'image, création de flocons, etc.). Cette grande variété sera mise en avant, notamment à titre de promotion de cette spécialité auprès, entre autres, des filles. Les Trophées sont remis aux lauréats lors de la « Journée nationale NSI ».

La première édition de la « Journée nationale NSI », renouvelée chaque année en lien avec des partenaires associatifs et industriels, soutenu notamment par le ministère, a eu lieu le 7 juin 2022, sous le signe de l'orientation, et fut un véritable succès. Elle s'est adressée aux élèves (collèges, lycées) en recherche d'information pour leur orientation scolaire, ainsi qu'à l'ensemble des prescripteurs (parents, professeurs, acteurs de l'orientation, etc.) dans un objectif de promotion des formations et des métiers du numérique et de l'informatique auprès des jeunes en orientation. Tous les acteurs du numérique en région, en établissements d'enseignement secondaires et supérieur, en entreprises, en associations, en laboratoires, en centres de recherche sont incités à mener des actions à destination des élèves et du corps enseignant, dans le cadre de cette manifestation tels que des visites de sites, d'ateliers pratiques, d'interventions en classe, de forums des métiers, de salons d'orientation. »

Actions pour maintenir l'attractivité de la spécialité mathématiques

Dans sa réponse à la Cour sur les actions envisagées pour maintenir l'attractivité de la spécialité mathématiques, la DGESCO indique que « le sentiment de compétences en mathématiques des élèves a une incidence importante dans les choix d'orientation. Le choix de l'enseignement de spécialité mathématiques dans le cycle terminal de la voie générale, au lycée en particulier et dans les filières scientifiques, que ce soit dans les voies technologique, professionnelle ou dans le choix de poursuites d'études après le baccalauréat, est souvent déterminé par le genre ou l'origine sociale des élèves. En effet, à évaluations égales, les jeunes filles et les élèves des milieux modestes choisissent moins fréquemment les filières mathématiques et scientifiques que les jeunes garçons et les élèves des milieux favorisés.

Des ateliers de sensibilisation, visant à élever l'ambition de ces élèves, peuvent être menés par les équipes pédagogiques, en complément du renforcement en mathématiques, en s'appuyant sur les ressources développées par l'Office national d'information sur les enseignements et les professions (ONISEP) ou sur les associations partenaires de l'école.

De plus, depuis septembre 2018, la mission mathématiques et la DGESCO ont permis aux clubs, en lien avec les mathématiques au sein des établissements scolaires, de gagner en visibilité. L'objectif de cette première démarche a été d'identifier ces clubs afin de les valoriser, de favoriser leur pérennisation et d'en faire de véritables leviers dans les politiques éducatives des académies. Un des intérêts de la création des clubs de mathématiques est, également, de développer la place des filles dans les cursus et les carrières scientifiques grâce à la lutte contre les stéréotypes de sexe. Les clubs de mathématiques sont proposés aux élèves

et peuvent devenir des lieux privilégiés de mixité, vecteurs d'égalité, qui plus est en promouvant une activité dans laquelle filles et garçons concourent équitablement sur le plan de la réflexion et du raisonnement, sans considérations de sexe. Il permet, ainsi, de nuancer, d'interroger, et de combattre, dans la pratique, les stéréotypes de genre.

Par ailleurs, il convient d'évoquer le fait que, pour l'année scolaire 2022-2023, l'enseignement des mathématiques pourra être suivi, à titre facultatif, par ceux de nos élèves qui n'ont pas choisi la spécialité mathématiques. Ainsi, l'option 1 h 30 leur sera offerte, dès la rentrée 2022, dans les enseignements communs de première générale, ce qui leur ouvrira les portes, le cas échéant, de l'option mathématiques complémentaires en classe de terminale. Des dispositions définitives seront mises en place à la rentrée 2023, dans l'objectif de renforcer la place des mathématiques au lycée général et d'y assurer un enseignement pour tous.

Enfin, des Assises nationales des mathématiques, en 2022-2023, seront envisagées afin de mener une réflexion plus exhaustive sur l'enseignement des mathématiques dans le second degré ».

Actions pour réduire les disparités de genre constatées

Dans sa réponse à la Cour sur les actions envisagées pour réduire les disparités de genre constatées dès le lycée pour la spécialité NSI et, dans une moindre mesure, en mathématiques, la DGESCO indique que « [...] la lutte contre les inégalités et l'accompagnement à l'orientation doivent s'inscrire dans un projet global à tous les niveaux du système éducatif. Ils ne peuvent être considérés comme des actions à mener en plus des missions d'enseignement ou comme des actions ponctuelles isolées. C'est dans cet esprit, qu'à l'issue du rapport « Faire de l'égalité filles-garçons une nouvelle étape dans la mise en œuvre du lycée du XXI^e siècle » (S. Béjean, C. Roiron et J C Ringard), l'objectif national cible de 30 % au moins de mixité d'ici cinq ans dans les enseignements de spécialité de la voie générale, les séries technologiques et les filières post-baccalauréat a été fixé pour les établissements et, d'ici 2028, au moins 40 % dans le secteur des sciences et du numérique. La question de la mixité dans les formations professionnelles fait également l'objet d'attention afin que les choix ne soient plus déterminés par le genre et que tous les domaines d'activité puissent être occupés par les filles et les garçons. La promotion de l'égalité entre les filles et les garçons et la prévention des stéréotypes sociaux ou de genre sont au cœur de ces démarches partagées avec les régions par le biais de leurs compétences pour l'information sur les métiers et les formations. Tout au long du parcours des élèves, ces démarches s'inscrivent dans une approche proactive de développement de compétences à s'orienter et de déconstruction des représentations genrées

des métiers qui permettent à l'élève de construire des choix ambitieux, diversifiés et éclairés, et de devenir acteur de son parcours de formation.

Le soutien à l'ambition des filles à suivre des enseignements scientifiques, notamment informatiques et mathématiques, fait déjà l'objet d'un travail partenarial, en complément de l'action pédagogique des enseignants et des personnels, notamment d'orientation. De nombreuses associations, dont l'objet est d'œuvrer à renforcer la mixité dans le champ du numérique et de l'informatique, interviennent selon des modalités variables auprès du public scolaire : séances de sensibilisation dans les classes (témoignages d'étudiantes ou de professionnelles, participation à des temps dédiés à l'orientation) ; proposition d'actions éducatives (concours, activités hors temps scolaire) ; tutorat et mentorat.

Aujourd'hui, en sus des associations spécialisées dans le numérique et l'informatique, toutes les associations généralistes œuvrant pour la mixité et l'égalité dans les sciences se mobilisent sur la question du numérique et de l'informatique, pour mettre en lumière les filières et l'opportunité qu'offrent ces secteurs (« Femmes et mathématiques », « Femmes et sciences », « Femmes ingénieures », « Elles bougent » pour en citer quelques-unes). Le ministère soutient également l'association « Becomtech », qui a développé, depuis quelques années, le programme « Jump in Tech », stage intensif d'été ouvert à des filles, ainsi que la structuration d'une communauté d'ambadrices, accompagnée dans leur ambition de poursuivre des études dans l'informatique et le numérique. Enfin, le ministère soutient la fondation « Femmes@numérique », qui accompagne, par exemple, le déploiement du programme « Numérique pour elles » en Occitanie.

L'ensemble de ces actions ciblées s'inscrit dans une démarche visant à bâtir, dès l'école primaire et tout au long du collège, une confiance des filles et des garçons dans leurs compétences. La mise en œuvre d'une pédagogie égalitaire en est une condition et s'acquiert par le biais d'une solide formation, engagée dès l'apprentissage du métier et tout au long de la carrière. À cet égard, la mise en œuvre du cahier des charges pour un continuum de formation à l'égalité, à la fois dans les Instituts nationaux supérieurs du professorat et de l'éducation (Inspé) et dans les plans académiques de formation (PAF), est un enjeu pour les prochaines années.

Par ailleurs, la labellisation « Égalité filles-garçons » des collèges et des lycées, nouveau dispositif visant à intégrer la politique d'égalité à la fois dans le pilotage de l'établissement, dans la formation des personnels, dans l'action pédagogique et éducative, et enfin dans la structuration de partenariats avec le monde associatif et professionnel, a

été conçue comme un levier pour faire progresser la mixité et atteindre une orientation moins marquée par des déterminismes de genre.

Enfin, dans le cadre de la mise en place prochaine d'une demi-journée « Avenir » consacrée à la découverte des métiers, l'enjeu est d'informer davantage les collégiens sur la diversité des métiers et du monde professionnel pour plus d'égalité et de mixité. Les actions qui seront mises en œuvre dans ce cadre pourront promouvoir les métiers et les filières du numérique, ainsi que faire connaître la réalité du secteur professionnel numérique ».

II. Filières de formation initiale de l'enseignement supérieur (éléments de cartographie complémentaires du corps de texte)

**Tableau n° 46 : évolution des effectifs étudiants inscrits
en mathématiques (maths) et en informatique (info) en licence,
licence professionnelle et DUT sur la période 2016 à 2020**

Session	Inscrits en info	dont femmes	Inscrits en maths	dont femmes	Inscrits en maths et info	dont femmes	Inscrits	dont femmes
2014	27 661	3 196	9 899	3 142	37 560	6 338	792 378	435 361
2015	29 231	3 431	11 360	3 606	40 591	7 037	814 880	447 563
2016	30 613	3 567	12 973	4 061	43 586	7 628	826 989	454 577
2017	31 490	3 826	13 779	4 398	45 269	8 224	844 382	466 081
2018	32 718	4 009	14 277	4 650	46 995	8 659	868 785	482 686
2019	34 198	4 426	16 952	5 571	51 150	9 997	891 655	501 838
2020	34 239	4 490	16 354	5 390	50 593	9 880	931 401	530 854
Évolution annuelle 2006 à 2020 (en %)	3 %	4 %	8 %	6 %	4 %	5 %	2 %	2 %
Évolution annuelle 2018 à 2020 (en %)	2 %	6 %	7 %	8 %	4 %	7 %	4 %	5 %
Moyenne 2006 à 2020	27 710	3 403	10 873	3 555	38 582	6 958	79 3130	441 635

Source : Méthode sémantique de la Cour des comptes appliquée aux libellés des diplômes et formations ainsi qu'aux spécialités de DUT répertoriés dans les données du MESR relatives aux inscriptions des étudiants²⁰²

²⁰² Cf. <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-principaux-diplomes-et-formations-prepares-etablissements-publics/>

**Tableau n° 47 : évolution des parts de femmes inscrites
en mathématiques (maths) et en informatique (info) en licence,
licence professionnelle et DUT sur la période 2016 à 2020.**

<i>Session</i>	Part des inscrits en info	Part des inscrits en math	Part des inscrits en maths et info	Part des femmes inscrites en info	Part des femmes inscrites en maths	Part des femmes inscrites en maths et info	Part des femmes inscrites	Écart à la moyenne (en points)
2014	3,5	1,2	4,7	11,6	31,7	16,9	54,9	- 38
2015	3,6	1,4	5	11,7	31,7	17,3	54,9	- 37,6
2016	3,7	1,6	5,3	11,7	31,3	17,5	55	- 37,5
2017	3,7	1,6	5,4	12,1	31,9	18,2	55,2	- 37
2018	3,8	1,6	5,4	12,3	32,6	18,4	55,6	- 37,2
2019	3,8	1,9	5,7	12,9	32,9	19,5	56,3	- 36,8
2020	3,7	1,8	5,4	13,1	33	19,5	57	- 37,5
<i>Évolution annuelle 2006 à 2020 (en points)</i>	0,0	0,1	0,1	0,1	- 0,5	0,0	0,0	0,0
<i>Évolution annuelle 2018 à 2020 (en points)</i>	0,0	0,1	0,0	0,4	0,2	0,6	0,7	- 0,1
<i>Moyenne 2006 à 2020 (en points)</i>	3,5	1,3	4,8	12,3	32,9	18,0	55,7	- 37,7

Source : Méthode sémantique de la Cour des comptes appliquée aux libellés des diplômes et formations ainsi qu'aux spécialités de DUT répertoriés dans les données du MESR relatives aux inscriptions des étudiants²⁰³.

²⁰³ Cf. <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-principaux-diplomes-et-formations-prepares-etablissements-publics/>

Annexe n° 11 : cartographie de la formation continue en IA et du profil de ses apprenants avec une focale sur l'enseignement supérieur public

Analyse de l'offre de formation continue en IA par la Cour

L'offre de formation continue en IA a été analysée à travers les données relatives à 241 727 formations certifiantes proposées dans le cadre de la plateforme « Mon Compte Formation »²⁰⁴ et mises à disposition par la Caisse des dépôts et consignations. Depuis la fin 2019, cette plateforme permet aux salariés et aux personnes sans emploi de se former à travers la mobilisation des crédits de leur compte personnel de formation (CPF). Les formations certifiantes sont déposées par des organismes de formation et sont inscrites à l'un des deux répertoires nationaux administrés par France Compétences le répertoire spécifique (RS) ou le répertoire national des certifications professionnelles (RNCP).

Dans « Mon Compte Formation », les formations sont caractérisées par des codes, dont un se rapporte spécifiquement à l'IA²⁰⁵. 70 formations continues peuvent ainsi être identifiées dans le domaine de l'IA²⁰⁶. L'emploi de la méthode sémantique développée dans le cadre de l'évaluation permet d'identifier 191 formations en IA²⁰⁷, dont près de 80 % étaient déjà identifiées dans le cadre de la codification spécifique²⁰⁸.

Si les formations dans le domaine de l'informatique et des données²⁰⁹ représentent environ 16 % du catalogue de « Mon Compte Formation », les formations continue en IA représentent moins de 0,1 % de l'offre de formation continue proposée tous secteurs confondus.

²⁰⁴ <https://www.moncompteformation.gouv.fr/>

²⁰⁵ Le code employé pour "intelligence artificielle" est 31028 dans le cadre du thésaurus "Formacode" édité par le Centre Inffo, une association sous tutelle du ministère en charge de la formation professionnelle.

²⁰⁶ L'assiette de l'analyse est donnée par l'ensemble des codes formacode classés de 1 à 5 pour l'ensemble des formations.

²⁰⁷ Détection de 200 formations dont neuf faux positifs retraités manuellement, soit 191 formations au total. Les faux positifs sont liés au domaine du traitement de l'image ou à des formations qui utilisent des IA comme medium en appui à la formation mais qui ne sont pas des formations à l'IA.

²⁰⁸ Sur les 70 formations continues identifiées par la codification explicite, 55, soit près de 80 %, sont retrouvées par la méthode sémantique. Les 15 formations qui ne sont pas retrouvées par la méthode sémantique ne mentionnent pas des termes d'IA.

²⁰⁹ « Informatique, traitement de l'information, réseaux de transmission des données ».

En croisant les données de « Mon Compte Formation » avec les annuaires des établissements du MESR et du MENJS²¹⁰, il apparaît que les formations continues en IA sont pour l'essentiel portées par des entités en dehors des acteurs « traditionnels » de l'enseignement (entre 75 % et 85 % pour le secteur privé ; 15 % à 25 %²¹¹ par des établissements d'enseignement supérieur ; 0 % par des établissements d'enseignement scolaire).

Au niveau national, seulement 26 établissements d'enseignement supérieur²¹² proposent des formations continues en IA. Les cinq établissements d'enseignement supérieur les plus actifs dans ce domaine sont l'Université Paris Dauphine (six formations), l'ISAE-SUPAERO (quatre formations), le Groupe des écoles nationales d'économie et de statistique, l'Université Lyon 2 et l'Université d'Aix Marseille (trois formations pour chaque établissement).

Analyse de la demande en lien avec la Caisse des dépôts et consignations

À partir des formations en IA identifiées par la Cour, la Caisse des dépôts a pu retracer la demande correspondante en mesurant l'évolution du nombre de dossiers de demande validés par les usagers. Dans la mesure où une formation peut être annulée par un organisme de formation ou un apprenant, il s'agit d'une estimation haute de la demande car tous les dossiers validés ne donneront pas nécessairement lieu à une formation effective.

Au global, secteurs publics et privés confondus, il apparaît que moins de mille personnes ont été formées à l'IA depuis le 1^{er} janvier 2020, un volume représentant moins de 0,1 % de la demande de formation continue proposée tous secteurs confondus²¹³.

Les établissements d'enseignement supérieur publics (ESRP) ont représenté seulement 11,6 % des dossiers de formation validés, soit au plus une centaine de personnes effectivement formées à l'IA.

²¹⁰ Open data des deux ministères.

²¹¹ L'annuaire des établissements d'enseignement maintenu par le MESR n'étant pas totalement exhaustif, une correction manuelle a été opérée pour certains opérateurs de formation qui peuvent être assimilés à des acteurs de l'enseignement supérieur tels que les écoles suivies par ministère chargé de l'économie ou certaines écoles de commerce.

²¹² Seul 26 établissements d'enseignement supérieur (ou assimilés) proposent des formations continues dans ce domaine : 17 établissements dans l'annuaire du MESR (sur un total de 254 ; soit 7 %) et neuf autres établissements d'enseignement supérieur.

²¹³ S'agissant de l'IA, 962 dossiers ont été validés depuis janvier 2020, avec un taux d'annulation de 9 %, soit un total de 877 dossiers hors annulation. Ces nombres sont à rapporter à un peu plus de 1 M de dossiers validés depuis janvier 2020 sur la même période

(source : https://opendata.caissedesdepots.fr/explore/dataset/moncompteformation_formations_engagees/information/).

Les apprenants concernés par les formations en IA dispensées par les établissements de l'ESRP sont pour l'essentiel des hommes (un quart de femmes seulement) et des personnes jeunes et d'âge intermédiaire (37 % de 26-35 ans et 42 % de 36-45 ans), avec une faible part de demandeurs d'emploi (11 %). S'agissant des modalités de présence aux formations, aucune n'est proposée intégralement à distance (43 % des apprenants sont formés sur place et 57 % en mixte).

Si les effectifs d'apprenants formés en IA par le secteur privé tendent à s'accroître considérablement depuis 2021, la contribution du secteur public à la formation continue en IA stagne (voir tableau dans le corps de texte sur la formation continue en IA).

Annexe n° 12 : évolutions des effectifs de formateurs publics dans les secteurs de l'enseignement supérieur où l'IA est fortement représentée

Données analysées

D'après la consultation menée par la Cour, les sections CNU où l'IA prévaut sont l'informatique (section n° 27), les mathématiques appliquées et applications des mathématiques (n° 26) ainsi que le génie informatique, automatique et traitement du signal (n° 61).

Le MESR propose des données en séries longues relatives aux effectifs nationaux d'enseignants-chercheurs et d'enseignants titulaires avec le détail des sections CNU concernées²¹⁴.

La Cour a donc retraité les données du MESR de manière à suivre l'évolution des profils des formateurs publics dans les domaines où les experts de l'IA sont sur-représentés.

Principaux résultats

Tableau n° 48 : évolutions des effectifs d'enseignants-chercheurs permanents détaillés par catégorie dans les principaux secteurs scientifiques où l'IA est représentée.

<i>Rentrée</i>	Effectif des personnels enseignants	Effectif des personnels enseignants dans les secteurs où l'IA est représentée	dont effectif des maîtres de conférences	dont effectif des professeurs d'université	dont effectif des autres enseignants dans le supérieur	Part des personnels enseignants dans les secteurs où l'IA est représentée (en %)	Part des maîtres de conférences dans ces secteurs (en %)
2010	69 131	6 804	4 759	2 045	0	9,8	69,9
2011	69 585	6 870	4 793	2 077	0	9,9	69,8
2012	69 669	6 925	4 814	2 111	0	9,9	69,5
2013	69 977	7 001	4 851	2 150	0	10	69,3
2014	70 019	7 039	4 879	2 160	0	10,1	69,3
2015	69 657	7 193	4 817	2 171	205	10,3	67
2016	69 670	7 216	4 838	2 188	190	10,4	67

²¹⁴ Voir <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-enseignants-titulaires-esr-public-national/information/> (dernière mise à jour en février 2022).

Rentrée	Effectif des personnels enseignants	Effectif des personnels enseignants dans les secteurs où l'IA est représentée	dont effectif des maîtres de conférences	dont effectif des professeurs d'université	dont effectif des autres enseignants dans le supérieur	Part des personnels enseignants dans les secteurs où l'IA est représentée (en %)	Part des maîtres de conférences dans ces secteurs (en %)
2017	68 547	7 079	4 756	2 145	178	10,3	67,2
2018	68 617	7 030	4 724	2 154	152	10,2	67,2
2019	68 624	7 061	4 738	2 171	152	10,3	67,1
2020	68 484	7 032	4 714	2 165	153	10,3	67
Moyenne 2010 à 2020	69 271	7 023	4 789	2 140	94	10	68
Évolution 2010 à 2020	- 1 %	3 %	- 1 %	6 %		0,5	- 2,9
Évolution 2010 à 2019	- 1 %	4 %	0 %	6 %		0,5	- 2,8
Évolution 2018 à 2020	0 %	0 %	0 %	1 %	1 %	0,1	- 0,2

Source : Cour des comptes d'après les données du MESR relatives aux effectifs nationaux d'enseignants-chercheurs et d'enseignants titulaires. Les sections CNU prises en considération sont l'informatique (section n° 27), les mathématiques appliquées et applications des mathématiques (n° 26) ainsi que le génie informatique, automatique et traitement du signal (n° 61)

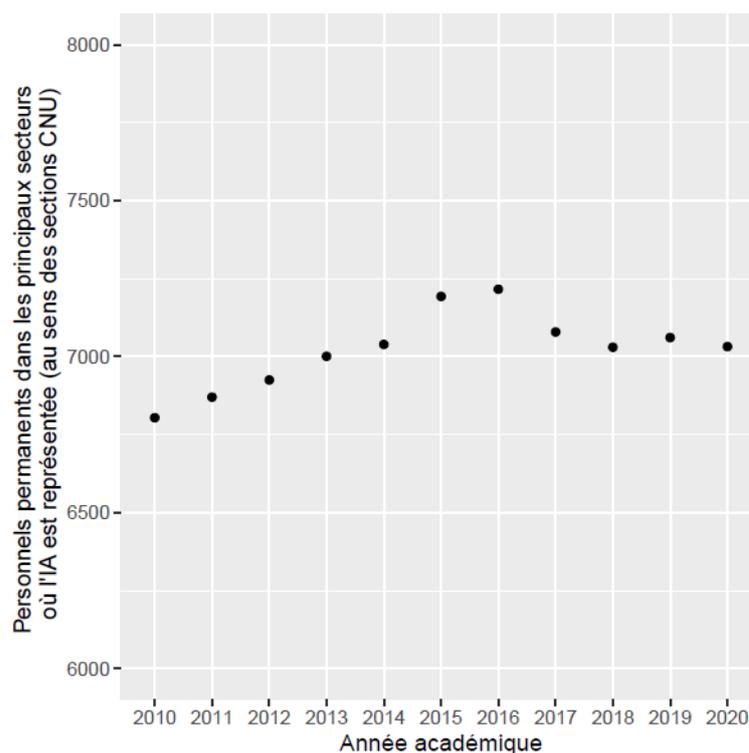
Tableau n° 49 : évolutions des effectifs d'enseignants-chercheurs permanents détaillés par genre et par nationalité dans les principaux secteurs scientifiques où l'IA est représentée

Rentrée	Effectif des personnels enseignants dans les secteurs où l'IA est représentée	dont effectifs des femmes	dont effectifs des français	dont effectifs des européens non-français	dont effectifs des extra-européens	dont effectifs des non-français (sans précision)	Part des femmes dans les secteurs où l'IA est représentée (en %)	Part des français dans ces secteurs (en %)
2010	6 804	1 558	5 765	0	0	1 039	22,9	84,7
2011	6 870	1 576	5 810	0	0	1 060	22,9	84,6
2012	6 925	1 587	5 836	0	0	1 089	22,9	84,3
2013	7 001	1 599	5 874	0	0	1 127	22,8	83,9
2014	7 039	1 613	5 885	0	0	1 154	22,9	83,6
2015	7 193	1 642	6 041	0	0	1 152	22,8	84
2016	7 216	1 660	6 054	399	763	0	23	83,9
2017	7 079	1 640	6 396	326	356	0	23,2	90,4
2018	7 030	1 615	6 371	324	329	0	23	90,6

<i>Rentrée</i>	Effectif des personnels enseignants dans les secteurs où l'IA est représentée	dont effectifs des femmes	dont effectifs des français	dont effectifs des européens non-français	dont effectifs des extra-européens	dont effectifs des non-français (sans précision)	Part des femmes dans les secteurs où l'IA est représentée (en %)	Part des français dans ces secteurs (en %)
2019	7 061	1 631	6 402	333	322	0	23,1	90,7
2020	7 032	1 640	6 330	331	331	0	23,3	90
<i>Moyenne 2010 à 2020</i>	7 023	1 615	6 069	156	191	602	23	86
<i>Évolution 2010 à 2020 (en % ou en points)</i>	3 %	5 %	10 %			- 100 %		0,4
<i>Évolution 2010 à 2019 (en % ou en points)</i>	4 %	5 %	11 %			- 100 %		0,2
<i>Évolution 2018 à 2020 (depuis la SNIA ; en % ou en points)</i>	0 %	2 %	- 1 %	2 %	1 %		1	0,3

Source : Cour des comptes d'après les données du MESR relatives aux effectifs nationaux d'enseignants-chercheurs et d'enseignants titulaires. Les sections CNU prises en considération sont l'informatique (section n° 27), les mathématiques appliquées et applications des mathématiques (n° 26) ainsi que le génie informatique, automatique et traitement du signal (n° 61). Les effectifs pour lesquelles la nationalité n'est pas connue sont limités (40 en 2000, 4 en 2019, 6 en 2018, 1 en 2017 et 0 pour les années antérieures)

Graphique n° 51 : évolutions des effectifs d'enseignants-chercheurs permanents dans les principaux secteurs scientifiques où l'IA est représentée



Source : Cour des comptes d'après les données du MESR relatives aux effectifs nationaux d'enseignants-chercheurs et d'enseignants titulaires. Les sections CNU prises en considération sont l'informatique (section n° 27), les mathématiques appliquées et applications des mathématiques (n° 26) ainsi que le génie informatique, automatique et traitement du signal (n° 61). Ces secteurs représentent 80 % des répondants à la consultation de la Cour auprès de la communauté scientifique en IA

Annexe n° 13 : bilan du financement des projets relatifs aux formations en IA dans le cadre de l'AMI CMA

**Tableau n° 50 : positionnement des projets en IA dans le cadre
de l'AMI CMA**

		Dossiers en IA	Ensemble des dossiers	Part des dossiers en IA
<i>Levée 1 (février 2022)</i>	nombre de dossiers déposés	9	50	18,0 %
	Nombre de dossiers lauréats	7	29	24,1 %
	Taux de sélection en nombre	78 %	58 %	
	Montants totaux demandés (en M€)	88	386,6	22,8 %
	Montants totaux alloués (en M€)	54,5	255,4	21,3 %
	Taux de sélection en montants	61,9 %	66 1 %	
<i>Levée 2 (juillet 2022)</i>	Nombre de dossiers déposés	8	105	7,6 %

Sources : Cour des comptes d'après les réponses de l'Agence nationale de la recherche. Il est à noter que ce tableau indique uniquement les projets déposés dans le cadre stricte de l'AMI CMA. Il existe opérationnellement une convergence avec les lauréats de Deffinum, qui permet de financer des projets de Dispositifs de Formation Innovante fondés sur le NUMérique

Annexe n° 14 : compte-rendu de l'atelier des parties prenantes du 19 juillet 2022

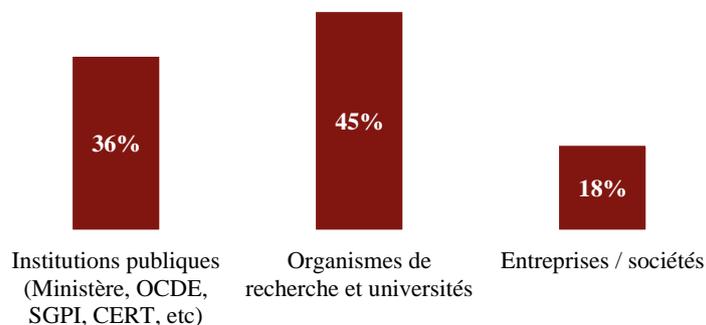
Méthodologie - Animation d'un panel de parties prenantes de l'écosystème de l'IA en France

Le 19 juillet 2022, l'Ifop a animé, pendant une demi-journée dans les locaux de la Cour des Comptes, un panel de 24 représentants de l'écosystème de la recherche en IA repartis de la façon suivante.

Ces derniers ont été invités – dans l'un des trois groupes d'experts constitué par l'Ifop et la Cour – à s'exprimer en détail sur deux des cinq grands axes de recommandation de la Cour des comptes, avant de présenter une synthèse de leurs échanges lors d'une réunion plénière en présence du groupe de contrôle de la Cour.

Ce dispositif a été conçu pour répondre aux objectifs suivants :

Graphique n° 52 : répartition des participants par institutions



Source : Cour des comptes

- échanger autour des principaux enjeux de l'écosystème de la recherche en IA à horizon 2030, des forces et faiblesses de la stratégie nationale de recherche, et mettre en débat des constats la Cour des Comptes ;
- aborder la question du positionnement de la France en matière d'IA, notamment en ce qui concerne ses partis-pris éthiques (IA de confiance, etc.) ;
- recueillir le point de vue des experts au sujet des recommandations de la Cour des comptes, obtenir des éléments d'optimisation et d'opérationnalisation des grands axes présentés aux experts.

Une nécessaire refonte du mode de gouvernance de l'écosystème

Confrontés à la question de la gouvernance de l'écosystème de l'IA en France, les experts évoquent un enjeu principal : la nécessité de faire émerger une vision stratégique commune entre des acteurs très différents, d'où l'intérêt d'imaginer un mode de gouvernance plus ouvert et plus intégré qu'aujourd'hui.

Dans cet objectif, les experts formulent plusieurs recommandations :

- la simplification d'une architecture de gouvernance aujourd'hui jugée trop complexe et peu lisible, avec une clarification des rôles et des périmètres d'action attribués à chacun ;
- la recherche d'opérationnalité, avec des acteurs décisionnaires en prise avec les réalités du terrain, qu'il s'agisse du monde de la recherche ou de l'industrie, et susceptibles de participer au déploiement effectif de la stratégie nationale ;
- une évolution vers un schéma plus intégré pour obtenir un vrai « effet écosystème ». Concrètement, il s'agit surtout de faciliter l'échange d'informations et la collaboration de différents types d'acteurs – publics et privés, orientés vers la formation, la recherche ou l'industrialisation, etc. – pouvant avoir tendance à fonctionner en silo ;
- l'élargissement du périmètre de l'écosystème en intégrant davantage l'échelon européen et en identifiant des partenaires extra-européens compatibles avec les valeurs de la France dans le domaine de l'IA ;
- la mise en place d'indicateurs clairs, précis, faciles à suivre dans le temps mais aussi plus complets que les outils actuels. Se pose par exemple la question de développer des indicateurs européens, notamment en ce qui concerne le co-financement, de se doter d'indicateurs permettant d'intégrer la création de valeur et l'impact des financements. Sur ce dernier point, plusieurs experts suggèrent d'introduire une distinction entre les projets orientés « cœur IA » et les projets plus « applicatifs » pour permettre à toute la chaîne de bénéficier des financements²¹⁵.

²¹⁵ Toutefois, la construction de nouveaux indicateurs paraît particulièrement délicate en matière d'IA, avec une difficulté propre au secteur dans la mesure où il s'agit de technologies extrêmement « diffuses » et embarquées dans des projets hybrides très divers. Il semble dès lors difficile d'imaginer un « seuil technologique objectif » à partir duquel un projet doit être intégré dans les indicateurs.

L'impératif d'une révision des instruments et des logiques de financement

Concernant les enjeux de financements, les représentants de l'écosystème évoquent prioritairement la question des montants alloués, avec des niveaux qui demeurent, en dépit des efforts consentis, très largement inférieurs aux ambitions que s'est donnée la France dans le domaine de l'IA. Deux pistes paraissent intéressantes pour envisager des améliorations à ce niveau : le développement de financements européens, aujourd'hui marginaux, et surtout des financements privés, actuellement en-deçà de leur plein potentiel.

En outre, au-delà de la question des montants, s'exprime aussi l'idée de faire évoluer les outils de financement du secteur pour les adapter à ses contraintes.

Plusieurs pistes sont proposées par les experts :

- la simplification des dispositifs de financement aujourd'hui jugés trop complexes en raison notamment de l'imbrication des différents échelons (France, UE), et plus largement d'un manque de lisibilité (concernant par exemple les contraintes de statut des demandeurs, la possibilité de cumuler plusieurs dispositifs, etc.) ;
- la création d'outils de financement plus flexibles pour permettre aux projets d'évoluer dans des directions qui n'étaient pas anticipées à l'origine, de laisser certaine liberté d'ajustement aux porteurs de projet²¹⁶ ;
- de nouveaux leviers de financement plus à même de faire fusionner les approches fondamentales (notamment les PEPR) avec des projets plus applicatifs dans l'objectif de fluidifier les transferts technologiques vers l'industrie ;
- la mise en place d'instruments capables de supporter une part de risque avec des projets exploratoires ou atypiques, sortant des feuilles de route traditionnelles (à l'instar du projet « *Big science* »).

En outre, ces différents dispositifs de financement doivent, dans la mesure du possible, respecter un principe de continuité, avec l'idée parfois évoquée de « sanctuariser » des montants minimaux dans le but d'envoyer un signal positif et cohérent aux acteurs de l'écosystème.

²¹⁶ Sur ce point, le PEPR est cité en exemple puisqu'il offre une certaine liberté d'ajustement et permet de financer des projets qui n'avaient pas été anticipés.

Mieux répondre aux enjeux actuels et futurs de la formation

Du point de vue des experts, les enjeux de formation nécessitent une réaction urgente des pouvoirs publics pour faire face à la pénurie de professionnels et de chercheurs dans le domaine l'IA. Toutefois, l'impératif de « massification » de la formation pose la question de la contribution respectives des différents acteurs du domaine – ONR, centres d'excellence et établissements supérieurs de recherche –, avec l'idée qu'il pourrait être contre-productif dans certains cas de contraindre des chercheurs de haut-niveau à réaliser des actions de formation (avec des risques pour l'attractivité des postes, et surtout en termes de coût d'opportunité).

Sur ce point, plusieurs axes de réflexion sont proposés par les experts :

- créer des ponts entre la recherche en laboratoire et la recherche appliquée en entreprise, en apportant notamment plus de flexibilité dans les conditions de mise en disponibilité des chercheurs ;
- davantage intégrer le regard des industriels dans les choix stratégiques des acteurs de la formation en IA pour mieux répondre aux besoins du secteur ;
- encourager la transdisciplinarité avec d'autres domaines, comme les sciences cognitives, pour attirer de nouveaux profils d'étudiants, et en particulier les profils féminins aujourd'hui largement sous-représentés dans le domaine ;
- inscrire les formations en IA dans la durée pour répondre à la problématique des formations non-pérennes. En effet, aux yeux des experts rencontrés, beaucoup de nouvelles formations voient le jour, mais trop peu se développent suffisamment pour atteindre une certaine masse critique et faire référence dans le domaine ;
- clarifier la mission des centres d'excellence et leur durée dans le temps pour envisager des « relais » à moyen terme ;
- encourager les actions de formation dès le plus jeune âge, à la fois pour susciter des vocations, mais aussi pour dédramatiser le fonctionnement de l'IA et favoriser son acceptabilité sociale ;
- valoriser la stratégie et renforcer l'attractivité de l'écosystème de la recherche en France.

Le travail de mise en valeur de l'écosystème français, et plus spécifiquement son attractivité auprès des chercheurs et des entreprises, est envisagé par tous les acteurs rencontrés comme une variable clé de succès de la stratégie nationale. Cette thématique soulève néanmoins la question des arbitrages à réaliser en termes de communication et d'investissement,

avec l'idée qu'il peut être judicieux de prioriser une petite quantité d'acteurs comme les 3IA pour faciliter leur identification, notamment à l'étranger, mais avec le risque de d'invisibiliser les autres structures.

Pour répondre à cet impératif de mise en visibilité de l'écosystème français, plusieurs pistes sont envisagées :

- appliquer des correctifs aux éléments pouvant bloquer les ambitions des jeunes chercheurs ou dégrader l'attractivité de la France dans le domaine de la recherche. Sur ce point, les représentants de l'écosystème évoquent notamment le manque de visibilité des possibilités de financements, la rigidité administrative qu'implique le statut d'enseignant-chercheur et la faiblesse des rémunérations ;
- mettre en place une plateforme nationale centralisant les offres de d'emploi dans le domaine (mais aussi, les possibilités de financement pour la recherche) ;
- promouvoir des « thématiques d'excellence » par les pouvoirs publics dans le but de positionner la France et de mobiliser des filières ;
- la mise en place d'un organisme « référent » en la matière bien identifié en France et à l'étranger pour centraliser l'information et la communication.

Annexe n° 15 : synthèse des travaux de parangonnage international

La Cour, en s'appuyant sur les données fournies par les services scientifiques des ambassades françaises, a examiné les politiques publiques menées en matière d'IA dans le monde et plus particulièrement au sein de dix pays : l'Allemagne, le Canada, les États-Unis, la Finlande, Israël, l'Italie, les Pays-Bas, le Japon, le Royaume-Uni, et la Suisse. La Cour a aussi examiné les différents volets de l'action de l'Union européenne.

Il en ressort que les objectifs poursuivis par la stratégie française sont similaires à ceux mis en œuvre dans la plupart des plans nationaux. Il est possible d'identifier sept priorités poursuivies recensées dans le tableau suivant, qui présente également des exemples d'actions menées dans certains pays étudiés.

Tableau n° 51 : éléments de comparaison des politiques publiques menées en matière d'IA

<i>Priorités poursuivies</i>	Politiques publiques en IA des pays étudiés par la Cour
<i>Le renforcement d'un nombre ciblé de centres de compétences transdisciplinaires pour la recherche en IA</i>	<ul style="list-style-type: none"> - Au Canada, trois centres de recherche et innovation en IA ont été constitués dès 2017 par l'Institut canadien de recherches avancées (CIFAR) chargé de mettre en œuvre la stratégie canadienne, dans les trois grands centres canadiens de l'intelligence artificielle : l'<i>Alberta Machine Intelligence Institute</i> (Amii) à Alberta, le MILA à Montréal et l'Institut vecteur à Toronto ; - Aux États-Unis, sept "AI institutes" ont été créés en 2020 et 11 en 2021, répartis dans des universités situées dans différents États, avec des financements alloués par la <i>National Science Foundation</i> ; - En Allemagne, outre le Centre allemand de recherche en intelligence artificielle (DFKI), le Ministère fédéral de l'éducation et de la recherche soutient cinq autres centres de compétence pour la recherche en intelligence artificielle, situés à Berlin (BIFOLD), Dresde/Leipzig (ScaDS.AI), Dortmund/St. Augustin (ML2R), Munich (MCML) et Tübingen (Tübingen AI Center) ; - Aux Royaume-Uni, l'<i>Alan Turing Institute</i> fondé en 2015 à l'initiative du gouvernement britannique est le principal centre d'excellence de la recherche dans ce domaine ; - Au Japon, la stratégie a fixé comme objectif la structuration d'un réseau rassemblant les institutions engagées dans la R&D en IA (universités et instituts de recherche). Les trois instituts composant son cœur sont : le <i>RIKEN Center for Advanced Intelligence Project</i>, le <i>National Institute of Advanced Industrial Science and Technology center</i> et le <i>National Institute of Information and Communications Technology AI center</i> ; - En Finlande, le principal centre sur lequel s'appuie la stratégie est le Centre finlandais d'intelligence artificielle (FCAI) créé par l'Université

<i>Priorités poursuivies</i>	Politiques publiques en IA des pays étudiés par la Cour
	d'Aalto et l'Université d'Helsinki et réunissant des universités, des acteurs industriels et des acteurs publics.
<i>La création de chaires</i>	Le financement de chaires en IA est en particulier un levier essentiel de la stratégie canadienne mise en œuvre par le CIFAR pour attirer et retenir des chercheurs de renommée mondiale. En 2018 et 2019, 80 chaires ont ainsi été créées en partenariat avec les trois instituts canadiens en IA, ainsi que plusieurs universités et hôpitaux. En 2018, l'Allemagne s'est fixée un objectif de création de 100 chaires en IA, et a annoncé en mai 2022 que cet objectif était atteint.
<i>Le développement la formation</i>	Les politiques publiques en IA étudiées par la Cour visent notamment à créer de nouvelles écoles doctorales pour ensuite développer la formation à tous les niveaux et dans une approche pluridisciplinaire.
<i>La croissance des capacités de calcul</i>	Le développement des capacités de calcul est également identifié comme un élément prioritaire de l'ensemble des stratégies étudiées par la Cour. Les États-Unis disposent des capacités de calculs les plus importantes : le <i>Summit scientific supercomputer</i> du laboratoire national Oak Ridge (DoE) et le <i>Frontier supercomputer</i> font ainsi partie des plus puissants ordinateurs conçus pour accélérer l'innovation en IA. Frontera (NSF) est un système HPC pour le monde académique. Au sein de l'UE, le premier super calculateur pré-exaflopique développé par EuroHPC a été inauguré en Finlande.
<i>Le renforcement de l'attractivité pour les chercheurs et les formateurs</i>	Différents leviers des stratégies nationales participent à cette priorité transversale, notamment les centres interdisciplinaires, les programmes de chaires et le renforcement des capacités de calcul. Quatre facteurs d'attractivité sont principalement identifiés : la durabilité de l'emploi, sa qualité, la rémunération et les conditions générales de vie dans le pays. L'étude de la Cour a permis de mettre en évidence des politiques volontaristes menées dans certains pays. Le programme de chaire canadien prévoit ainsi que la moitié des chaires revienne à des chercheurs venant de l'étranger. En Finlande, quatre initiatives ont plus particulièrement été mises en place : une procédure accélérée pour l'obtention des permis de travail/résidence (i), l'adoption de mesures en faveur du maintien en Finlande des étudiants internationaux, comme l'extension de la validité du permis de séjour (ii), le projet HEI LIFE ²¹⁷ financé par le ministère de l'éducation et le ministère de la culture qui vise à créer des modèles nationaux pour les établissements d'enseignement supérieur finlandais afin de soutenir l'intégration des universitaires et du personnel internationaux et de leurs familles (iii), un projet de recherche auquel participe le ministère des affaires économiques et de l'emploi qui vise à déterminer comment renforcer l'attractivité de la Finlande auprès des experts internationaux et des expatriés afin d'endiguer la pénurie de main-d'œuvre qualifiée qui constitue un obstacle à la croissance dans de nombreux secteurs (iv).
<i>Le développement de la coopération partenariale internationale</i>	Toutes les stratégies nationales encouragent, sous certaines conditions d'excellence, la coopération internationale. Il en résulte deux mouvements : le développement de la coopération bilatérale, et celui de réseaux, à l'initiative d'un ou plusieurs centres de recherche. Ils offrent des

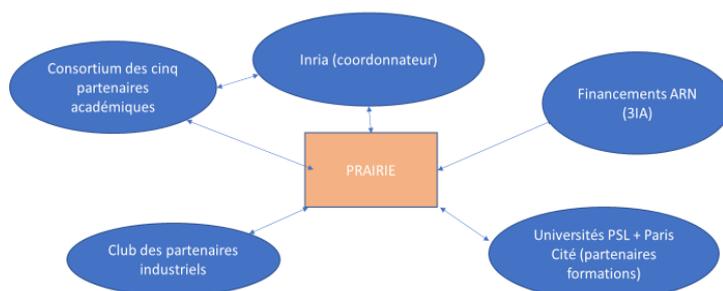
²¹⁷ *Higher Education Institutions' Support for International Academics and Staff - Living, Integration, Family and Employment in Finland.*

<i>Priorités poursuivies</i>	Politiques publiques en IA des pays étudiés par la Cour
	possibilités d'échanges très enrichissantes au bénéfice des chercheurs, mais aussi des étudiants. Le rayonnement d'un centre de compétence se mesure aussi dans la participation à ces réseaux dont les critères d'admission fixés dans leurs chartes.
<i>Le transfert des résultats de la recherche vers le monde économique</i>	<p>Les États-Unis ont une avance importante, et les GAFAMI sont très impliqués dans la diffusion de l'IA.</p> <p>Au Canada, sur le budget de 443,8 M\$ canadiens lié au renouvellement de la Stratégie pancanadienne en matière d'IA, annoncée par le gouvernement en 2021, plus de 40 % des fonds seront consacrés à la commercialisation (applications), soient 185 M\$ pour favoriser la mise sur le marché et la valorisation des technologies et innovations dans l'IA. L'étape du passage du laboratoire au grand public prend le plus souvent la forme de la création d'entreprise.</p> <p>Dans les États fédéraux, les états fédérés investissent également. En Allemagne, Bade-Wurtemberg a élaboré une stratégie numérique interministérielle dès 2017 (<i>Digitalisierungsstrategie – digital@bw</i>) sous l'égide du ministère pour le numérique du Bade-Wurtemberg. Les actions en faveur du transfert de technologie s'appuient sur le renforcement d'un écosystème régional en IA (<i>Cyber Valley</i>) et européen (réseau européen ELLIS) dense, l'implication forte des industriels notamment par l'intermédiaire de fondations (fondation Dieter Schwarz, fondations Hector), la création de quatre Science Data Centers, financés à hauteur de 8 M€.</p> <p>Au Pays-Bas, dans le cadre l'appel à projets de 2021 du Fonds national de croissance néerlandais (FNC), le consortium AINed a obtenu un financement à hauteur de 276 M€ pour réaliser des projets d'intelligence artificielle. L'objectif de ce programme d'investissement est d'apporter une contribution significative à la compétitivité internationale du pays en ce qui concerne le développement et l'application de l'IA dans les domaines suivants : IA embarquée ; les systèmes d'IA hybrides ; infrastructures contrôlées par l'IA ; IA pour la langue néerlandaise ; la personnalisation et la protection de la vie privée ; le partage des données. Le FNC repose sur une logique spécifique, subventionnant 50 % des dépenses induites par les projets, le reste devant émanant d'autres fonds, publics ou privés.</p>

Source : Cour des comptes

Annexe n° 16 : monographie PRAIRIE

Organigramme n° 1 : sociogramme de la gouvernance



Source : Cour des comptes

Tableau n° 52 : tableau récapitulatif des principaux éléments clé 3IA PRAIRIE

Coordination	Inria
Partenaires	Membres académiques : PSL, Université Paris Cité, CNRS, Institut Pasteur, Inria Industriels membres : Amazon, Google, Google deepmind, Facebook,, Nokia Bell labs, Janseen, Pfizer Innovation, Faurecia, Naver Labs, Stellantis, valeo, Vertex, Idemia Partenaires associés : Dxomark, Avatar medical, Kayrros, Implicity, Sonio
Nombre de chercheurs	129
Nombre de chaires	51
Nombre de doctorants	67
Puissance de calcul	Accès au calculateur « Jean Zay »

Source : Cour des comptes d'après données PRAIRIE

Contexte institutionnel et genèse du projet

Le projet PRAIRIE a été amorcé par Inria et les établissements de l'ESR (PSL et Paris Cité) en 2017 lors de la montée des enjeux de recherche en intelligence artificielle et mis en pause du fait des travaux de la mission Villani en 2018. PRAIRIE, coordonné par Inria, est régi par un accord de consortium entre cinq partenaires académiques (Université Paris Sciences & Lettres, CNRS, Inria, Université Paris Cité, Institut Pasteur). Les objectifs du projet sont de contribuer de manière significative, par une recherche fondamentale portées par des établissements partenaires d'excellence, au progrès des connaissances en IA diffusées librement dans la communauté scientifique, mener une recherche tournée vers l'entreprise et les applications, adossée à des partenaires industriels internationaux, contribuer à la formation dans le domaine de l'IA, avoir un impact socio-économique significatif.

Les recherches mettent un fort accent sur les applications de l'IA à la biologie et la médecine avec les porteurs de chaires issus de l'Institut Pasteur, de l'Université Paris Cité et de l'Institut Curie. Des recherches interdisciplinaires ont vu le jour dans le cadre de PRAIRIE entre porteurs de chaires et chercheurs des laboratoires des tutelles académiques, dans le cadre, notamment, de l'application de techniques d'apprentissage automatique et de vision artificielle à l'astronomie et à l'astrophysique. Mais également dans l'application de méthodologies venues de la physique statistique et l'apprentissage machine. Un autre volet concerne la collaboration avec des chercheurs issus des sciences humaines et sociales avec la mise en place d'un groupe de travail « Éthique et IA » et des collaborations avec les chercheurs dans le groupe AIS (*Artificial Intelligence and Society*). Ou encore la chaire « Justice sociale et IA ».

PRAIRIE participe à des formations des universités « Paris Cité » (UPC) concernant la médecine et « Paris Sciences & Lettres » (PSL). La master mathématiques-Vision-Apprentissage (MVA) est sûrement la formation la plus réputée. Un double objectif est affiché : augmenter le vivier d'étudiants sur les fondements de l'IA, en informatique et en mathématiques, diffuser les techniques de l'IA dans l'ensemble des disciplines pouvant bénéficier de l'exploitation de masses de données et des techniques d'apprentissage automatique.

PRAIRIE participe également à la formation continue tout au long de la vie par des masters exécutifs (executive master statistique et big data, executive master intelligence artificielle et sciences des données), la

certification en « Sciences des données » et des Masterclasses en IA, l'IA en médecine et par les outils « Illumens ²¹⁸».

La direction de PRAIRIE considère que la priorité est la qualité reconnue de l'activité de recherche dans le cœur de l'IA. PRAIRIE a un positionnement prudent sur l'éthique et sur la frugalité. Cependant, des travaux prennent en compte ce sujet, comme par exemple en optimisation distribuée pour exploiter au mieux les ressources de communications entre agents, ou en traitement d'image pour limiter le nombre de paramètres qui doivent être appris grâce à l'utilisation de modèles hybrides mêlant modélisation physique et apprentissage profond.

Enjeux institutionnels et pérennisation

La pérennité de l'institut est un sujet qui préoccupe sa direction. Le démarrage a été difficile (notamment en période de confinements lié à la pandémie), mais l'institut est dorénavant en vitesse de croisière. Aujourd'hui, l'inconnue du risque d'arrêt de l'institut empêche de capter des industriels. L'arrêt des financements de l'État aurait des conséquences très négatives sur la dynamique du projet. Les 3IA ne peuvent se passer des crédits d'État aujourd'hui.

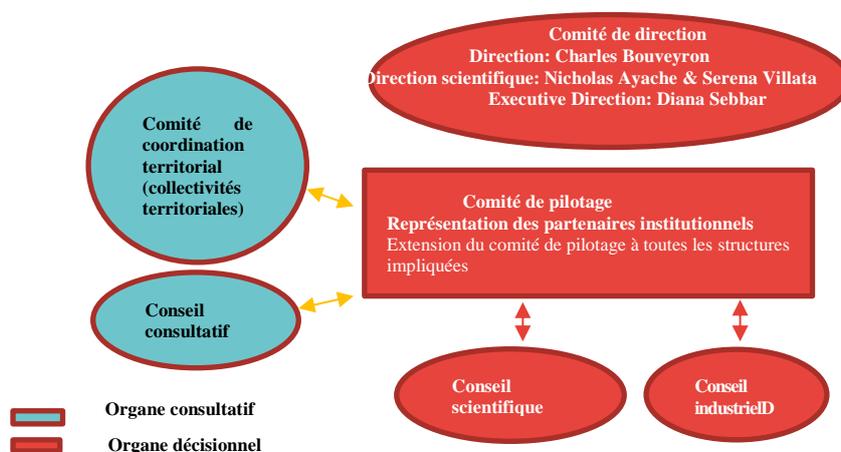
L'absence de personnalité morale de l'institut était perçue comme positive, agile. Cependant, la direction estime que le manque d'autonomie de la structure peut poser des problèmes. Le fonctionnement de PRAIRIE est lié aujourd'hui à la personnalité des cadres.

Au niveau régional, PRAIRIE et les trois instituts IA hors 3IA (DATAIA, SCAI, Hi! PARIS) ont soumis une réponse commune (AI4IDF) à l'appel à propositions « Domaines de Recherche et d'Innovation Majeurs » (DIM) de la région sous l'impulsion d'Inria. Il s'agit pour les quatre instituts de structurer et de stimuler la communauté Île-de-France en IA, en offrant aux partenaires industriels et internationaux une vision unifiée des forces en présence. Le programme s'articulant autour de deux piliers : une composante doctorale et des équipements scientifiques de haut niveau.

²¹⁸ Illumens : formation médicale par la simulation en santé

Annexe n° 17 : monographie 3IA Nice Sophia-Antipolis

Organigramme n° 2 : sociogramme de la gouvernance



Source : Cour des comptes

Tableau n° 53 : tableau récapitulatif des principaux éléments clé du 3IA Nice Sophia-Antipolis

Coordination	Université Côte d'Azur
Partenaires	CNRS, Eurecom, Inria, Inserm, Skema business school
Nombre de chercheurs ²¹⁹	61 chercheurs (titulaires de chaires + post-doctorants recrutés par l'institut)
Nombre de chaires	47 dont trois chaires internationales
Nombre de doctorants	66
Puissance de calcul	Nœud de cluster GPU acheté par le 3IA pour renforcer les capacités de calcul et mis à disposition en priorité pour les chaires 3IA et les pools d'ingénieurs 3IA Tech Pool et possibilité d'accéder au supercalculateur Jean Zay

Source : Cour des comptes d'après les données 3IA Nice Sophia-Antipolis

²¹⁹ Indiquer à la place de cette note les critères d'inclusion.

Activités de l'institut

L'écosystème niçois-Sophia Antipolis-Valbonne, à la suite de l'Idex, présente les caractéristiques favorables à l'implantation d'un centre d'excellente 3IA : une masse critique suffisante d'universités, des centres de recherche de renom, un tissu dense d'entreprises et de start-up et une histoire ancienne de plus de cinquante ans dans le développement de la technologie numérique alimentée par de nombreux partenariats internationaux noués. Le 3IA Nice Côte d'Azur a permis selon les partenaires institutionnels, les entreprises et les collectivités locales entendus, une véritable structuration de l'écosystème autour d'un « guichet unique » en matière d'IA.

La recherche de l'Institut 3IA Côte d'Azur s'organise autour de quatre axes de recherche : IA fondamentale (Axe 1), IA pour la médecine computationnelle (Axe 2), IA pour la biologie computationnelle (Axe 3) et IA pour les territoires intelligents (Axe 4). L'interdisciplinarité se décline, notamment dans les domaines de la santé et de la ville, de la manière suivante : attribution de chaires à six chercheurs ou professeurs dans des disciplines autres que le cœur de l'IA ou de l'IA d'application (en santé ou droit par exemple) de 20 % des bourses doctorales et post-doctorales ; organisation fréquente d'événements scientifiques et rencontres pour décloisonner la science ; création de pôles d'ingénieurs pour initier des projets interdisciplinaires.

Concernant la formation, le 3IA Nice Côte d'Azur propose des programmes pilotes pour « *tester des nouvelles combinaisons de programmes* ». Vingt programmes sont certifiés tels qu'un diplôme universitaire « AI et santé » à destination des personnels de santé ou encore une formation deux fois par an sur cinq ans pour les ingénieurs et les techniciens. Les porteurs de chaires 3IA doivent enseigner 32 h par an, les doctorants 64 h. L'institut 3IA a permis de créer plus de 8 500 h de formation depuis 2019 (+ 1 200 étudiants/an en IA, sur des modules non spécialistes ; +1 100 étudiants/an sur des programmes spécialisés en IA ; + 300 professionnels/an en IA ; + 5 000 citoyens (enfants et adultes) par le biais de conférences, séminaires et ateliers de vulgarisation).

L'institut a permis de rendre plus attractifs des postes permanent au niveau local. Des chercheurs internationaux ont été recrutés pour pouvoir trois chaires internationales (deux Anglais, les Pr. Sebastien Ourselin et Pr. David Wales et un Italien, le Pr. Marco Gori). Leur obligation est d'être présents au moins douze mois à Nice sur les quatre années de leur chaire.

Le 3IA a noué des partenariats à l'international avec le MILA, l'Université Laval (Canada), l'Université d'Utah (États-Unis), l'Université de Santiago (Chili), BFK (Italie), Insight (Irlande), USP (Brésil).

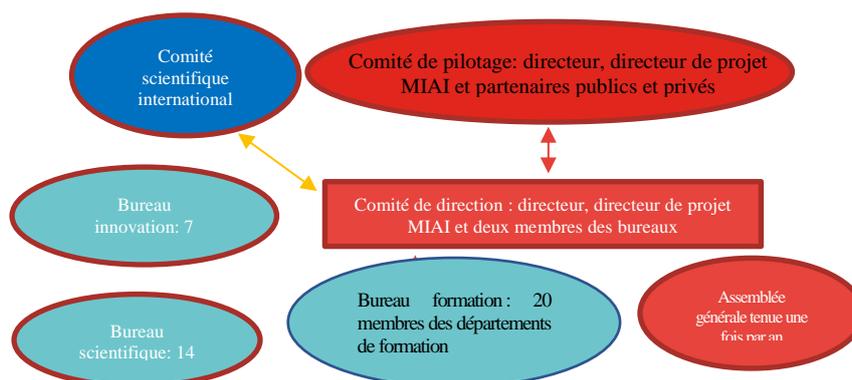
La question de l'éthique est traitée de façon transverse, notamment dans les formations des doctorants proposées par les Écoles doctorales d'Université Côte d'Azur, par les comités d'éthique des établissements partenaires. Il existe une collaboration avec la Chaire UNESCO d'Éthique du Vivant et de l'Artificiel (EVA) portée par le Pr. Vanessa Nurock. Le Pr Serena Villata, titulaire d'une chaire 3IA, est membre du Comité National Éthique du Numérique. L'Université Côte d'Azur a créé l'observatoire OTESIA, un observatoire des impacts de l'IA et un espace d'échange sur les enjeux de l'éthique de l'IA. La frugalité est au cœur du 3IA Côte d'Azur car le 3IA s'intéresse à des problématiques d'IA sous contrainte et applicables dans le monde réel, potentiellement en mode embarqué (satellites, véhicules autonomes).

Enjeux institutionnels et pérennisation

- rôle de guichet et de pilotage unique territorial ;
- pérennité du modèle 3IA ;
- formation et enseignement en IA.

Annexe n° 18 : monographie 3IA MIAI Grenoble

Schéma n° 7 : sociogramme de gouvernance



Source : Cour des comptes d'après les données du MIAI

Tableau n° 54 : tableau récapitulatif des principaux éléments clé de MIAI Grenoble

Coordinateur	Université Grenoble-Alpes
Partenaires	Membres académiques : CEA, CHU Grenoble Alpes, CNRS, Grenoble École de management, INP UGA, Inria, Inserm Partenaires et parties prenantes : Floralis UGA filiale, Fondation INP Grenoble, Institut Carnot LSI et Institut Carnot Inria, Minalogic, Linksium, la région Auvergne-Rhône-Alpes 89 entreprises engagées avec le MIAI pour 150 projets collaboratifs
Nombre de chercheurs	200 chercheurs
Nombre de chaires	30 (dont trois chaires en codirection internationale avec Stanford (USA, la Belgique et Georgia Institute of technology – États-Unis) - entre 220 000 € et 457 000 € alloués par l'ANR par chaire - 4,5 membres académiques par chaire en moyenne
Nombre de doctorants	145 (55 sur fonds ANR et 90 sur fonds privés) et 40 post doctorants et 14 ingénieurs
Puissance de calcul	Possibilité d'accéder au supercalculateur Jean Zay

Source : Cour des comptes d'après les données MIAI

Activités de l'institut

La candidature du 3IA Grenoble a été préparée par un comité d'une dizaine de scientifiques qui a identifié les domaines prioritaires en IA et des échanges avec les partenaires industriels déjà mobilisés dans le cadre de l'Idex fondés sur un historique lié au développement de la technologie numérique et de la micro-électronique. Pour embarquer l'ensemble de la communauté, il a été constitué des 31 chaires fortes de quatre à cinq personnes permanentes, ce qui a été une façon de créer un compromis entre ces deux injonctions. Un travail de fusion pour coupler « cœur » et « application » de l'IA s'est engagé. Les quatre objectifs du MIAI sont de développer une recherche interdisciplinaire en IA, former les étudiants et les professionnels à l'IA et ses applications, soutenir l'innovation en IA au profit des grands groupes, TPE, PME et start-up, et informer et interagir avec les citoyens au sujet de l'IA.

MIAI repose sur deux grands thèmes : cœur IA et IA et applications. Le premier thème comprend quatre axes de recherche : architectures matérielles et embarquées pour l'IA, apprentissage et raisonnement, perception et interactions, IA et société. Le deuxième thème comprend trois axes de recherche : IA et santé, IA pour l'environnement et l'énergie, industrie 4.0. Le MIAI s'est appuyé sur l'existant tant pour la recherche que pour la formation. Mais désormais, l'offre de formation s'amplifie et se concentre sur le développement de la formation niveau Licence 3 alors que jusqu'à présent les formations développées relevaient du niveau BAC +4 et au-delà (+ 2 600 personnes formées à l'IA en 2021 et 2022 ; + 1 250 formées à l'IA en 2018 et 2019).

Partenariats privés :

- augmentation du nombre de collaborations (de 48 en 2018 à 89 en 2021) ;
- création du centre IA d'ATOS (200 membres d'ici fin 2024-Grenoble, Paris, Canada) ;
- création de deux laboratoires communs avec Orange (santé) et Constellium (alliages) ;
- renforcement des activités en IA à Grenoble (HPE, Schneider electric) ;
- création de huit start-up (10 de plus en cours, trois autres "jeunes pousses" ont rejoint MIAI).

Embauches et impacts individuels

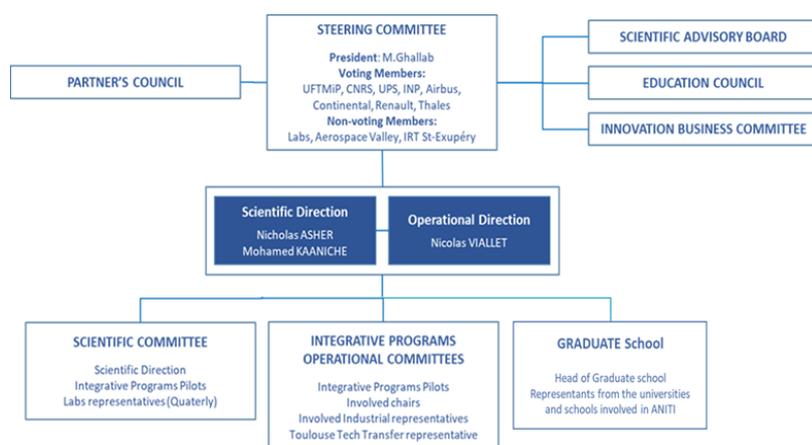
- 44 % (25 %) de doctorantes/doctorants et 60 % (70 %) de post-doctorants étrangers (diplôme précédemment acquis à l'étranger) ;
- *Theodore Christakis*: comité national sur l'éthique des données et de l'IA ; membre du Data Transfer Experts Council (Royaume-Uni) ;
- *Thierry Ménissier* : membre du comité exécutif.

Enjeux institutionnels et pérennisation

- image de l'IA à la française ;
- coordination des 3IA et collaboration entre instituts.

Annexe n° 19 : monographie 3IA ANITI Toulouse

Schéma n° 8 : Sociogramme de gouvernance



Source : ANITI

Tableau n° 55 : tableau récapitulatif des principaux éléments clé d'ANITI

<i>Coordination</i>	Université fédérale Toulouse Midi-Pyrénées
<i>Partenaires</i>	Membres académiques : les trois université de Toulouse, les principales écoles d'ingénieurs (INP, ISAE-SUPAERO, INSA, ENAC) Environ 30 partenaires industriels dont : Airbus, Safran, Continental, Thalès, Renault, Soprasteria, SNCF, BPCE
<i>Nombre de chercheurs</i>	108
<i>Nombre de chaires</i>	24
<i>Nombre de doctorants</i>	90
<i>Puissance de calcul</i>	Calculateurs des écoles d'ingénieurs (IRT, ISAE) / OSIRIM / Calmip

Source : Cour des comptes d'après données ANITI

Contexte institutionnel et genèse du projet

La candidature à l'AAP 3IA, marqué par l'échec de l'Idex, a d'abord été préparée par les directeurs de trois laboratoires LAAS²²⁰, IRIT²²¹ et IMT²²² et présenté aux partenaires éventuels (Thalès et Airbus) et au président de « l'Institut de Recherche Technologique Saint Exupéry »²²³. La réponse à l'AAP a été élaborée par de Nicholas Asher (chercheur en informatique) et Philippe Walter (ingénieur d'Airbus et de Aerospace Valley). Un comité scientifique a ensuite été mis en place. La gouvernance est inclusive avec les acteurs économiques à l'exception du *Scientific Advisory Board*.

La direction d'ANITI ne revendique pas l'interdisciplinarité de la recherche en IA comme centrale. Cependant, des thèmes issus du regroupement de cas d'usages industriels permettent des échanges scientifiques pour des chaires ne participant pas à la même communauté (statistique/optimisation, assimilation des données/analyse de sensibilité). Des travaux ont également lieu entre disciplines éloignées dans la cadre de l'IA hybride ou de la réglementation européenne (chaire mêlant juristes et mathématiciens).

L'offre de formation en intelligence artificielle sur le site toulousain est portée par tous les établissements, universités et écoles d'ingénieurs. L'institut « 3 IA » ANITI met en synergie les parcours en master ingénieur autour d'un projet commun. 51 formations sont labellisées et près de 350 modules de cours (soit près de 3 400 étudiants à la rentrée 2021, contre près de 2 000 à la rentrée 2018).

Pour attirer des talents, l'institut utilise des techniques de « *academic headhunting*²²⁴ » répandues dans les universités américaines. ANITI recherche des chercheurs de renommée internationale sur des thématiques précises privilégiées, puis organise une campagne de recrutement sur le site. Pour pérenniser le recrutement, l'institut se tourne vers les partenaires académiques tels que le CNRS et son processus de recrutement externe. L'attractivité de la chaire ANITI, avec son « *package* » ressources humaines et rémunération, est positive.

Trois chaires d'ANITI étudient des questions éthiques liées à l'IA. ANITI est responsable des questions éthiques dans l'utilisation de l'IA

²²⁰ LAAS-CNRS : Laboratoire d'Analyse et d'Architecture des Systèmes.

²²¹ IRIT : Institut de Recherche en Informatique de Toulouse.

²²² IMT : Institut de Mathématiques de Toulouse.

²²³ IRT Sant Exupéry : fondation de recherche privée qui promeut la valorisation de la recherche vers l'industrie.

²²⁴ Headhunting : chasseur de têtes.

pour le projet européen « COALA ICT 38 ». Projet financé dans le cadre de l'appel européen « H2020 ICT-38 » - Intelligence artificielle pour la fabrication, visant à développer une IA centrée sur l'humain digne de confiance dans la fabrication. L'institut a une approche d'architectures modestes, explicables et avec des garanties. Et non pas des architectures de grands calculateurs avec des milliards d'hyper paramètres visant une IA généraliste. Ces architectures sont plus frugales.

Enjeux institutionnels et pérennisation

La stratégie nationale a permis de doter les chaires de décharges d'enseignements et de moyens leur permettant de se focaliser sur leurs activités de recherche en proposant des personnels contractuels de recherche et des décharges d'enseignement.

L'appartenance à ANITI leur a permis de donner de la visibilité et offre des opportunités et des sollicitations leur permettant de contribuer ou de répondre aux appels à projet (par exemple les projets Taylor, Coala, Tuples, COCOBOT).

Les délais, et l'enveloppe budgétaire jugée limitée, de l'ANRT²²⁵ pour traiter les demandes de thèses CIFRE²²⁶ ces deux dernières années ont rendu difficile le lancement de certains projets, voire ont conduit à les annuler.

Le supercalculateur « Jean Zay » est peu utilisé par les chercheurs d'ANITI. Des calculateurs d'architectures petites et moyennes sont utilisées (OSIRIM ou Calmip). Les doctorants et post-doctorants utilisent les calculateurs locaux (celui de IRIT ou de l'ISAE-SUPAERO).

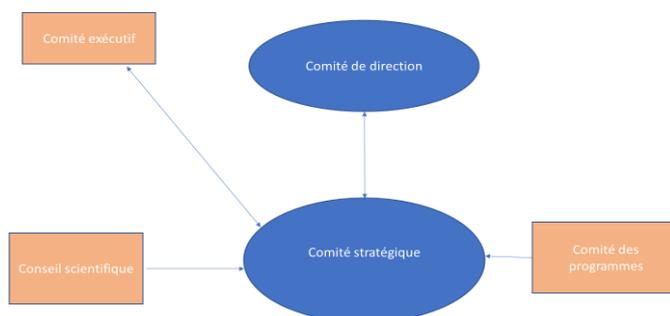
Ce qui pose problème est la lisibilité sur le long terme. Une crainte s'exprime de l'arrêt du modèle 3IA.

²²⁵ ANRT : Association nationale de la recherche et de la technologie.

²²⁶ CIFRE : convention industrielle de formation par la recherche. Le dispositif CIFRE permet aux entreprises de bénéficier d'une aide financière pour recruter de jeunes docteurs.

Annexe n° 20 : monographie DATAIA

Organigramme n° 3 : sociogramme de gouvernance



Source : Cour des comptes

Tableau n° 56 : tableau récapitulatif des principaux éléments clé

<i>Coordinateur</i>	Université Paris-Saclay
<i>Partenaires</i>	Membres académiques : Université Paris-Saclay (dont Évry et Versailles-Saint-Quentin), CentraleSupélec, ENS Paris-Saclay, AgroParisTech Organismes de recherche : CEA, Inria, INRAE, CNRS Établissements partenaires : Institut Mines-Télécom Business School, IFP Énergies nouvelles Partenaires industriels
<i>Nombre de chercheurs</i>	42 laboratoires (1 000 chercheurs sur Université Paris-Saclay)
<i>Nombre de chaires</i>	10 en IA
<i>Nombre de doctorants</i>	450 dont 45 financés sur UDOPIA (ANR)
<i>Puissance de calcul</i>	Puissance de calculs dans laboratoires / Calculateur Fusion-Ruche (CentraleSupélec / ENS Paris-Saclay) / Investissements dans des moyens de calcul (projet Dim AI4IDF)
<i>Lieu propre identifié</i>	Hébergé dans le bâtiment d'Inria du site de Saclay

Source : Cour des comptes d'après données DATAIA

Contexte institutionnel et genèse du projet

À la fin de l'année 2016, à l'initiative de plusieurs établissements (Inria, Telecom Paris, École Polytechnique, Université Paris-Sud, CentraleSupélec), un groupe de travail s'est constitué avec pour objectif de répondre à l'appel à projets (AAP) « Institut de Convergence » sur la thématique des sciences de données pour la société. Le projet I2-DRIVE (ancien nom de DATAIA) a été accepté et financé en vague 2 de l'AAP « Institut de Convergence ». La contractualisation avec l'ANR a été faite en octobre 2017 pour une fin en 2025. Le 1^{er} janvier 2021, DATAIA est devenu l'Institut d'Intelligence Artificielle de l'Université Paris-Saclay. À son lancement, DATAIA était centré sur des thématiques dites « cœur IA » avec des applications sociétales (sciences humaines et sociales) puis a développé des recherches dans les domaines suivants : IA et bio-santé, aspects éthiques et sociaux de l'IA, IA et sciences humaines et sociales, fondements mathématiques de l'IA, IA pour les autres sciences (physique), IA et environnement, IA de confiance-sécurité.

Initialement, l'Institut ne possédait pas de volet formation au niveau licence ou master. Néanmoins, dès le démarrage du projet, il avait dans son programme la co-organisation des écoles d'été et d'hiver à destination des doctorants. Plus récemment, l'Institut a réalisé la cartographie des enseignements en IA dispensés essentiellement dans les trois *Graduates schools* (GS) : informatique et sciences du numérique (ISN), mathématiques, Sciences de l'ingénierie et des systèmes (SIS). Les GS portent les diplômes de niveau master et doctorat. Au total, 530 unités d'enseignements relatives à l'IA sont proposées à l'Université Paris Saclay pour 67 parcours du niveau Bac+1 à Bac+5. Une massification des enseignements de l'IA est prévue dans le contexte du projet « SaclAI-Schools » dans le cadre de l'AMI CMA. Est mis en place un système de professeurs attachés avec des charges de 48 heures d'enseignement annuelle. Il y a une forte concurrence pour attirer les talents dans le domaine de l'intelligence artificielle. DATAIA, adossé à l'Université Paris-Saclay, se positionne sur son interdisciplinarité dans le domaine de la recherche. Les chercheurs en IA peuvent se situer à l'interface de l'IA et d'autres disciplines comme la biologie (bio-informatique), la physique des particules ou le droit. La renommée de l'Université et de ses laboratoires permet également une attractivité forte. DATAIA finance un programme de professeurs invités sur une année universitaire complète permettant une implication réelle des invités dans l'éco-système de l'Université Paris-Saclay et favorisant la création de lien de collaboration solides et pérennes.

En matière d'éthique, quatre projets sont associés à cette thématique (GDP-ERE, Bad Nudge-Bad Robot ?, Vadore, YPPOG). Cet enjeu éthique

est traité sous des angles distincts, depuis les aspects juridiques (utilisation des données sensibles médicales, recherche d'emplois), ou mathématiques dans des domaines variés (agents conversationnels, données patients, aide à la recherche d'emploi, impact des pesticides sur la santé). La frugalité est une problématique récente en IA. Plusieurs laboratoires du périmètre DATAIA ont des équipes de recherche sur ces thématiques au niveau de l'impact écologique du traitement numérique et du stockage massif.

Enjeux institutionnels et pérennisation

Le fait de ne pas être labellisé 3IA convient à l'Institut. Son écosystème se structure et se fédère par le choix de projets de financements collaboratifs pluridisciplinaires et impliquant au moins deux établissements fondateurs. Par opposition, des financements pour des chaires individuelles auraient pu impacter la dynamique fédérative de l'Institut.

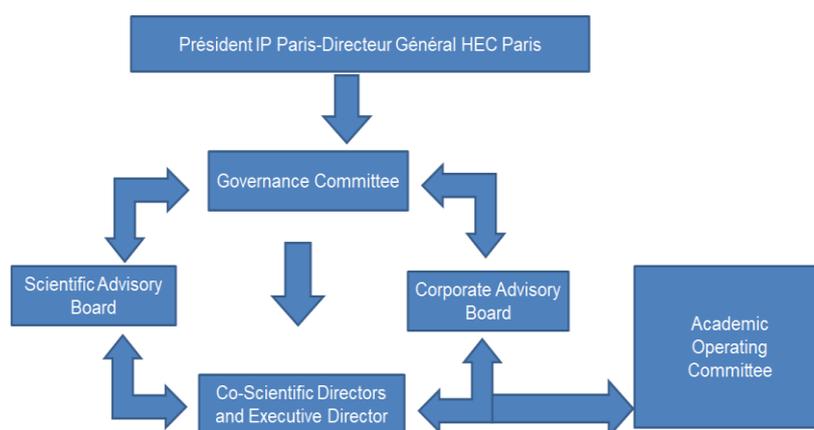
Le travail de structuration nationale a permis à l'Institut de nouer des liens avec les autres instituts nationaux, particulièrement ceux d'Île-de-France et d'organiser des événements communs et des réponses communes à des appels à projets tel que DIM²²⁷ « AI4IDF » de la région Île-de-France dans lequel les quatre Instituts d'intelligence artificielle de la région (DATAIA, Hi! PARIS, PRAIRIE et SCAI) sont impliqués. DATAIA, qui n'a pas de collaboration avec le calculateur « Jean Zay », est co-porteur, dans le cadre du projet DIM « AI4IDF », d'un investissement dans des moyens de calcul.

Cependant, la direction de DATAIA estime que les chaires individuelles, qui permettent à une petite équipe réunie autour d'un chercheur de développer une recherche de qualité, rendent plus difficile la structuration de la communauté.

²²⁷ DIM : domaine d'intérêt majeur pour 2022-2026 soutenu par la région Île-de-France centré sur l'IA.

Annexe n° 21 : monographie Hi! PARIS

Organigramme n° 4 : sociogramme de gouvernance



Source : Cour des comptes

Tableau n° 57 : tableau récapitulatif des principaux éléments clé

<i>Coordinateurs</i>	HEC Paris et Institut Polytechnique de Paris
<i>Partenaires</i>	Membres d'IPP : École Polytechnique, ENSTA, ENSAE, Télécom Paris, Télécom SudParis Partenaires industriels mécènes : L'Oréal, Capgemini, TotalEnergies, Kering, Rexel, Schneider Electric, Vinci
<i>Nombre de chercheurs</i>	170
<i>Nombre de chaires</i>	10 chaires + 5 fellowships
<i>Nombre de doctorants</i>	25
<i>Puissance de calcul</i>	Relation avec GENCI / Calculateur de TotalEnergies
<i>Lieu propre identifié</i>	Oui / Bâtiment de Télécom Paris

Source : Cour des comptes d'après les données Hi! PARIS

Contexte institutionnel et genèse du projet

Hi! PARIS a été lancé le 15 septembre 2020 par ses deux membres fondateurs que sont l'École des hautes études commerciales (HEC Paris) et l'Institut Polytechnique de Paris (IPP). Inria a rejoint le centre en tant que nouveau membre en juillet 2021. Il s'agit d'accélérer le développement des fondateurs dans le domaine de l'intelligence artificielle et de la science des données au « *service de la science, de l'économie et de la société avec pour ambition de se placer au meilleur niveau international* ». Hi! PARIS a l'ambition de « *contribuer au développement de la souveraineté numérique de la France et de l'Europe, assurant la compétitivité de ses entreprises, l'efficacité de ses administrations, et la qualité de vie des citoyens* ». Le centre ambitionne de devenir une destination « *de premier choix* » pour la recherche, la formation et l'innovation en intelligence artificielle et sciences de données et ses applicatifs.

Sur la formation, le centre propose un ensemble d'initiatives en collaboration avec les directions de la formation d'IPP et HEC Paris et soutient les objectifs de développement de nouveaux programmes d'enseignement, et une forte augmentation du nombre d'étudiants avec une double compétence data/IA et business. Hi! PARIS organise des événements du type « Hi !ckaton²²⁸ » ou data bootcamp. Les formations en IA qui bénéficient des actions pédagogiques de Hi! PARIS concernent des Masters 2 et des Mastère spécialisés, par exemple en « data science », « data science et business », « data et IA », « data analytics and corporate finance », « international finance », « probabilités et finances ».

Le programme phare du centre Hi! PARIS est le programme de recrutement de nouveaux professeurs de niveau international (chaires) et de rétention (*fellowship*). Une large part du budget (les deux tiers) est consacrée à ce programme, et principalement pour le recrutement.

L'enjeu éthique est abordé dans des projets de recherche à HEC Paris, de manière directe (élaboration de méthodes d'évaluation d'algorithmes sous l'angle éthique) ou indirecte (impact de la pratique juridique de la conformité algorithmique). À IP Paris, des travaux sont menés sur les problématiques du biais en apprentissage profond et les problèmes de « vie privée ». Par ailleurs, éthique en IA et sciences des données est au cœur de plusieurs enseignements (par exemple, dans le MBA/EMBA d'HEC Paris, le cours « *data privacy : the legal and ethical use of human data* » ou dans le master DSB le cours « *data et ethics* »). L'apprentissage automatique frugal est traité dans des travaux de

²²⁸ Hackaton : compétition entre étudiants sur la base d'un même jeu de données.

recherche : » frugalité des données », « frugalité d'apprentissage » et, ou, « d'inférence » pour limiter le coût associé au processus d'apprentissage, en particulier les ressources de calcul et de « frugalité du modèle ».

Enjeux institutionnels et pérennisation

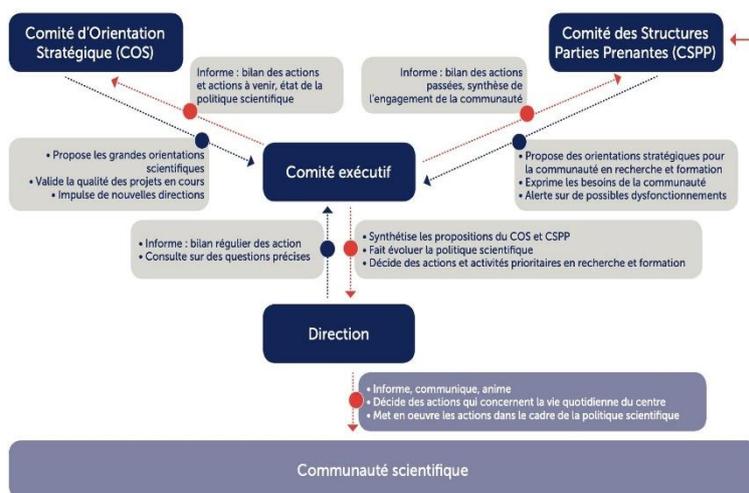
De création très récente, le 15 septembre 2020, Hi! PARIS a été invité par les autres centres IA de la région parisienne à participer à l'appel à projets DIM « IA4IdF » de la région Île-de-France. Le projet a été retenu par le conseil régional en février 2022.

Avec l'absence de labellisation, les dirigeants du centre ont envisagé et créé un montage financier différent, basé sur le mécénat d'envergure, à l'image de ce qui existe en Amérique du nord. En l'absence de financements publics liés à la labellisation, le centre souhaite la mise en place d'un principe de financement public équivalent au budget privé obtenu.

La direction du centre considère que l'absence de labellisation a été un signal négatif pour le tissu socio-économique, l'empêchant de nouer des relations contractuelles avec les grands acteurs de l'IA à l'instar de « PRAIRIE ». La logique partenariale de Hi! PARIS est axée sur les acteurs industriels nationaux. Ces partenaires sont des entreprises dont le cœur de métiers n'est pas le numérique (à l'exception de Capgemini).

Annexe n° 22 : monographie SCAI

Organigramme n° 5 : sociogramme de gouvernance



Source : SCAI

Tableau n° 58 : tableau récapitulatif des principaux éléments clé

Coordinateur	l'Idex Sorbonne Université.
Partenaires	Structure de l'Alliance Sorbonne Université (UTC, MNHN, INSEAD, Inria, CNRS, APHP, Inserm, CEA) Principaux partenaires industriels et institutionnels : APHP, Criteo, EDF, ESSILOR, LNE, Mairie de Paris, Pierre Fabre dermo-cosmétique, SIEMENS, SUEZ, TOTAL, VALEO, Mila Institut québécois IA, Université Laval, UC Berkeley
Nombre de chercheurs	129 + invités
Nombre de chaires	Trois chaires
Nombre de doctorants	38
Puissance de calcul	Cluser de calcul GPU (puissance de 1 800 Teraflops) / Accès au super ordinateur Jean Zay / Mais accès difficile

Source : Cour des comptes d'après les données SCAI

Contexte institutionnel et genèse du projet

L'ambition de SCAI, créé par Sorbonne université en juin 2019, est de contribuer significativement à l'excellence de la recherche et de la formation en intelligence artificielle, en favorisant les échanges entre enseignants-chercheurs, chercheurs, étudiants et partenaires industriels. Il a pour objectif de devenir la vitrine du savoir-faire de l'Alliance Sorbonne Université dans le domaine de l'intelligence artificielle et de constituer une porte d'entrée pour les partenaires académiques et industriels. SCAI est un organe interne à l'Idex Sorbonne Université dont l'existence ne dépend pas d'un appel à projets national ou d'un programme spécifique avec une durée déterminée. Son financement est assuré pour partie par l'Idex mais aussi par des ressources propres (projets ANR, partenariats industriels, etc.).

Les quatre grands axes de recherche affichés par SCAI depuis sa création sont les suivants : Mathématiques, informatique, robotique, Santé, médecine et biologie, Climat, environnement, Univers, Humanités numériques. Le positionnement volontairement interdisciplinaire de l'institut est incarné par le programme doctoral, qui a permis de financer 35 contrats doctoraux en trois ans (10 *via* l'ANR et 25 *via* l'Alliance Sorbonne Université). La quasi-totalité de ces contrats ont concerné des sujets de recherche à l'interface entre deux voire trois disciplines (mathématiques, informatique, physique, biologie, médecine, lettres, musique, etc.).

La formation est principalement portée par la faculté des sciences, à travers quatre parcours de masters, tous ces parcours étant adossés à des laboratoires. Il existe une unité d'enseignement dédiée à l'IA en L1 destinée aux étudiants ayant eu un enseignement en informatique au lycée (60 étudiants). Les objectifs 2022-2028 en termes de formation sont les suivants : + 50 % dans les masters IA, renforcement de l'offre IA du département de mathématiques, ouverture d'une mineure « données » en L2/L3 (60 étudiants). Monter à 90 étudiants pour l'UE L1, + 50 % formation continue.

Les principaux aspects éthiques comprennent l'impact socio-économique et environnemental, la protection des données, les droits fondamentaux des citoyens, les valeurs européennes, ainsi que la transparence de la technologie de l'intelligence artificielle. Par ailleurs, un cours en ligne européen a été développé, en partenariat avec les universités de Copenhague et Milan (alliance 4EU+), qui est opérationnel depuis septembre 2022 (*Thinking AI: bringing together ethical, legal and social aspects of AI*). L'institut a enfin organisé, durant l'été 2021, une école virtuelle conjointe avec l'OBVIA (Observatoire international sur les impacts

sociétaux de l'IA et du numérique) portant sur le traitement des aspects éthiques, et plus largement sociaux, des systèmes d'IA responsables.

Enjeux institutionnels et pérennisation

Le choix a été d'opter d'emblée pour une structuration locale, agile et réactive, pour laquelle une labellisation 3IA aurait probablement été moins adaptée. Une fois les acquis locaux consolidés, SCAI a alors pu accélérer sa stratégie de développement national. De ce point de vue, l'absence de labellisation 3IA fut à la fois une force et une faiblesse. Une force dans la mesure où l'autonomie a permis de tisser des partenariats propres et d'être maîtres des actions, tant sur la forme que sur le fond, sans les contraintes des 3IA. Une faiblesse car l'absence de label rend moins lisible l'institut dans un environnement très concurrentiel et en évolution rapide. Trois ans après sa création, l'institut semble avoir sa place dans le paysage de l'intelligence artificielle française, comme en témoignent par exemple les demandes croissantes d'associations pour participer à des appels à projets nationaux et internationaux.