

Analyse des publications scientifiques d'Inria

**Contribution à l'évaluation de l'Institut national
de recherche en sciences et technologies du
numérique (Inria)**

Septembre 2023

Sommaire

Préambule	4
Synthèse	5
Introduction.....	7
Évolutions des publications et des copublications	7
Profil des publications et spécialisation en informatique.....	9
Indicateurs d'impact des publications.....	11
Annexes	13
Annexe 1 – Base de données et méthode.....	13
Annexe 2 – Classification disciplinaire	15
Annexe 3 – Couverture par le WoS des actes A et A* de CORE.....	16
Annexe 4 – Les institutions comparées	18

Préambule

Ce rapport, élaboré par l'Observatoire des sciences et techniques (OST) et le Département d'évaluation des organismes du Hcéres, constitue une contribution à l'évaluation de l'Institut national de recherche en sciences et technologies du numérique (Inria) pour la période 2018-2022 ; il analyse les publications scientifiques de l'Institut. L'objet de l'analyse, ainsi que le choix des institutions étrangères et françaises comparées à Inria ont été définis en concertation avec Inria. Une version préliminaire du rapport a été transmise à l'institut avant la finalisation de son rapport d'auto-évaluation.

Le Hcéres, signataire de la Déclaration de San Francisco sur l'évaluation de la recherche (Dora), a contribué en 2022 à la préparation de l'appel de Paris sur l'évaluation de la recherche puis a rejoint la Coalition on advancing research assessment (CoARA) ; il est donc attentif à bien positionner l'analyse des indicateurs quantitatifs dans le cadre de l'évaluation globale de la recherche. L'évaluation d'Inria est une évaluation qualitative qui s'appuie sur un usage responsable des indicateurs quantitatifs. Ainsi, la présente analyse des indicateurs quantitatifs n'est pas l'évaluation elle-même : elle est une contribution à l'évaluation d'Inria. Tout en étant pleinement conscient qu'il ne couvre pas toutes les productions d'Inria, ni ne décrit tous les impacts de ses productions, le Hcéres a élaboré ce rapport avec l'objectif qu'il apporte des informations utiles à la fois au comité d'évaluation et à Inria lui-même.

Synthèse

Ce rapport analyse les publications scientifiques d'Inria et les compare à celles d'une quinzaine d'institutions françaises et étrangères ayant des activités de recherche dans les mêmes domaines. Certaines sont spécialisées dans les domaines de l'informatique, comme l'institut Alan Turing, DFKI ou ETRI. D'autres sont pluridisciplinaires et de taille beaucoup plus importante, comme les universités Oxford ou Stanford. Les institutions comparées sont originaires de pays à hauts revenus intensifs en recherche.

Quatre corpus de publications ont été constitués pour chacune des institutions comparées : l'ensemble des publications toutes disciplines confondues, le périmètre de la discipline informatique, celui du domaine de l'intelligence artificielle (IA) et le corpus des actes de conférence classées A et A* par la *Computing Research and Education Association of Australasia*, CORE, association de départements universitaires en informatique d'Australie et Nouvelle Zélande. Afin de tenir compte de l'importance des copublications, nationales et internationales, les indicateurs sont généralement calculés à partir d'un décompte des contributions des institutions, autrement dit, une fraction de publication est attribuée à une institution en fonction du nombre total de signataires.

Evolution des publications durant la période 2017-21

L'épidémie de covid a affecté le nombre de publications mondiales en informatique, avec une baisse sensible en 2020 et un redressement en 2021. Au cours de la période, l'ensemble de l'échantillon étudié voit sa part mondiale de publications en informatique diminuer, mais les parts du CNRS et d'Inria baissent plus que celles d'autres institutions (d'environ 25 %). Les deux organismes français conservent néanmoins les premiers rangs de l'échantillon, avec respectivement 1,1 et 0,3 % du total mondial. Le paysage est un peu différent dans le domaine de l'IA. Le CNRS est aussi l'institution qui publie le plus, mais avec une part mondiale des publications en IA de 1 % en 2021. Inria est en 4^{ème} position avec une part de 0,23 %, derrière Stanford et MIT.

Profil disciplinaire des publications

Inria publie dans différentes disciplines au-delà de l'informatique, notamment en sciences pour l'ingénieur, mathématiques, biologie fondamentale et physique. L'indice de spécialisation permet de comparer l'engagement relatif d'une d'institution dans un domaine. Il rapporte la part du domaine dans le total des publications d'une institution, à cette même part pour le total des publications mondiales - la valeur neutre pour le monde étant 1. Inria est très fortement spécialisé en informatique, avec un indice de spécialisation supérieur à 6 en 2020, mais l'institut est pratiquement aussi spécialisé en mathématiques. Inria est beaucoup moins spécialisé en sciences pour l'ingénieur, mais cette discipline publiant beaucoup plus à l'échelle mondiale, elle représente tout de même la deuxième en nombre de publications pour Inria.

Les universités et organismes de recherche pluridisciplinaires de l'échantillon tendent à ne pas être spécialisés en informatique. Stanford a cependant un indice un peu supérieur à 1 et le MIT un indice proche de 2. Le CNRS a une part de publications en informatique près de 20 % supérieure à la moyenne mondiale. Les institutions ayant les indices les plus élevés, entre 5 et 9 fois la moyenne mondiale, sont cependant les organismes plus centrés sur l'informatique et de plus petite taille.

La spécialisation en IA est corrélée à la spécialisation en informatique, mais certaines institutions sont clairement relativement plus présentes en IA. C'est tout particulièrement le cas de l'institut Alan Turing et dans une moindre mesure de RIKEN, de l'université d'Oxford et de DFKI. Au sein de l'informatique, Inria n'apparaît pas particulièrement engagé dans l'IA, avec un indice de spécialisation inférieur (5,4 contre 6,5 en informatique).

Mesures de l'impact scientifique des publications

L'analyse de deux mesures d'impact à partir des citations des publications donne des résultats convergents pour les différentes institutions. Pour l'indice d'impact moyen normalisé, la valeur mondiale de référence est 1. L'indice est calculé pour le total des publications et les trois corpus spécifiques du rapport.

Quatre institutions ont des indices d'impact moyens supérieurs à 2 en informatique et en IA : les universités Stanford et Oxford, le MIT et la société Max Planck. Pour Stanford, Oxford et la société Max Planck, ces indices sont sensiblement supérieurs à ceux de l'ensemble de leurs publications toutes disciplines confondues. C'est l'inverse pour Inria qui a un indice de 1,2 pour l'ensemble de ses publications et de 1,1 pour l'informatique et l'IA.

L'indice d'activité dans le décile des publications les plus citées au monde est complémentaire en se concentrant sur les publications les plus citées. Cet indicateur fournit un positionnement similaire des institutions comparées. Les cinq institutions ayant les indices les plus élevés (Stanford, MIT, Max Planck, Alan Turing et Oxford), comme celles qui ont les indices les plus faibles, sont les mêmes. L'indicateur pour Inria est un peu plus élevé en informatique, 20 % au-dessus de la moyenne mondiale contre 10 % pour l'impact moyen.

Introduction

Ce rapport analyse les publications scientifiques de différentes institutions à partir des données de la base de publications de l'OST, version enrichie du Web of Science (WoS) de Clarivate Analytics qui est présentée par l'annexe méthodologique (annexe 1). Inria a signalé au Hcéres les revues et conférences que l'organisme considère comme particulièrement pertinentes pour son domaine et l'OST a vérifié que la base fournit une couverture suffisante de ces supports.

Le rapport distingue quatre périmètres de publications pour ses analyses. Quatre corpus de publications ont donc été constitués pour chacune des institutions comparées. Le premier est constitué de l'ensemble des publications des institutions, toutes disciplines confondues. Le deuxième regroupe les publications de la discipline informatique identifiée dans la base de l'OST à partir des 254 catégories WoS. Cette discipline regroupe 10 catégories, 7 mentionnant explicitement computer science (artificial intelligence, cybernetics, hardware & architecture, information systems, software engineering, theory & methods, interdisciplinary applications), les 3 autres étant logic, medical informatics, robotics et telecommunications (voir annexe 2). Le troisième corpus concerne les publications dans le domaine de l'intelligence artificielle (IA), identifiées à partir d'une analyse thématique croisée avec l'appartenance à des domaines de la discipline informatique ; l'approche est précisée dans l'annexe méthodologique. Enfin, le quatrième corpus correspond à celui des actes de conférence classées A et A* par la Computing Research and Education Association of Australasia, CORE, association de départements universitaires en informatique d'Australie et Nouvelle Zélande. L'association publie une liste de conférences dans le domaine et les classe en catégories sur la base de l'expertise de ses membres. Les classes A* et A regroupent les conférences les plus sélectives ; la classe A* compte 7,5 % du total des 804 conférences répertoriées en 2023 et la classe A, 14,6 %¹. La constitution de ce corpus et la couverture de ces actes par la base de l'OST sur la période de l'analyse sont précisées à annexe 3.

Le rapport comporte trois parties. La première présente l'évolution durant la période 2017-21 des publications et des copublications d'Inria, ainsi que des institutions de comparaison. Ces institutions sont présentées de façon succincte en annexe 4. La deuxième partie analyse le profil disciplinaire de ces institutions, notamment en informatique et dans le domaine de l'intelligence artificielle. La troisième partie présente des indicateurs d'impact des publications des institutions sur les différents corpus.

La période de cinq ans d'analyse des publications correspond aux années 2017 à 2021. Du fait de la date d'actualisation de la base de publications de l'OST utilisée, l'année 2021 comprend environ 95 % des données.

Évolutions des publications et des copublications

Le tableau 1 fournit le nombre total de publications sur la période considérée de chacune des institutions comparées en compte entier (encadré 1 et annexe 1). Il distingue celles qui sont spécialisées dans les domaines de l'informatique de celles qui ont un périmètre pluridisciplinaire. Ces dernières peuvent aussi avoir un périmètre moins large que les grandes universités, comme l'ensemble des instituts Fraunhofer ou RIKEN. Les institutions spécialisées ont un nombre de publications nettement inférieur aux institutions pluridisciplinaires. Ces dernières sont même très diverses, que ce soit du fait de leur positionnement plus ou moins appliqué et orienté vers les sciences de l'ingénieur, ou de leur taille absolue.

Les publications des institutions françaises sont comptabilisées sur la base des publications des unités de recherche dont elles sont tutelle ou cotutelle. Autrement dit, il ne s'agit pas des seules publications signées par les personnels de ces institutions, mais par les chercheurs affiliés aux différents laboratoires. C'est ce qui explique notamment le nombre très élevé de publications attribuées au CNRS. Avec plus de 12 500 publications sur la période, soit près de 2 520 par an, Inria a le nombre de publications le plus élevé des institutions spécialisées comparées.

Tableau 1 - Nombre total de publications en compte entier, toutes disciplines, 2017-21*

Institutions spécialisées		Institutions pluridisciplinaires			
DFKI	1 086	Fraunhofer	13 556	Stanford U.	62 103
CWI	1 188	RIKEN	16 523	Oxford U.	63 712
Alan Turing Institute	1 534	CEA	28 813	Max Planck	63 829
NICT	2 969	U. McGill	37 590	U. Toronto	77 895
ETRI	3 849	MIT	44 883	CNRS	272 152
INRIA	12 584				

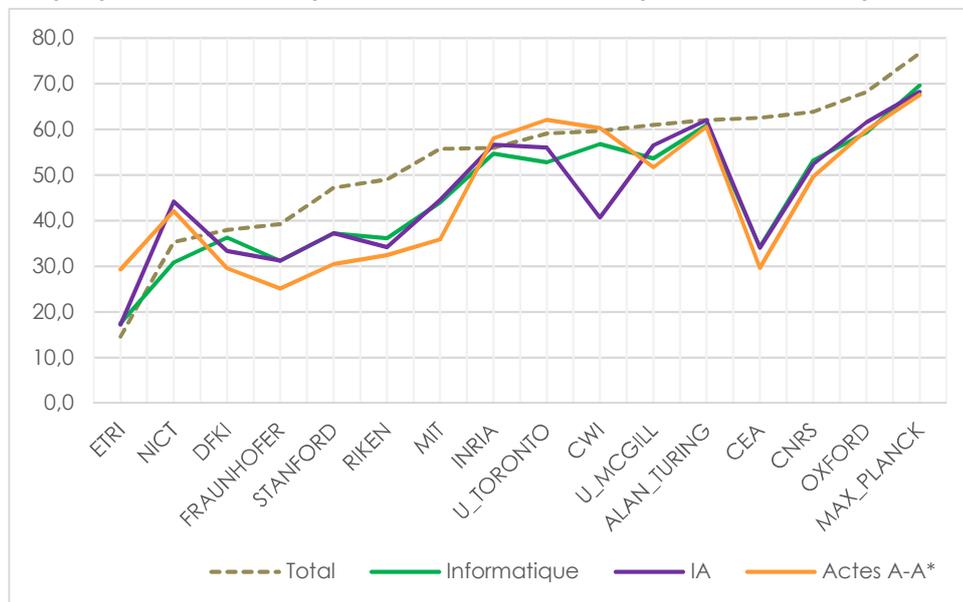
* Pour l'année 2021, la base de publications est complète à environ 95 %

Source : Base OST, WoS, calculs OST

¹ Voir le portail des conférences classées sur le site de CORE, <https://www.core.edu.au/conference-portal>

Pour toutes les institutions, les publications ont généralement plusieurs auteurs, y compris des coauteurs étrangers. La part des copublications internationales rapporte le nombre des publications comportant des affiliations à l'étranger au total. Le graphique 1 souligne que la part des copublications internationales varie sensiblement entre institutions. Toutes disciplines confondues, elle varie de 15 % pour ETRI à 77 % pour Max Planck. Ces variations s'expliquent à la fois par des caractéristiques nationales (les grands pays et les pays asiatiques tendent à avoir des taux plus faibles) et par le fait que les recherches appliquées sont généralement menées dans un cadre plus local. Ainsi, les instituts Fraunhofer ont-ils un taux de copublications internationales beaucoup plus faible que la société Max Planck.

Graphique 1 – Part des copublications internationales par institution, compte entier, en %, 2017-21*



* Pour l'année 2021, la base de publications est complète à environ 95 %

Source : Base OST, WoS, calculs OST

Les parts de copublications internationales des quatre corpus sont corrélées, mais pour certaines institutions le domaine informatique est moins internationalisé. En France, c'est le cas pour le CNRS et le CEA, mais pas pour Inria. Cela s'explique sans doute par la part importante de disciplines très internationalisées pour ces deux opérateurs français (sciences de l'univers, biologie fondamentale par exemple). Parmi les institutions spécialisées dans le domaine informatique, Inria a des publications moins internationalisées que l'institut Alan Turing ou CWI, mais sensiblement plus que ETRI.

Afin de tenir compte du fait que les publications sont très largement des copublications, les indicateurs de ce rapport sont généralement calculés à partir d'un décompte des publications en compte fractionnaire, défini à l'encadré 1.

Encadré 1 - Dénombrer les publications : compte entier et compte fractionnaire

En compte entier, une publication est comptée entièrement (1) pour une institution dès lors que celle-ci apparaît dans la liste des adresses d'affiliation. Le compte entier rend compte de la participation de chacun des signataires à la publication. Dans la mesure où une publication est comptée autant de fois qu'il y a d'adresses d'affiliation, la somme de ces participations génère des doublons et le compte entier n'est pas additif.

Afin de tenir compte des différentes contributions, le compte fractionnaire attribue une fraction de publication à chacune des institutions signataires en comptabilisant $1/n$ pour chacune, n étant le nombre total d'affiliations figurant sur la publication. Le compte fractionnaire est additif et permet de calculer des parts de publications, par exemple au sein d'un pays, sans doublons. Il permet aussi de tenir compte des pratiques de collaboration différentes selon les domaines de recherche, et d'éliminer des biais dans les comparaisons entre domaines (les domaines où les collaborations sont nombreuses comptent plus de publications, mais pas nécessairement plus de contributions individuelles).

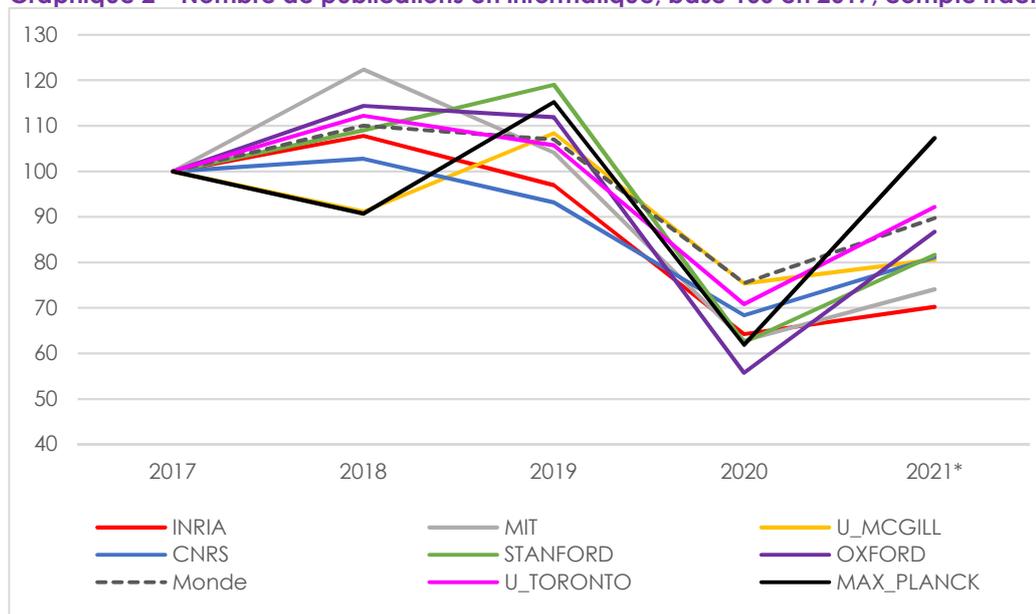
Le compte fractionnaire est aussi utilisé entre disciplines ou catégories disciplinaires pour tenir compte du fait que les supports de publications peuvent être attribués à deux voire plusieurs catégories.

Source et compléments : méthode en annexe 1.

Le graphique 2 permet de suivre l'évolution du nombre de publications en informatique des institutions comparées, indépendamment de leur taille. Il montre que la période du covid a sensiblement affecté les publications mondiales en informatique. Les institutions de l'échantillon ont été plus ou moins touchées, mais les évolutions sont les mêmes : baisse sensible en 2020 et redressement en 2021. Rappelons que la base de l'OST n'est pas complète pour 2021 au moment où les indicateurs ont été calculés. C'est notamment le cas pour les

actes de conférences et avec une année complète, l'année 2021 compte en fait un peu plus de publications que niveau 2019 (tableau 2). Inria apparaît parmi les institutions dont le nombre de publications se tasse relativement au monde et aux institutions de l'échantillon, le phénomène débutant en 2019. Cette évolution en compte fractionnaire pourrait être en partie due à une augmentation de la part de copublications internationales ; elle était plus faible pour Inria durant la période d'évaluation précédente².

Graphique 2 – Nombre de publications en informatique, base 100 en 2017, compte fractionnaire



* Année complète à environ 95 %

Source : Base OST, WoS, calculs OST

En complément, le tableau 2 fournit le nombre de publications mondiales et le nombre que compte les différents corpus étudiés dans ce rapport. La baisse du nombre de publications l'année de la déclaration du covid est la plus nette pour les articles parus dans des actes de conférences. Elle se répercute au niveau de l'ensemble de l'informatique.

Tableau 2 - Nombre de publications mondiales *, 2017-21**

Publications mondiales	2017	2018	2019	2020	2021**	2017-21**
Total toutes disciplines	2 087 023	2 156 506	2 234 142	2 310 341	2 550 138	11 338 150
Informatique	164 814	172 554	184 985	150 552	134 772	807 678
Part du total	7,9%	8,0%	8,3%	6,5%	5,3%	7,1%
Intelligence artificielle	64 583	74 486	86 886	79 951	83 092	388 998
Part du total	3,1%	3,5%	3,9%	3,5%	3,3%	3,4%
Articles dans actes CORE AA*	25 785	28 377	30 361	22 909	20 565	127 997
Part du total	1,2%	1,3%	1,4%	1,0%	0,8%	1,1%

* Compte fractionnaire sur la dimension disciplinaire (pour toutes disciplines, égal au compte entier)

** 2021 complète à environ 95 %

Source : Base OST, WoS, calculs OST

Profil des publications et spécialisation en informatique

Inria publie dans différentes disciplines au-delà de l'informatique, notamment en sciences pour l'ingénieur, mathématiques, biologie fondamentale et physique³. Le tableau 3 fournit les indices de spécialisation d'Inria dans les cinq grandes disciplines dans lesquelles l'organisme est le plus engagé. L'indice rapporte la part de la discipline dans le total des publications d'Inria, à cette même part pour le total des publications mondiales. Inria est très fortement spécialisé en informatique, avec un indice de spécialisation de 7, mais l'institut est pratiquement aussi spécialisé en mathématiques (et en leur sein, les mathématiques appliquées). Inria est beaucoup moins spécialisé en sciences pour l'ingénieur, mais cette discipline publiant beaucoup plus à l'échelle mondiale, elle représente tout de même la deuxième en nombre pour Inria.

² Proche de 50 % contre 58 % en 2017-21.

³ L'évaluation quinquennale précédente avait permis de détailler la part des disciplines dans les publications d'Inria.

Tableau 3 – Indice de spécialisation d'Inria pour ses 5 premières disciplines de spécialisation, 2017-21*

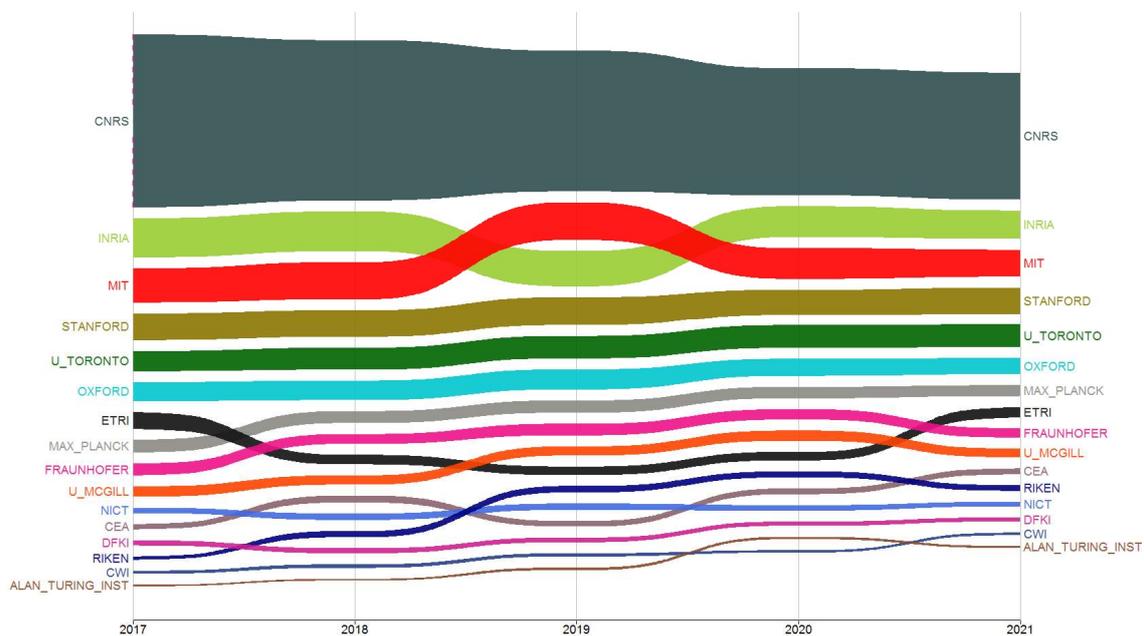
	2017	2018	2019	2020	2021*
Informatique	6,29	6,45	5,86	6,34	7,17
Mathématiques	4,99	5,61	5,50	6,61	7,15
Sciences pour l'ingénieur	1,34	1,25	1,37	1,55	1,73
Physique	0,57	0,51	0,58	0,68	0,85
Biologie fondamentale	0,53	0,45	0,55	0,64	0,51
Toutes disciplines	1,00	1,00	1,00	1,00	1,00

* Année complète à environ 95 % ; compte fractionnaire

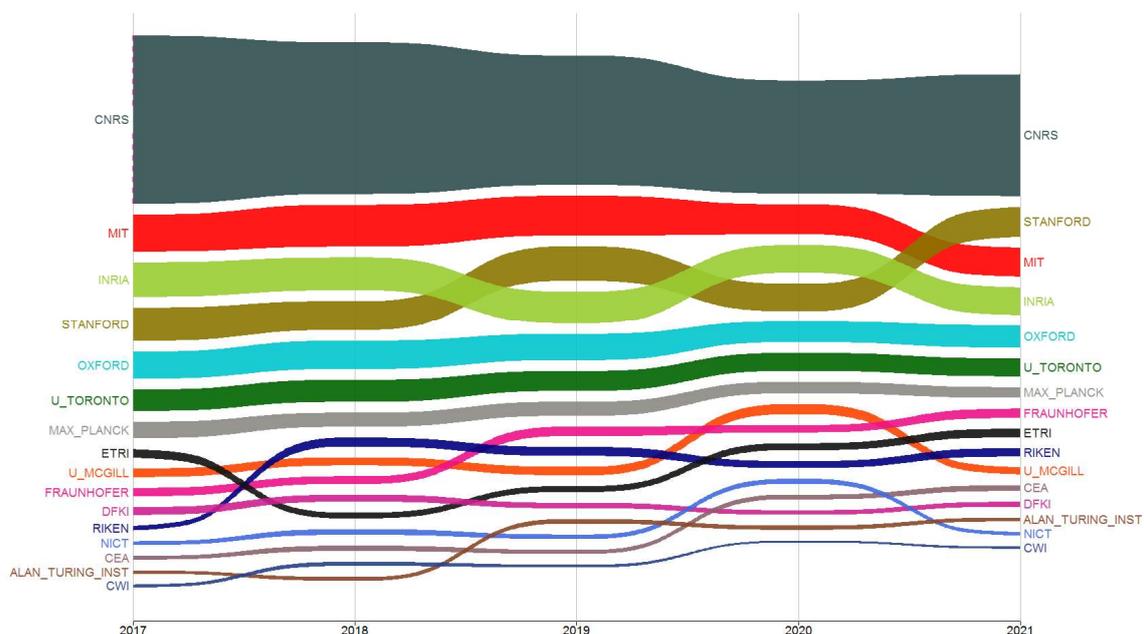
Source : Base OST, WoS, calculs OST

Le graphique 3a fournit la part mondiale des publications en informatique. L'ensemble des institutions voit sa part mondiale diminuer, ce qui peut être en partie dû à l'absence d'institutions de pays émergents, dont la part tend à augmenter dans toutes les disciplines. Les parts mondiales du CNRS et d'Inria baissent plus que celles d'autres institutions (d'environ 25 %), mais conservent les premiers rangs, avec respectivement 1,1 % et 0,3 %.

Graphique 3a – Part mondiale de publications en informatique par institution, compte fractionnaire, 2017-21*



Graphique 3b – Part mondiale de publications en IA par institution, compte fractionnaire, 2017-21*



* Année 2021 complète à environ 95 %

Source : Base OST, WoS, calculs OST

Le paysage est un peu différent dans le domaine de l'IA (graphique 3b). Le CNRS est aussi au premier rang, mais avec une part mondiale de 1 % en 2021. Inria est en 4ème position avec une part de 0,23 %, derrière Stanford et MIT.

Le tableau 4 permet de préciser l'engagement relatif des institutions en informatique et en IA. L'indice de spécialisation rapporte la part d'une discipline dans le total des publications d'un pays ou d'une institution à cette même part dans le total des publications mondiales. La valeur neutre pour le monde est donc égale à 1. Les universités et organismes de recherche pluridisciplinaires tendent à ne pas être spécialisés en informatique. Stanford a cependant un indice un peu supérieur à 1 et le MIT un indice proche de 2. Le CNRS a une part de publications en informatique de près de 20 % supérieure à la moyenne mondiale. Les institutions ayant les indices les plus élevés, entre 5 et 9 fois la moyenne mondiale, sont cependant les organismes plus centrés sur l'informatique et de plus petite taille. L'indice de spécialisation d'Inria en informatique est de 6,5.

Tableau 4 – Indice de spécialisation en informatique et en IA, 2017-21*

	Informatique	IA
U_TORONTO	0,58	0,49
U_MCGILL	0,59	0,48
MAX_PLANCK	0,59	0,62
CEA	0,66	0,46
OXFORD	0,73	0,93
STANFORD	1,05	1,12
RIKEN	1,12	1,45
CNRS	1,15	1,00
FRAUNHOFER	1,81	1,35
MIT	1,90	1,87
NICT	4,45	3,26
ETRI	4,69	2,93
ALAN_TURING	4,98	8,73
INRIA	6,51	5,38
CWI	6,77	5,99
DFKI	9,11	10,66

* Pour l'année 2021, la base de publications est complète à environ 95 %

Source : Base OST, WoS, calculs OST

La spécialisation en IA est corrélée à la spécialisation en informatique, mais certaines institutions sont clairement relativement plus présentes en IA. C'est tout particulièrement le cas de l'institut Alan Turing et dans une moindre mesure de RIKEN, de l'université d'Oxford et de DFKI. Au sein de l'informatique, Inria n'apparaît pas particulièrement engagé dans l'IA, avec un indice de spécialisation inférieur. Le constat est le même pour le CNRS et le CEA.

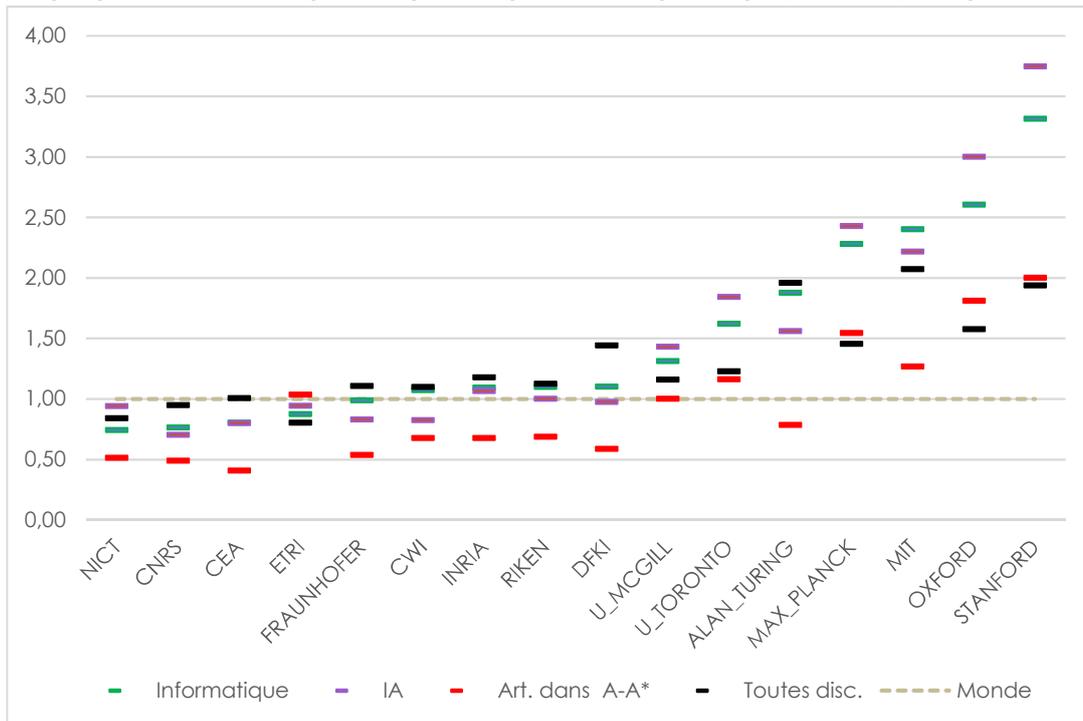
Indicateurs d'impact des publications

Le graphique 4 présente les indices d'impact normalisés pour le total des publications et les trois corpus spécifiques. Il est ordonné sur l'indice d'impact en informatique. Les indices sont assez bien corrélés, même si celui du corpus des articles parus dans les actes de conférences A-A* sont généralement inférieurs. Ce phénomène concernant le domaine de l'informatique correspond d'ailleurs à l'observation plus générale d'une tendance à moins citer les articles d'actes par rapport aux articles parus dans des revues. L'écart entre les deux types de support s'accroît pour les institutions dont les indices d'impact sont les plus élevés. Dans le cas de l'institut Alan Turing, l'indice est même inférieur à la moyenne mondiale (1) pour les actes de conférences alors que les indices pour les articles de revues sont supérieurs à 1,5. La petite taille de l'institution et donc le faible nombre d'articles d'actes de conférences pourrait jouer un rôle dans cette dispersion.

Trois autres institutions ont des indices d'impact supérieurs à la moyenne mondiale pour les articles de revues et inférieurs pour les articles d'actes, mais avec des écarts plus faibles : DFKI, RIKEN et Inria. Pour ce dernier, les articles de revues en informatique et IA ont un indice d'impact 10 % au-dessus de la moyenne mondiale, mais un indice de 30 % inférieur pour les actes de conférences.

Quatre institutions ont des indices d'impact supérieurs à 2 en informatique et IA : Stanford U., Oxford U., le MIT et la société Max Planck. Pour Stanford U., Oxford U et la société Max Planck, ces indices sont sensiblement supérieurs à ceux de l'ensemble de leurs publications toutes disciplines confondues.

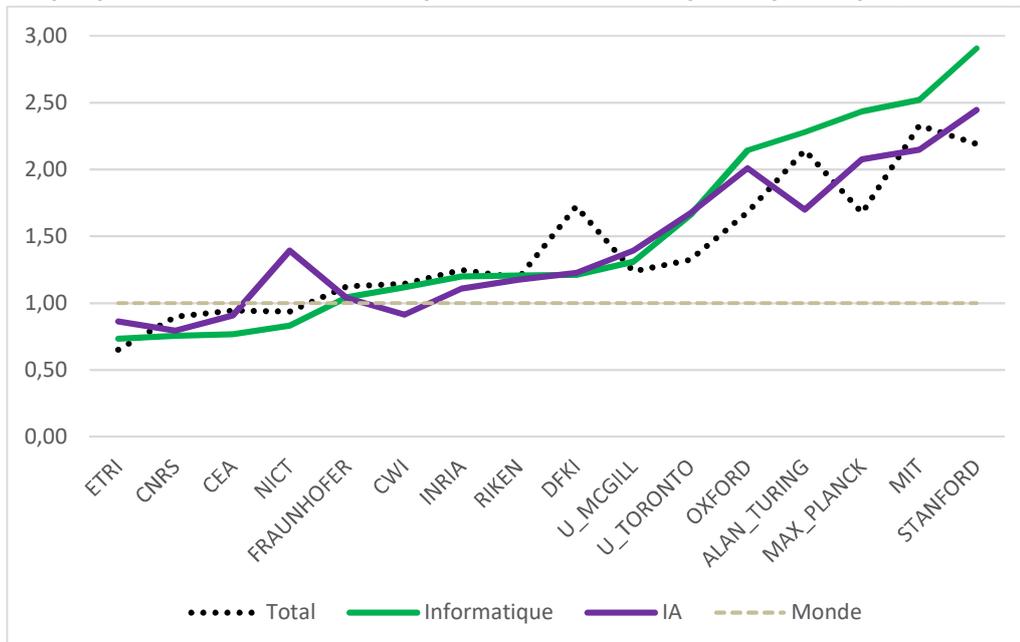
Graphique 4 – Indice d'impact moyen des publications par corpus, 2017-20*, compte fractionnaire



* Pour l'année 2021, la base est complète à environ 95 %, donc l'année 2020 n'est pas complète pour les citations
 Source : Base OST, WoS, calculs OST

Le graphique 5 présente en complément les indices d'activité des institutions dans le décile des publications les plus citées au monde. Il est ordonné sur l'indice d'activité en informatique, comme le graphique 4. On retrouve à peu de choses près le même positionnement des institutions. Les cinq institutions ayant les indices les plus élevés, supérieurs à 2, sont les mêmes pour cet indicateur plus sélectif : Stanford, MIT, Max Planck, Alan Turing et Oxford. De même pour les cinq institutions ayant les indices les plus faibles. L'indicateur pour Inria est un peu plus élevé en informatique, 20 % au-dessus de la moyenne mondiale contre 10 % pour l'impact moyen.

Graphique 5 – Indice d'activité des publications dans le top 10 % par corpus, 2017-20*, compte fractionnaire



* Pour l'année 2021, la base de publications est complète à environ 95 %, donc l'année 2020 n'est pas complète pour les citations
 Source : Base OST, WoS, calculs OST

Annexes

Annexe 1 – Base de données et méthode

La base de données

L'analyse s'appuie sur des données de la base OST qui enrichit la base de données source, le Web of Science (WoS) de Clarivate Analytics, avec des données complémentaires de nomenclature et de repérage institutionnel. La base de l'OST a été actualisée en avril 2022 ; l'année de publication la plus récente disponible est 2021 pour laquelle les données sont en moyenne complètes à 95 %. Les actes de conférences tendent à être entrés en base dans le WoS après les articles de revues et il est probable que la complétude est inférieure en 2021 pour ce type de support de publication.

La base WoS recense les revues scientifiques et les actes de conférences qui répondent à une série de critères en matière de qualité éditoriale comme l'évaluation par les pairs et en matière d'influence académique à l'échelle mondiale. Elle a une bonne couverture pour les disciplines internationalisées et une moins bonne couverture dans certaines disciplines appliquées et dans les disciplines à forte tradition nationale. Néanmoins, la couverture de la base évolue et de nouvelles revues y sont intégrées chaque année suivant le processus de sélection mis en place par Clarivate Analytics.

Les publications de la base OST correspondant aux index SCI-Science Citation Index Expanded, SSCI-Social Sciences Citation Index, A&HCI-Arts & Humanities Citation Index, CPCI-Conference Proceedings Citation Index (S et SSH) sont prises en compte. Les indicateurs ne sont calculés que sur les types de documents articles, reviews, proceedings papers. Les documents pour lesquels manque une partie des informations (catégories du WoS, pays...) comme les publications rétractées ne sont pas pris en compte.

Les corpus de publications

Les indicateurs sont calculés à partir de quatre périmètres de publications scientifiques.

1. Le total des publications, toutes disciplines confondues.
2. La discipline informatique, constituée des catégories suivantes de la nomenclature du WoS : 7 catégories *computer science (artificial intelligence, cybernetics, hardware & architecture, information systems, interdisciplinary applications, software engineering, theory and methods)*, ainsi que *medical informatics, logic, robotics, telecommunications*. La liste est aussi fournie en annexe 2.
3. Un corpus de publications en intelligence artificielle (IA). Il est constitué d'un croisement entre un corpus thématique large en IA et des catégories du WoS. Le corpus large a été construit à partir de mots-clés caractéristiques des techniques d'IA et de leurs applications, validés par des experts du domaine. Les publications retenues au sein de la base de publications de l'OST ont au moins un de ces mots-clés dans le titre et/ou le résumé et/ou les mots-clés auteurs. La restriction est obtenue en exigeant que les publications soient aussi indexées dans les catégories suivantes du WoS : IA, cybernétique, matériel et architecture informatiques, systèmes d'information, ingénierie logicielle, théorie et méthodes informatiques, mathématiques appliquées.
4. Les articles publiés dans les actes de conférences (proceedings) des listes A et A* de l'association CORE <https://www.core.edu.au/conference-portal>. L'OST a vérifié que la base WoS contient une part satisfaisante des actes de ces listes. La présence des actes est précisée à l'annexe 3.

Repérage des adresses des établissements

L'identification des publications auxquelles un établissement français participe repose sur le repérage annuel des adresses d'affiliation organisé par l'OST. Pour les institutions étrangères, l'OST a utilisé le champ *organization enhanced* du WoS après un échange avec l'éditeur sur sa fiabilité.

La production d'Inria est identifiée en consolidant les documents mentionnant Inria dans l'adresse d'affiliation des auteurs et s'appuie sur le repérage effectué par Inria dans le cadre de la production annuelle d'indicateurs par l'OST pour les documents budgétaires du ministère en charge de la recherche⁴. Le processus est le même pour le CNRS et le CEA.

Types de compte des publications

Considérée d'un point de vue institutionnel et géographique, une publication scientifique comporte souvent plusieurs lignes d'adresses, car elle a été produite par des chercheurs d'établissements ou de laboratoires différents. Se pose donc la question de déterminer comment prendre en compte la publication pour chacun des établissements ayant participé ou contribué à sa production. Le compte entier privilégie le point de vue de la participation à la production scientifique : chacune des publications auquel l'institution a contribué est comptabilisée 1 pour cette dernière, quel que soit le nombre total d'adresses d'affiliation des auteurs.

⁴ Indicateurs pour le programme 172 de la Lof.

Le compte fractionnaire privilégie le point de vue de la contribution aux publications scientifiques. Il donne à la contribution un poids égal à la proportion du nombre d'adresses d'affiliation relatives à l'établissement dans l'ensemble des adresses mentionnées sur chaque publication. Par construction, le total des poids attribués aux affiliations institutionnelles de la publication est égal à 1. Les nombres de publications mesurés dans ces conditions sont sommables entre institutions, ce qui n'est pas le cas pour le nombre de publications en compte entier, car il y a des doublons entre institutions. Le compte fractionnaire est appliqué sur le nombre d'adresses d'affiliation, donc ne tient pas compte du phénomène des cotutelles, une unité en cotutelle apparaissant sur une seule ligne d'affiliation.

De même, considérée d'un point de vue disciplinaire, une publication est souvent rattachée à plusieurs catégories disciplinaires et deux logiques de comptage peuvent être adoptées, l'une attribuant entièrement la publication à chacune des catégories, l'autre comptant la publication pour $1/n$, n étant le nombre de catégories où la publication est indexée. Le WoS permet de distinguer plus de 254 *subject categories*, qui sont les mailles disciplinaires les plus fines dans la base utilisée. La plupart du temps, la publication hérite des catégories disciplinaires de la revue ou du support dans lequel elle a été publiée. À titre d'exemple, une publication issue d'une revue indexée dans deux catégories sera, soit comptabilisée pour 1 dans chacune des deux disciplines considérées (compte entier), soit pour une moitié dans chacune (compte fractionnaire disciplinaire).

Lorsqu'on adopte le point de vue d'une institution, il peut être pertinent de privilégier la perspective de la participation, donc le compte entier. C'est ce qui est généralement fait pour comptabiliser les copublications et mesurer la participation d'une institution à une collaboration. Dans d'autres cas, le compte entier peut affecter les possibilités de comparaison, entre disciplines, pays ou institutions. Dans ce cas, c'est le compte fractionnaire qui est préféré.

Le nombre total de publications fourni au début de ce rapport est en compte entier. Par la suite, en dehors des indicateurs de copublication, qui sont calculés en compte entier, les indicateurs sont calculés en compte fractionnaire. Le fractionnement combine les fractionnements sur les adresses d'affiliation et les catégories disciplinaires. Le compte fractionnaire est additif à toutes les échelles et pour tous les niveaux de nomenclature.

Indicateurs

Les indicateurs calculés dans ce rapport sont définis comme suit.

Part nationale / mondiale des publications	Pourcentage des publications de l'acteur, en tenant compte de la somme de ses poids contributifs pour chaque publication. Ce calcul ne peut être effectué qu'en compte fractionnaire, car le compte entier génère des doublons (voir Nombre de publications).
Copublications internationales	Publications co-signées par une institution ou un pays et au moins une institution étrangère. Le calcul est fait en compte entier.
Part des publications d'un domaine pour une institution	Répartition des publications par domaine disciplinaire au sein du corpus des publications identifiées pour chaque établissement (calcul fractionnaire). La répartition correspondante peut être comparée à celle d'une institution ou du monde.
Indice de spécialisation dans un domaine	Rapport entre le pourcentage de publications dans le domaine disciplinaire considéré au sein de l'institution et ce même pourcentage pour une zone de référence (dans ce rapport, il s'agit du monde). Le calcul est fait en compte fractionnaire. Un indice supérieur à 1 indique une spécialisation dans le sous-domaine considéré (respectivement une non spécialisation pour un indice inférieur à 1).
Indice d'impact des publications	L'indice d'impact d'un établissement est la moyenne des scores de citation normalisés de ses publications. La méthode consiste à calculer un score normalisé (par catégorie WoS, type de document et année) pour chaque publication, de façon à obtenir une mesure comparable pour tous les articles. Un indice d'impact supérieur à 1 signifie que les publications de l'établissement sont plus citées en moyenne que les publications du même domaine dans le monde, en considérant le même laps de temps pour les citations (le temps passé jusqu'à la dernière année prise en compte).
Indice d'activité dans la classe des 10 % des documents les plus cités	La part des publications d'un acteur dans la classe des 10 % des documents les plus cités est définie par la proportion de ses publications qui appartiennent aux 10 % des publications les plus citées au monde. Du fait des ex aequo, notamment les documents non cités, la part mondiale diffère légèrement de 10 % en plus ou en moins. L'indice d'activité d'une institution dans la classe des 10 % les plus cités est défini par la part des publications de cet acteur dans la classe des 10 % des documents les plus cités rapportée à la part des publications mondiales dans cette classe. Un acteur dont 15 % des publications appartiennent à la classe des 10 % des documents les plus cités aura un indice d'activité dans cette classe de 1,5. La valeur neutre de l'indice vaut 1.

Annexe 2 – Classification disciplinaire

La classification en onze disciplines dans la base OST résulte d'une agrégation de catégories de recherche établies par Clarivate Analytics. (254). Les onze disciplines sont :

- Biologie appliquée-écologie,
- Biologie fondamentale,
- Chimie,
- Informatique
- Mathématiques,
- Physique,
- Recherche médicale,
- Sciences de l'univers,
- Sciences humaines,
- Sciences pour l'ingénieur,
- Sciences sociales.

Les revues peuvent être rattachées à plusieurs domaines de recherche et éventuellement par agrégation, à plusieurs disciplines. Les articles de revues multidisciplinaires (Nature, PNAS US et Science notamment) sont distribués dans les différentes disciplines en fonction de leurs sujets. Dans les 3 premiers chapitres de l'étude, les indicateurs sont calculés pour les disciplines.

Le tableau suivant fournit la correspondance entre une discipline et les catégories de recherche scientifique qu'elle agrège pour les disciplines les plus importantes pour Inria.

Grandes disciplines	Libellés des domaines de recherche dans le WoS	
INFORMATIQUE	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE COMPUTER SCIENCE, CYBERNETICS COMPUTER SCIENCE, HARDWARE & ARCHITECTURE COMPUTER SCIENCE, INFORMATION SYSTEMS COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS COMPUTER SCIENCE, SOFTWARE ENGINEERING	COMPUTER SCIENCE, THEORY & METHODS MEDICAL INFORMATICS LOGIC ROBOTICS TELECOMMUNICATIONS
MATHEMATIQUES	MATHEMATICS MATHEMATICS, APPLIED	STATISTICS & PROBABILITY MATHEMATICS, INTERDISCIPLINARY APPLICATIONS
SCIENCES POUR L'INGENIEUR	AUTOMATION & CONTROL SYSTEMS CONSTRUCTION & BUILDING TECHNOLOGY ENERGY & FUELS ENGINEERING, AEROSPACE ENGINEERING, CHEMICAL ENGINEERING, CIVIL ENGINEERING, ELECTRICAL & ELECTRONIC ENGINEERING, INDUSTRIAL ENGINEERING, MANUFACTURING ENGINEERING, MARINE ENGINEERING, MECHANICAL ENGINEERING, MULTIDISCIPLINARY	ENGINEERING, OCEAN ENGINEERING, PETROLEUM IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY MECHANICS METALLURGY & METALLURGICAL ENGINEERING MINING & MINERAL PROCESSING NUCLEAR SCIENCE & TECHNOLOGY OPERATIONS RESEARCH & MANAGEMENT SCIENCE REMOTE SENSING THERMODYNAMICS TRANSPORTATION SCIENCE & TECHNOLOGY
PHYSIQUE	ACOUSTICS INSTRUMENTS & INSTRUMENTATION OPTICS PHYSICS, APPLIED PHYSICS, ATOMIC, MOLECULAR & CHEMICAL PHYSICS, CONDENSED MATTER	PHYSICS, FLUIDS & PLASMAS PHYSICS, MATHEMATICAL PHYSICS, MULTIDISCIPLINARY PHYSICS, NUCLEAR PHYSICS, PARTICLES & FIELDS SPECTROSCOPY

Annexe 3 – Couverture par le WoS des actes A et A* de CORE

La liste des conférences CORE a été obtenue depuis le portail [CORE](#)⁵. La liste des conférences a été limitée aux conférences classées A et A* et aux *Fields of Research* du groupe 46 *Information and Computing Sciences* de l'[ANZSRC-2020](#)⁶. Au total, 246 conférences correspondent à ces critères.

Afin de retrouver dans la base WoS les communications associées à ces conférences CORE A - A*. Deux approches ont été utilisées :

- Recherche du nom de la conférence (par mots-clés) dans le support des publications. Cette approche fonctionne lorsque la conférence possède son propre support pour la publication des communications. Les requêtes ont été adaptées de manière à exclure les faux négatifs.
- Recherche du nom de la conférence dans le champ *Conférence* du WoS. Cette approche cible notamment les conférences qui publient leurs communications dans un numéro spécial d'une revue.

Ces approches ont permis de récupérer des communications pour 226 conférences CORE sur les 246 classées A et A*, soit 91,8% (Tableau A1). Pour certaines de ces conférences, aucune communication n'a été trouvée pour certaines éditions, sans que l'on sache si cela est dû à la non tenue de la conférence cette année-là, ou à la non indexation de ces communications par le WoS. La liste des 20 conférences qui n'ont ramené aucune communication sur la période 2017-21 est fournie par le Tableau A2.

Tableau A1 : Couverture des conférences CORE A et A* par le WoS

Trouvées dans le WoS	CORE A et A*
OUI	226
NON	20
% non trouvées	8,13%

Tableau A2 : Liste des conférences non trouvées dans le WoS entre 2017 et 2021

Conférences CORE A/A* non trouvées dans le WoS

ACM CONFERENCE ON OBJECT ORIENTED PROGRAMMING SYSTEMS LANGUAGES AND APPLICATIONS
ANNUAL MEETING OF THE COGNITIVE SCIENCE SOCIETY
NORTH AMERICAN ASSOCIATION FOR COMPUTATIONAL LINGUISTICS
PRIVACY ENHANCING TECHNOLOGIES SYMPOSIUM (WAS INTERNATIONAL WORKSHOP OF PRIVACY ENHANCING TECHNOLOGIES)
EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING
INTERNATIONAL CONFERENCE ON EMBEDDED WIRELESS SYSTEMS AND NETWORKS (WASEUROPEAN CONFERENCE ON WIRELESS SENSOR NETWORKS)
MEASUREMENT AND MODELING OF COMPUTER SYSTEMS
UNIX SYMPOSIUM ON INTERNET TECHNOLOGIES
IEEE INTERNATIONAL SYMPOSIUM ON ARTIFICIAL LIFE
LOGIC PROGRAMMING AND AUTOMATED REASONING
CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING
THEORETICAL ASPECTS OF RATIONALITY AND KNOWLEDGE
EUROMICRO CONFERENCE ON REAL-TIME SYSTEMS
INTERNATIONAL WORKSHOP ON APPROXIMATION ALGORITHMS FOR COMBINATORIAL OPTIMIZATION PROBLEMS
COMPUTER SUPPORTED COLLABORATIVE LEARNING
INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS
WORKSHOP ON ALGORITHM ENGINEERING AND EXPERIMENTS
ADVANCES IN MODAL LOGIC
CONFERENCE ON INNOVATIVE DATA SYSTEMS RESEARCH
INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS

⁵ <http://portal.core.edu.au/conf-ranks/>

⁶ Australian and New Zealand Standard Research Classification (ANZSRC).
<https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-research-classification-anzsrc/2020>

On obtient entre 22 000 et 30 500 communications par année, entre 2017 et 2021 (Tableau A3). Le nombre de communications collectées croit d'abord entre 2017 et 2019, puis diminue jusqu'à 2021, ce qui coïncide avec la pandémie de covid-19 (voir aussi le corps du texte sur l'évolution temporelle des publications en général).

Tableau A3 : Nombre de communications des conférences CORE A et A*, par année

Année	Nb d'actes de colloque
2017	25 959
2018	28 642
2019	30 461
2020	23 035
2021	22 070
Total	130 167

Annexe 4 – Les institutions comparées

Cette annexe présente de façon succincte et par ordre alphabétique les institutions pour lesquelles les publications scientifiques ont été analysées dans le rapport. Il convient de considérer avec précaution les informations fournies qui proviennent de différentes sources accessibles en ligne ; elles n'ont pas été validées par les institutions considérées et les données sont parfois incomplètes. Ces données ne sont donc pas toujours comparables.

ALAN TURING INSTITUTE

<https://www.turing.ac.uk/>

L'Alan Turing Institute est un institut de recherche britannique spécialisé dans le domaine de la science des données, de l'intelligence artificielle et de la recherche en informatique.

Fondé en 2015, l'Alan Turing Institute joue un rôle essentiel dans le développement de l'intelligence artificielle éthique et de l'analyse de données, tout en travaillant sur des applications pratiques dans des domaines tels que la santé, les sciences sociales, l'économie, l'environnement et bien d'autres.

En 2021, l'Alan Turing Institute employait environ **1 000 personnes**, y compris des chercheurs postdoctoraux, des doctorants, du personnel de soutien et des chercheurs affiliés, et disposait d'un budget de **59 M€**.

CEA

<https://www.cea.fr/>

Le Commissariat à l'énergie atomique et aux énergies alternatives (CEA) est un établissement public français sous la tutelle des ministres chargés de l'énergie, de la recherche, de l'industrie et de la défense. Le CEA est principalement axé sur la recherche scientifique et technologique dans les domaines de l'énergie nucléaire, des technologies de défense, de la recherche fondamentale en physique et en chimie, ainsi que sur les énergies alternatives et les technologies environnementales.

En 2022, le **CEA comptait 21 433 employés** et disposait d'un budget d'environ **5 800 M€**.

Les sciences et technologies du numérique sont essentiellement développées au sein du Laboratoire d'intégration des systèmes et des technologies (CEA-List) de la Direction de la recherche technologique du CEA. En 2018, l'effectif du CEA-List était de **842 chercheurs, personnels de soutien et doctorants** (107).

CNRS

<https://www.cnrs.fr/fr>

Le Centre national de la recherche scientifique (CNRS) est un organisme public de recherche pluridisciplinaire créé en 1939. Le CNRS est structuré en 10 instituts dont l'Institut des sciences de l'information et de leurs interactions (INS2I) au sein duquel sont principalement menées les recherches en informatique et intelligence artificielle. L'ensemble des personnels des 40 unités de recherche de l'INS2I (unités mixtes avec des universités et éventuellement d'autres organismes de recherche) comprend près de **10 000 chercheurs, enseignants-chercheurs, personnels de soutien et doctorants** dont **650 chercheurs et 400 personnels de soutien employés par le CNRS et 3 550 doctorants**.

Le CNRS compte 31 900 employés, dont 11 400 chercheurs, et dispose d'un budget d'environ **3 700 M€**.

CWI

<https://www.cwi.nl/en/>

Le *Centrum voor Wiskunde en Informatica* (CWI) est un centre de recherche national en mathématiques et informatique à Amsterdam, aux Pays-Bas, il est majoritairement financé par l'organisation néerlandaise pour la recherche scientifique. Le centre se concentre principalement sur quatre domaines de recherche fondamentale : Algorithmes, Données et systèmes intelligents, Cryptographie et sécurité, et Informatique quantique.

En 2022, le CWI employait **227 personnes**, dont 85 doctorants, et disposait d'un budget de **21 M€**.

DFKI

<https://www.dfki.de/en/web>

Le *Deutsches Forschungszentrum für Künstliche Intelligenz* (DFKI) est un institut de recherche en intelligence artificielle (IA) basé en Allemagne. Fondé en 1988, le DFKI est l'un des plus grands centres de recherche en IA en Europe. Il joue un rôle clé dans le développement et l'application de technologies avancées d'IA dans divers domaines, tels que la robotique, le traitement du langage naturel, la vision par ordinateur et l'apprentissage automatique. Le DFKI collabore étroitement avec l'industrie et les universités pour promouvoir l'innovation et la recherche de pointe en IA.

En 2021, le DFKI employait environ **1 500 personnes**, dont 610 doctorants, et disposait d'un budget de **76 M€**.

ETRI

<https://www.etri.re.kr/eng/main/main.etri>

L'*Electronics and Telecommunications Research Institute* (ETRI) est un institut de recherche basé en Corée du Sud, spécialisé dans les domaines de l'électronique, des télécommunications et de l'informatique. Fondé en 1976, ETRI développe des technologies avancées, notamment dans les domaines des communications sans fil,

de l'intelligence artificielle, de la robotique, de l'informatique quantique et de l'internet des objets. ETRI collabore activement avec des entreprises, des universités et d'autres instituts de recherche nationaux et internationaux. En 2020, l'ETRI employait environ **2 300 personnes**, et disposait d'un budget de **435 M€**.

FRAUNHOFER

<https://www.fraunhofer.de/en/institutes/institutes-and-research-establishments-in-germany/fraunhofer-groups/ict-technology.html>

Le *Fraunhofer-Gesellschaft* est un centre de recherche allemand créé en 1949 spécialisé dans la recherche en sciences appliquées, il regroupe 67 instituts. Plusieurs d'entre eux relèvent des champs disciplinaires de l'informatique et de l'intelligence artificielle. Le Fraunhofer ICT Group (information and communications technologies) fédère la vingtaine d'instituts impliqués dans les ICT, il emploie environ **4 300 personnes** et dispose d'un budget d'environ **275 M€**.

INRIA

<https://www.inria.fr/fr>

L'Institut national de recherche en sciences et technologies du numérique (Inria) est un établissement public de recherche créé en 1967 et sous tutelles des ministres en charge de la recherche et de l'industrie. Inria est le seul organisme français exclusivement dédié au numérique. Ses missions couvrent un large spectre en appui des politiques publiques de la recherche, de la formation et de l'innovation dans le numérique.

En 2022, sur les **4 742 personnes** travaillant pour les missions d'Inria, **2 805 étaient employées par Inria**, dont 619 doctorants. Le budget d'Inria était de **260 M€** en 2022.

MAX PLANCK

<https://www.mpg.de/en>

La Société Max Planck (*Max-Planck-Gesellschaft*) est une organisation de recherche scientifique fondée en 1948 et basée en Allemagne. Cette société à but non lucratif regroupe un réseau de plus de 80 instituts de recherche (MPI) répartis dans toute l'Allemagne, couvrant un large éventail de disciplines scientifiques, de la physique à la biologie en passant par la chimie, les sciences humaines et sociales.

Plusieurs instituts relèvent des champs disciplinaires de l'informatique au sens large, notamment le *MPI für Informatik* (MPI-INF), le *MPI für Software system* (MPI-SWS) et le *MPI für intelligenten Systemen* (MPI-IS).

L'ensemble de la société Max-Planck comprend 20 900 employés, dont 6 700 chercheurs, et dispose d'un budget d'environ **2 600 M€**.

MIT

<https://web.mit.edu/>

Massachusetts Institute of Technology (MIT), située à Cambridge, Massachusetts, aux États-Unis. Fondé en 1861, le MIT est réputé pour son excellence dans les domaines de la science, de la technologie, de l'ingénierie et des mathématiques. La recherche au MIT est structurée de manière à favoriser l'interdisciplinarité, l'innovation et la collaboration entre différentes facultés et départements, au sein de *centers, labs* et *programs*.

Les principales structures de recherche en informatique et intelligence artificielle (IA) au MIT sont le *Computer Science and Artificial Intelligence Laboratory* (CSAIL), l'un des plus grands laboratoires de recherche en informatique au monde ; le *MIT-IBM Watson AI Lab* ; le *Laboratory for Information and Decision Systems* (LIDS) qui se concentre sur la recherche en théorie de l'information, en systèmes de communication, en contrôle et en prise de décision ; le *Computer Science Theory Group* qui se penche sur les fondements théoriques de l'informatique, pour le développement de nouvelles approches en IA et en informatique ; le *MIT Quest for Intelligence* qui vise à comprendre l'intelligence sous toutes ses formes).

Le MIT compte 16 300 employés, dont 4 000 chercheurs, et dispose d'un budget d'environ **4 200 M€**.

NICT

<https://www.nict.go.jp/en/index.html>

Le *National Institute of Information and Communications Technology* (NICT) est le principal institut de recherche national du Japon dans le domaine de l'information et de la communication. Le NICT est une institution administrative indépendante créée en 2004, elle a pour mission de mener des activités de recherche et de développement dans le domaine des technologies de l'information et de la communication.

En 2020, le NICT employait environ **1 195 personnes** et disposait d'un budget de **231 M€**.

L'université d'OXFORD

<https://www.ox.ac.uk/>

L'Université d'Oxford créée en 1167 est la plus ancienne université britannique. L'université comprend une variété de collèges et de départements académiques, couvrant un large éventail de disciplines, de la littérature à la science, en passant par la médecine et le droit.

Les principales structures de recherche en informatique et intelligence artificielle (IA) sont le *Department of Computer Science* qui abrite de nombreux groupes de recherche travaillant sur une variété de sujets, tels que l'apprentissage automatique, la vision par ordinateur, la sécurité informatique, la bioinformatique et les systèmes distribués ; l'*Oxford Internet Institute* qui se concentre sur l'étude des implications sociales, politiques et économiques d'Internet et des technologies numériques ; l'*Oxford Robotics Institute* ; le *Future of Humanity*

Institute qui s'intéresse particulièrement à l'impact à long terme de la technologie sur l'humanité, y compris les questions liées à l'IA et à la sécurité informatique ; l'Université d'Oxford a établi des partenariats de recherche avec DeepMind, entreprise de recherche en intelligence artificielle.

L'université d'Oxford compte **14 900 employés, dont 7 000 chercheurs**, et dispose d'un budget de **3 190 M€**.

RIKEN

<https://www.riken.jp/en/about/>

Le Riken, ou l'Institut de recherche en sciences physiques et chimiques du Japon (*Rikagaku Kenkyūjo*) est l'une des principales organisations de recherche scientifique au Japon. Fondé en 1917, le Riken est un centre de recherche multidisciplinaire de renommée internationale qui se consacre à la recherche avancée dans une variété de domaines scientifiques, notamment la physique, la chimie, la biologie, l'informatique et l'intelligence artificielle.

Les principales structures de recherche en informatique et intelligence artificielle sont le *Riken Center for Advanced Intelligence Project* qui se concentre sur la recherche fondamentale et appliquée en IA, y compris l'apprentissage automatique, la vision par ordinateur, le traitement du langage naturel et d'autres domaines connexes ; le *Riken Center for Computational Science* qui joue un rôle crucial en particulier en ce qui concerne les supercalculateurs et les technologies de calcul de pointe nécessaires pour l'IA ; le *Riken Center for Advanced Photonics* qui se penche sur des domaines tels que l'imagerie médicale avancée et les systèmes optiques avancés, qui ont des applications potentielles en IA, en particulier en ce qui concerne le traitement d'images.

L'ensemble du Riken compte **3 000 employés** et dispose d'un budget d'environ **750 M€**.

L'université STANFORD

<https://www.stanford.edu/about/>

L'Université Stanford, est une université de recherche prestigieuse réputée pour sa proximité avec la *Silicon Valley*, ce qui en fait un centre majeur d'innovation technologique et d'entrepreneuriat. Elle abrite également des centres de recherche de premier plan dans des domaines tels que l'intelligence artificielle, la biotechnologie et les sciences de l'information.

Les principales structures de recherche de Stanford dans les domaines de l'informatique et de l'intelligence artificielle sont le *Stanford Artificial Intelligence Laboratory*, l'un des plus influents laboratoires en IA au monde ; le *Stanford Vision and Learning Lab* qui travaille sur des projets liés à la vision artificielle ; le *Stanford Natural Language Processing Group* spécialisé dans le traitement du langage naturel et le dialogue homme-machine ; le *Stanford Human-Centered Artificial Intelligence Institute* créé pour promouvoir la recherche en IA centrée sur l'humain, y compris l'éthique, la gouvernance et l'impact sociétal de l'IA ; le *Stanford Center for Artificial Intelligence in Medicine and Imaging*.

En 2021, **Stanford** comprenait **1 3800 employés** et disposait d'un budget d'environ **6 860 M€**.

L'université MCGILL

<https://www.mcgill.ca/about/fr>

L'Université McGill, fondée en 1821, elle est l'une des plus anciennes universités du Canada. McGill est réputée pour son excellence académique. McGill dispose de plusieurs structures de recherche en informatique et en intelligence artificielle (IA) dont les principales sont le *McGill Reasoning and Learning Lab* qui se concentre sur la recherche en apprentissage automatique et en traitement automatique du langage naturel ; le *McGill Centre for Intelligent Machines* qui se penche sur divers aspects de l'intelligence artificielle, y compris la robotique, la vision par ordinateur, le traitement du langage naturel et la recherche fondamentale en IA ; le *McGill Data Science and Artificial Intelligence Lab*, l'un des centres de recherche en IA les plus prestigieux au Canada ; le *McGill Human-Computer Interaction Lab* qui se consacre à la recherche sur l'interaction homme-machine.

En 2021, l'ensemble de **l'université Mc Gill** comprenait environ **10000 employés** et disposait d'un budget d'environ **1000 M€**.

L'université de TORONTO

<https://www.utoronto.ca/about-u-of-t>

L'Université de Toronto, fondée en 1827, est l'une des principales universités du Canada. L'université de Toronto est un acteur majeur dans le domaine de l'informatique et de l'intelligence artificielle, ses principales structures de recherche dans ces domaines sont le *Vector Institute for Artificial Intelligence*, l'un des principaux centres de recherche en IA au Canada ; le *Machine Learning Group*, qui se concentre sur la recherche en apprentissage automatique, apprentissage profond, traitement du langage naturel, vision par ordinateur et en statistiques ; le *Toronto AI Lab* qui rassemble des chercheurs de l'université et d'entreprises pour collaborer sur des projets de recherche ; le *Distributed Systems Group* spécialisé dans la recherche sur les systèmes informatiques distribués ; le *Citizen Lab*, laboratoire de recherche interdisciplinaire, qui se penche sur la sécurité en ligne, la censure sur Internet et la surveillance numérique ; le *Dynamic Graphics Project* qui se concentre sur la création et l'animation d'images en 3D ; le *Deep Learning Research Group* spécialisé dans les aspects fondamentaux de l'apprentissage profond.

En 2021, **l'université de Toronto** comptait environ **24 500 employés** et disposait d'un budget d'environ **2 200 M€**.



2 rue Albert Einstein
75013 Paris, France
T. 33 (0)1 55 55 60 10
hceres.fr

Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur