

Séance du 5 novembre 2024

Conférence-débat de Paul Groth avec Patrick Maestro et Michael Matlosz

INTELLIGENCE ARTIFICIELLE POUR LA RECHERCHE ET LA TECHNOLOGIE : UN REGARD EUROPÉEN

Paul Groth présente les enjeux des récentes avancées de l'intelligence artificielle (IA) pour la science. Il rappelle l'attribution en 2024 des prix Nobel de physique et de chimie à des recherches liées à l'IA et présente des applications concrètes, comme AlphaFold, ainsi que des innovations de moindre maturité mais tout aussi spectaculaires (des réseaux neuronaux sélectionnant les molécules ayant un potentiel antibiotique élevé ; la découverte de nouveaux matériaux potentiels grâce à l'apprentissage profond combiné à la théorie de la fonctionnelle de la densité ; un apprentissage par renforcement susceptible d'être utilisé pour contrôler la stabilité du plasma d'ITER...). Le conférencier évoque aussi les prémises d'une vision futuriste où des systèmes d'IA pourraient réaliser de la recherche de façon autonome. Il dresse un paysage de la recherche dominé par des grandes entreprises, les États-Unis et la Chine. Il aborde ensuite les risques de l'IA : manque de transparence, biais et données de mauvaise qualité, désinformation, inégalités dans la recherche, manque de lignes directrices, utilisations malveillantes. Pour y répondre, le rapport rendu par SAPEA fournit quatre recommandations : des cadres évolutifs, des normes de qualité assurant la reproductibilité, des investissements massifs dans l'infrastructure, et le respect des valeurs fondamentales de l'Europe. En conclusion, Paul Groth insiste sur le fait que l'IA est un accélérateur qui peut améliorer ou aggraver les évolutions en cours selon ses utilisations. Précédée d'une présentation de SAPEA par Patrick Maestro et d'une introduction par Michael Matlosz, cette conférence plaide pour que l'UE tire pleinement parti des ressources de l'IA pour la recherche scientifique et technologique.

Sommaire

Présentation de SAPEA	2
Propos introductifs	2
Exposé de Paul Groth	3
Discussion	9

Paul Groth est professeur de science algorithmique des données à l'université d'Amsterdam, où il dirige le laboratoire d'ingénierie des données intelligentes (INDElab). Docteur en informatique de l'université de Southampton (2007), il a effectué des recherches à l'université de Californie du Sud, à la Vrije Universiteit Amsterdam et à Elsevier Labs. Celles-ci portent sur les systèmes intelligents permettant de traiter des quantités massives de connaissances contextualisées diverses, avec un accent particulier mis sur la provenance et l'intégration des données et le partage des connaissances. Directeur scientifique du Data Science Center de l'UvA, il est aussi co-directeur scientifique de deux laboratoires de l'Innovation Center for Artificial Intelligence (ICAI) : le laboratoire AI for Retail (AIR) et le laboratoire Discovery. Il est co-auteur de *Provenance. An Introduction to PROV* (Springer 2013) et de *The Semantic Web Primer* (3rd Edition, MIT Press, 2012) ainsi que de nombreux articles académiques. Il était membre du groupe de travail de SAPEA qui vient de rendre un rapport sur les enjeux de l'IA pour la science.

Patrick Maestro est un chimiste et expert des matériaux qui s'intéresse particulièrement aux produits biosourcés avec les composés à base d'oxydes de terre rares que l'on retrouve aujourd'hui dans les lampes à basse consommation LED ou dans les catalyseurs de postcombustion des voitures. Directeur scientifique de Rhodia puis de Solvay, auteur d'une quinzaine de brevets, il s'est surtout consacré à faire collaborer recherche publique et monde industriel. Il a ainsi contribué à la création de plusieurs unités mixtes du CNRS en partenariat avec Solvay telles que le laboratoire « Polymères et matériaux avancés », le « laboratoire du futur » et le laboratoire E2P2L à Shanghai. Il est également membre de l'Académie des technologies depuis 2014 et sert aujourd'hui en tant que secrétaire général d'Euro-CASE, association des académies de technologies et d'ingénierie d'Europe (faisant partie de SAPEA).

Michael Matlosz est ingénieur chimiste diplômé de l'Institut de Technologie de l'État du New Jersey et docteur en génie chimique de l'université de Californie à Berkeley. Il a démarré sa carrière de chercheur à l'École Polytechnique Fédérale de Lausanne en Suisse. Il a ensuite été professeur et directeur de l'École nationale supérieure des industries chimiques de Nancy (ENSIC-Nancy), puis président directeur général de l'Agence nationale de la recherche (ANR). Il est aujourd'hui président de l'association EuroScience et professeur à l'ENSIC- Nancy. Il est membre de l'Académie des technologies depuis 2011 et sert aujourd'hui en tant que délégué aux relations internationales pour un mandat de deux ans.



Présentation de SAPEA

par Patrick Maestro

SAPEA (*Science Advice for Policy by European Academies*) est un projet européen dont le rôle est de fournir des avis au Groupe des conseillers scientifiques en chef de la Commission européenne, à travers principalement des rapports (une quinzaine depuis 2016). Il regroupe cinq réseaux d'Académies, dont Euro-CASE, c'est à dire le Conseil européen des Académies des sciences appliquées, des technologies et de l'ingénierie, qui est un réseau de 23 Académies regroupant autour de 6000 experts, dont fait partie l'Académie des Technologies. Les quatre autres réseaux sont Academia Europaea, ALLEA, FEAM et YASAS. SAPEA fonctionne surtout à la demande de la Commission mais peut aussi prendre l'initiative de rédiger un rapport. L'un des derniers rapports rendus est celui dont Paul Groth va vous entretenir.

Propos introductifs

par Michael Matlosz

La présentation du conférencier s'intitule « L'adoption réussie et opportune de l'Intelligence artificielle dans la science au sein de l'Union européenne », qui est aussi le titre du rapport rendu par SAPEA en avril 2024 à la Commission européenne. Je tiens à remercier le professeur Paul Groth, l'un de ses rédacteurs, de venir nous le présenter. Paul Groth est professeur en science des données à l'université d'Amsterdam. Il traitera aussi des récents développements dans le domaine de l'Intelligence artificielle (IA) pour la recherche, notamment aux États-Unis, un pays dont nous sommes tous deux originaires.



Exposé de Paul Groth

Cette présentation comportera deux parties. D'abord, je traiterai des développements récents de l'IA pour la science, puis j'en viendrai aux politiques européennes et à ce qui peut être fait dans ce cadre.

L'adoption de l'IA dans la science

Cette année, l'IA a gagné deux prix Nobel : le prix Nobel de physique 2024 (attribué à John J. Hopfield et Geoffrey E. Hinton, qui ont développé l'approche de l'IA par « réseaux de neurones ») et, ce qui est plus intéressant encore, le prix Nobel de chimie, décerné à AlphaFold, un système d'IA permettant de passer du séquençage des protéines à leur structure en trois dimensions. Au-delà de ces symboles, la manière accrue dont l'IA contribue à la recherche scientifique est rendue encore plus manifeste par la croissance du taux de publications scientifiques recourant à l'IA : on l'observe en physique, où ce pourcentage décolle à partir de 2020, comme dans les sciences de la vie, les sciences médicales et même les sciences sociales et les humanités. Une autre manière de mesurer à grande échelle l'adoption de l'IA est de demander aux chercheurs s'ils ont recours à l'IA. Entre 2020 et 2021, le taux de réponses positives en ingénierie passe ainsi de 19 à 21 %, en sciences des matériaux de 0 à 18 %, ce qui témoigne d'une transformation radicale des méthodes, tout comme en chimie où le taux passe de 2 à 19 %.

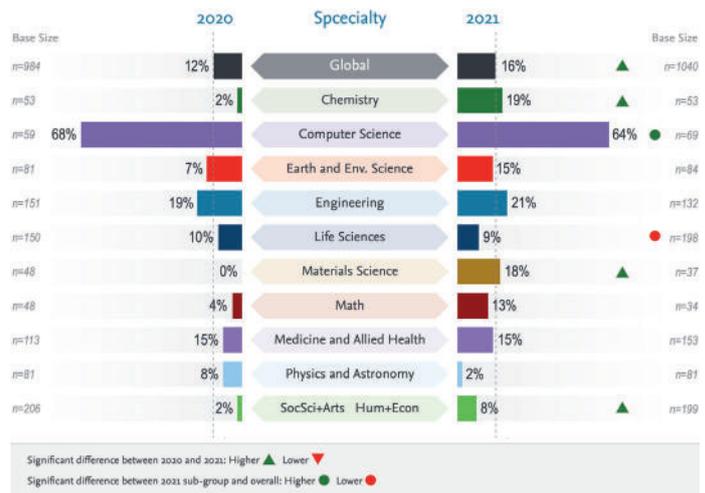


Figure 1 : Proportion des recherches ayant significativement recours à l'IA selon les disciplines

La figure 1 indique les chiffres de 2021. Depuis ces tendances se sont accentuées. On peut mesurer aussi l'accroissement du taux de projets ERC comportant de l'IA. En 2021, sa proportion était déjà assez considérable : 14 % en sciences sociales, 11 % en sciences de la vie, et 17 % en sciences physiques. Tous ces éléments témoignent de la croissance des usages de l'IA dans les sciences. En outre, ce qui confirme la perception d'une tendance de fond, c'est que chacun en fait l'expérience dans sa vie quotidienne. Que vous utilisiez un agent conversationnel, un moteur de recherche ou soyez confronté au système de recommandation d'un réseau social, vous avez adopté un usage de l'IA dans votre vie quotidienne, et cette familiarité croissante ne peut que favoriser la pénétration de l'IA dans la science.

L'enjeu de l'IA en science dans l'Union européenne

Tous ces éléments poussaient à déclencher le « mécanisme de conseil scientifique » (*Scientific Advice Mechanism*) de la Commission européenne, avant même que ne soit formulée une demande à SAPEA, à poser la question « comment accélérer l'adoption opportune de l'IA pour la science ? » En arrière-plan, il y a la conscience que les progrès ultra-rapides de l'IA sont essentiels au progrès technologique, mais avec différentes approches de sa régulation en Chine, aux États-Unis et en Europe. Du point de vue de l'UE, l'enjeu est de savoir comment maintenir sa compétitivité scientifique face aux autres acteurs globaux. Quelles recommandations pour l'IA dans la science ? Avant de les détailler, il faut souligner le nombre et la diversité des applications concrètes de l'IA en science : elle peut être utilisée pour concevoir des expériences, pour améliorer

des calculs mathématiques et théoriques, pour collecter des informations en sciences sociales, pour effectuer plus vite des simulations, pour communiquer et mieux présenter des résultats, etc. Voir Figure 2.



Figure 2 : Abondance des usages de l'IA en science. Source University of Amsterdam (Ghosh A., et al. 2023).

Un cas exemplaire : AlphaFold

AlphaFold est un modèle d'apprentissage profond qui s'inspire des mêmes techniques que celles qui sous-tendent les grands modèles de langage (LLM), c'est-à-dire l'architecture « transformeur » d'apprentissage profond, qui est typique des architectures sous-jacentes aux systèmes d'apprentissage profond de l'IA.

Voici les résultats de l'article original d'AlphaFold. Avec, en bleu, tout ce qui est prédit par AlphaFold, et en vert, le résultat expérimental. Vous pouvez voir à quel point les structures protéiques prédites à partir des séquences d'acides aminés correspondent aux résultats expérimentaux. Ce qui est encore plus intéressant, c'est que pour n'importe quelle autre séquence d'acides aminés, vous pouvez prédire une structure protéique susceptible de se produire dans le monde réel. Et la figure 3-a vous montre la comparaison avec le niveau de performance de tous les anciens algorithmes utilisés pour résoudre le problème du repliement des protéines. La compétition de longue haleine pour résoudre le problème du repliement des protéines a été remportée par AlphaFold : l'IA est à l'origine d'une augmentation spectaculaire de la précision et donc de la capacité à réaliser cette partie fondamentale de la biologie, qui est cruciale pour la découverte de médicaments, et aussi de nouveaux matériaux.

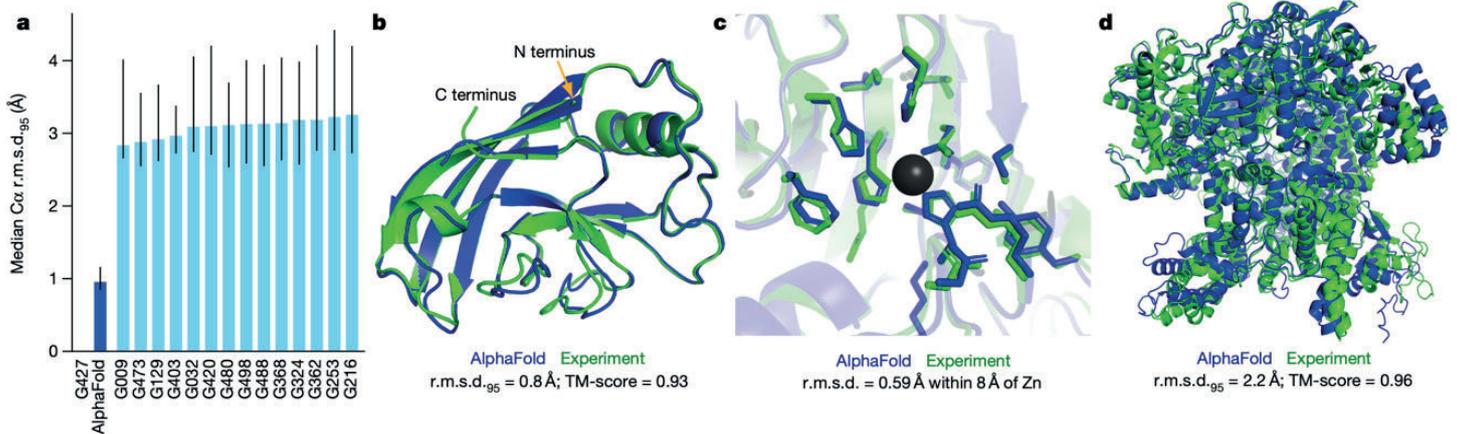


Figure 3 : AlphaFold. Source University of Amsterdam (Jumper J., Evans R., Pritzel A., et al. 2021).

Les antibiotiques du futur

Nous pouvons également étudier ce phénomène dans d'autres domaines de la biologie. Voici un article qui vient d'être publié cette année sur la découverte d'une nouvelle classe structurale d'antibiotiques. Ces chercheurs ont pris environ 12 millions de composés qu'il fallait « filtrer » pour voir s'ils ont un comportement antibiotique à l'égard d'une bactérie donnée. Ils utilisent alors des réseaux neuronaux (transformeur) basés sur des graphes qui sélectionnent les molécules en fonction du score de prédiction antibiotique, ainsi que de leur cytotoxicité. Ils essayent de prédire s'il

ya un sous-ensemble favorable. Ils sont ainsi parvenus à sélectionner environ 1200 nouveaux composés antibiotiques potentiels. Vous voyez ici une visualisation de l'espace chimique des milliers de milliards de combinaisons possibles que l'on essaye de réduire en fonction de leur efficacité antibiotique. En vert sont représentées, sur la Figure 4, les molécules aux scores élevés de prédiction antibiotique (et au score faible de cytotoxicité). Ensuite, on a testé un ou deux composés sur des modèles (souris) et ils ont une capacité substantielle à être des antibiotiques.

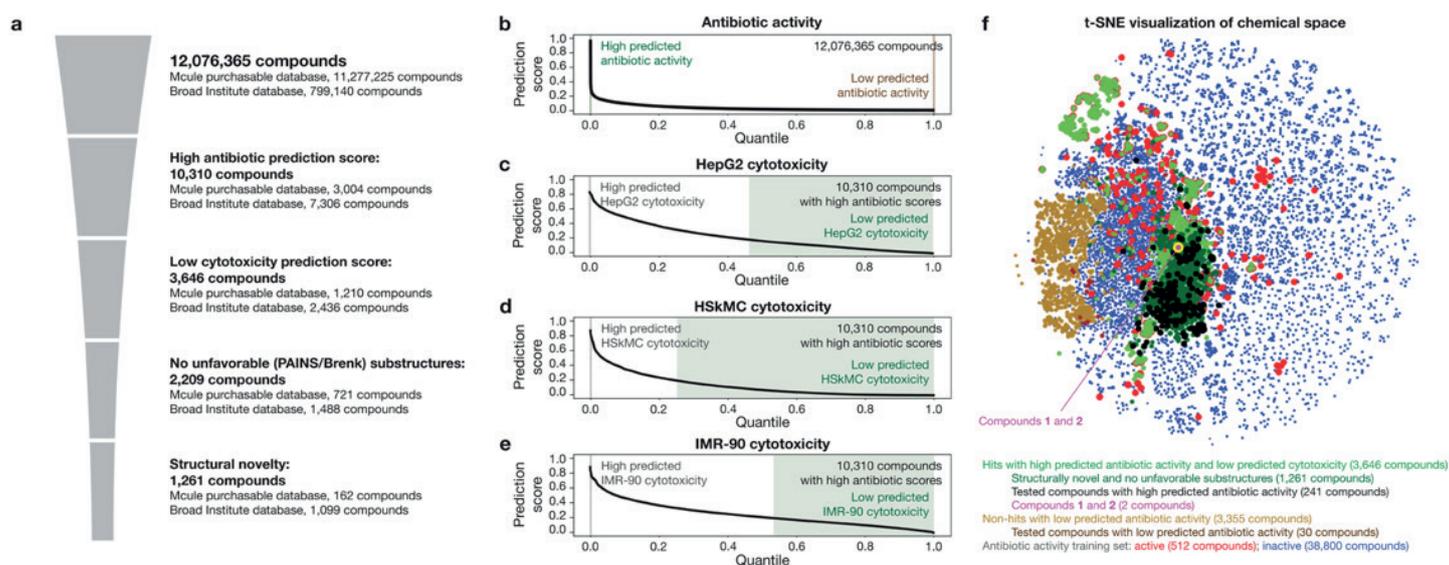


Figure 4: Découvrir de nouveaux antibiotiques. Source Wong F., Zheng E.J., Valeri J.A., et al. 2024.

Inventer de nouveaux matériaux

Dans un autre domaine, l'IA sert à découvrir de nouveaux matériaux potentiels à faible consommation d'énergie. Voici comment fonctionne le « pipeline » : il y a d'une part, encore une fois, un réseau neuronal prédisant les structures chimiques potentielles à partir du graphe du matériau ; et d'autre part, en parallèle, une recherche aléatoire, examinant les propriétés chimiques du matériau recherché. Au point de rencontre se trouve un algorithme classique, la théorie de la fonctionnelle de la densité, utilisée en physique et en chimie pour comprendre les structures des atomes et fournir un critère pour prédire la stabilité des structures. Les résultats de cette évaluation, les matériaux potentiels stables, sont réintroduits dans la base de données et notés, et ces informations sont utilisées dans l'algorithme d'apprentissage pour aider l'algorithme à apprendre ce qui constitue les matériaux potentiels. Quels sont les résultats ? Auparavant, nous avions connaissance d'environ 20 000 cristaux réputés stables. 48 000 de plus ont été déterminés par les approches informatiques antérieures. Avec l'apprentissage profond,

l'IA a produit 421 000 nouveaux matériaux potentiels qui sont des cristaux stables. Or 736 d'entre eux existaient déjà dans la littérature : ils ont été réalisés de manière indépendante par des équipes externes, et il y a parmi eux un nouveau matériau potentiellement supraconducteur.

La prévision météorologique

Les prévisions météorologiques sont très précises. La façon dont nous les faisons généralement de nos jours consiste à utiliser des quantités massives de superordinateurs pour calculer des équations très compliquées et obtenir des simulations météorologiques. Dans cet article, ils utilisent l'apprentissage profond pour créer ce qu'on appelle un « modèle de fondation », comme celui qui sous-tend les LLM. Ils ont pris des données provenant de GPS, de simulations temporelles, et ils ont pré-entraîné ces modèles. Ils ont automatiquement appris une représentation de la météo. Il est possible d'affiner ce modèle pour une tâche particulière, puis de l'utiliser pour des prévisions

à différentes résolutions. Ces résultats d'apprentissage profond sont extrêmement rapides : en moins d'une minute, ce modèle fait une prévision de la pollution atmosphérique mondiale à cinq jours et une prévision météorologique à haute résolution à 10 jours. Cela surpasse les outils de simulation en termes de vitesse et même de consommation d'énergie.

Le contrôle de la fusion nucléaire

Passons au contrôle des réactions nucléaires. Un réacteur à fusion de type Tokamak (ITER) est en train d'être construit dans le sud de la France. L'un des véritables défis de la fusion nucléaire réside dans le fait que son système magnétique à haute énergie doit contenir un plasma surchauffé. Or ce plasma est très dynamique et il est très difficile d'ajuster à la vitesse nécessaire le système magnétique et le système dans son ensemble pour le contenir. Ce qui est proposé, c'est un modèle d'IA reposant sur l'apprentissage profond par renforcement pour apprendre à effectuer ce contrôle. Ensuite, vous êtes en mesure d'effectuer le contrôle en boucle fermée de cette surface magnétique. Certains pensent que c'est l'un des rares moyens de maintenir une quantité suffisante de plasma à une vitesse suffisante pour obtenir une fusion continue. L'IA pourrait donc changer la donne pour la fusion nucléaire.

De l'expérimentation automatisée en boucle fermée à l'automatisation de la découverte

Jusqu'à présent, nous avons vu comment des chercheurs utilisent des IA pour améliorer la science. Et si un système d'IA pouvait réaliser une expérience de A à Z ? En l'occurrence, il s'agit d'un chimiste automatisé auquel on fournit une requête : « fais telle expérience de chimie ». Ce système planifiera automatiquement l'expérience. Il se connectera au matériel physique pour manipuler des pipettes. Il recherchera sur le web la documentation pour réaliser cette expérience. Il produira le code nécessaire à son exécution et disposera d'une boucle de rétroaction pour s'améliorer. Vous n'avez qu'à demander, et il génère automatiquement le code pour exécuter le protocole, vous indiquer les matériaux dont il a besoin, et faire ainsi toute l'expérience à votre place. Évidemment, c'est limité : cela fonctionne avec un appareil donné, et dans un environnement fermé particulier, mais c'est un résultat assez étonnant.

Enfin, on peut même envisager d'automatiser l'ensemble du « pipeline » de découverte scientifique. On prend un ensemble de données et on demande à une IA comment l'utilisation des terres urbaines a affecté l'invasion de différents types de plantes introduites en Catalogne. L'IA produit une procédure opérationnelle dont les étapes sont les suivantes : prendre cet ensemble de données, filtrer l'habitat avec le critère « urbain », effectuer une distribution binomiale, vérifier l'autocorrélation, effectuer un test statistique, puis produire une réponse finale. L'IA en conclut que l'utilisation des terres urbaines a diminué l'invasion par les plantes agroforestières, et propose un benchmark avec d'autres hypothèses. L'IA a donc réalisé une recherche scientifique de bout en bout. Mais, dans la science de tous les jours, l'intégration est plus modeste ; à l'université d'Amsterdam, par exemple, mon collègue Julian Evans, qui dirige le centre d'étude de la biodiversité, utilise l'IA pour en suivre l'évolution aux Pays-Bas.

La configuration internationale de la recherche en IA

Notre groupe de travail était dirigé par Anna Fabijanska et Andrea Emiliano Rizzoli. Il a recueilli, avec célérité, un nombre considérable d'informations auprès de trente-cinq experts, rédigé trois rapports d'atelier et effectué huit analyses documentaires systématiques sur une période d'environ trois ou quatre mois. Le rapport est donc très complet et la première chose qui en ressort est un « paysage ». L'IA est une technologie générique qui nécessite des infrastructures massives. C'est pourquoi l'industrie exerce un fort leadership dans la recherche : les articles les plus cités proviennent d'entreprises technologiques américaines, comme Google, Meta, Microsoft, etc., et ce sont les grandes entreprises, aux États-Unis et en Chine, qui dominent l'informatique disponible. Certes, il y a Leonardo, mais ce superordinateur est utilisé pour d'autres types d'informatique de haute performance (HPC), pas seulement pour de l'IA. Ce paysage est donc largement dominé par les États-Unis et la Chine. C'est aussi un paysage dont la régulation évolue rapidement et l'Europe, avec l'« EU AI Act », est un moteur de ces évolutions. Mais il faut tenir compte que l'évolution de la réglementation dans le monde entier rend la création de systèmes d'IA plus ou moins difficile et avec plus ou moins de propriétés. Voilà donc le paysage dans lequel nous nous trouvons : une technologie très importante, générique, dont les leaders s'imposent grâce à la puissance de calcul, dans un paysage réglementaire dynamique. Dans ce contexte, le rapport a mis en évidence l'utilisation croissante mais irrégulière de l'IA dans tous les domaines de la recherche.

Les risques globaux de l'IA pour la science

Nous manquons d'évaluations sur les effets globaux de l'IA et notamment d'études théoriques ou qualitatives de son impact sur le système scientifique dans son ensemble. Certes, nous disposons d'une vue d'ensemble sur la manière dont les gens utilisent l'IA, mais nous ne savons pas quel est son impact. Or l'utilisation de l'IA dans le domaine scientifique présente de nombreux défis et des risques globaux.

1. Le manque de transparence : Le premier défi est le manque de transparence. Des articles mentionnent l'utilisation de GPT-4 pour justifier leurs résultats. Or nous ne savons pas comment GPT-4 a été construit. Un rapport technique volumineux a été publié, mais nous ne savons pas exactement quelles sont les données qui ont été utilisées. Nous ne savons pas comment le modèle a été formé. Nous assistons à l'émergence de modèles de fondation en « *open source* », mais il y a vraiment un manque de transparence sur la façon dont ces modèles sont construits. La nature même des modèles de fondation obtenus par un apprentissage profond basé sur des transformeurs manque également de transparence. Vous ne savez pas quels paramètres provoquent quel résultat.

2. Les biais et données de mauvaise qualité : Nous constatons que les données récoltées à grande échelle déterminent les performances, de sorte que l'accès à un grand nombre de données est un avantage décisif, mais si vos données sont biaisées, votre modèle le sera aussi. Des données de mauvaise qualité produisent donc des résultats de mauvaise qualité. Cela peut extrapoler les biais existants qui sont déjà présents dans ces systèmes. Le World Wide Web, où beaucoup de ces données sont récoltées, est un produit de l'humanité, mais ce que l'humanité publie sur le World Wide Web n'est pas toujours ce qu'il y a de mieux. C'est un vrai problème, en particulier lorsque vous commencez à utiliser l'IA pour essayer de résoudre des problèmes qui dépassent les compétences de l'ingénierie physique, quand vous commencez à faire, par exemple, de la recherche à retombées potentiellement politiques ou de la recherche plus qualitative sur les personnes.

3. La désinformation : Dans l'édition scientifique, c'est déjà un vrai problème : la capacité de générer de faux articles, de faux résultats de recherche, et de surcharger le système d'évaluation par les pairs. Il s'agit simplement d'une surabondance d'informations et de la possibilité qu'elles passent à travers nos contrôles de publication de sorte que de mauvaises informations soient publiées.

4. Les inégalités dans la recherche : Les cartes de processeur graphique (à la base des réseaux neuronaux) coûtent cher. Même dans une institution comme l'université d'Amsterdam, dans un pays d'Europe du Nord plutôt prospère, où mes étudiants ont accès à des clusters, nous ne pouvons pas rivaliser avec les types de systèmes dont disposent d'autres chercheurs, de DeepMind par exemple.

La capacité à s'appuyer ou non sur les systèmes d'IA ne fait qu'exacerber les inégalités existantes au sein du système de recherche.

5. Le manque de connaissance et de lignes directrices : Un autre défi est celui de la formation. Comment les gens sauront-ils comment utiliser correctement ces systèmes? Quand et où faut-il introduire des lignes directrices? Et quand et comment utiliser ces systèmes de manière appropriée? Il n'existe pas beaucoup de lignes directrices exploitables et comme la technologie évolue très vite, il y a donc un manque de connaissances.

6. Les utilisations potentiellement dangereuses : J'ai mentionné la recherche d'antibiotiques dans l'espace chimique. Cela signifie que vous pouvez également rechercher des armes chimiques potentielles dans l'espace chimique. En réalité, vous pouvez rechercher dans l'espace chimique des protéines potentielles qui font ce que vous voulez ou ce que vous ne voulez pas. Le grand problème est que cette technologie est, dans un certain sens, facilement accessible et qu'elle le devient de plus en plus. Dès lors, comment se protéger contre un type d'utilisations potentiellement dangereuses? C'est l'un des aspects dont j'ai pris conscience en rédigeant ce rapport : l'abondance des utilisations potentiellement néfastes.

Les effets globaux de l'IA sur la société

Un grand nombre de ces défis ne sont pas nécessairement liés à la technologie en tant que telle. Mais ils ont à voir avec la manière dont nous utilisons cette technologie. L'impact sur les personnes, c'est là que nous avons le plus d'incertitude dans l'examen des données, mais nous pensons que les carrières et les emplois seront fortement impactés par l'IA. Les gens auront à l'avenir besoin de compétences numériques et de connaissances en matière d'IA pour poursuivre leur carrière. Mais des compétences différentes sont nécessaires pour ceux qui emploient l'IA et ceux qui développent l'IA, et probablement que vous n'avez pas besoin de savoir comment se forme un réseau neuronal, ni en quoi consiste un « transformeur », si vous êtes juste un utilisateur.

Mais nous devons autant développer notre compréhension de ce que les systèmes d'IA peuvent faire pour ces personnes profanes que pour les personnes désireuses de développer des systèmes d'IA à la pointe de la technologie. Nous avons constaté que les partenariats public-privé, en raison de la nature générique de l'IA, sont utiles, mais qu'il y a aussi des inquiétudes quant à l'appropriation des résultats de ces partenariats. Nous pensons enfin qu'il faut surveiller la capacité à automatiser complètement les systèmes d'IA ou l'ensemble du processus scientifique, et favoriser l'amélioration des capacités des personnes et des scientifiques.

Les quatre grandes recommandations :

1. La première est de **déployer des cadres qui évoluent avec l'environnement dynamique**, le rythme rapide de l'IA, et de renforcer l'utilisation de l'IA dans la recherche. Nous devrions donc disposer à la fois d'un cadre réglementaire, d'un cadre systémique et d'un cadre de financement pour y parvenir.

2. Nous pensons qu'il faut **développer des normes de qualité**. Nous avons donc besoin de normes sur ce que signifie construire des systèmes d'IA de haute qualité et aussi de normes sur la fourniture d'un accès aux systèmes d'IA de manière à ce que nous puissions appliquer le principe fondamental de la méthode scientifique, à savoir la reproductibilité.

3. Il est aussi nécessaire d'**investir dans la puissance de calcul**. Une partie de la proposition consistait à mettre en place un programme ou une infrastructure pour les systèmes d'IA (de type CERN) afin d'être à la pointe de l'infrastructure pour l'IA.

4. Enfin, l'UE entend **demeurer fidèle à ses valeurs** et ses principes fondamentaux, et il faut s'assurer que nous développons des environnements dans lesquels ces principes éthiques sont au cœur des systèmes d'IA produits en Europe.

Le rapport détaille plus précisément comment ces recommandations peuvent être mises en œuvre.

Conclusion : préserver l'attachement inconditionnel à la rigueur scientifique

Une chose difficile à faire dans ce rapport était d'isoler l'enjeu de l'IA pour la science, parce que l'IA est une technologie générique. Par conséquent, toutes les choses dont vous parlez à propos de n'importe quel système d'IA se retrouvent lorsque vous parlez de l'IA pour la science ou de l'IA pour n'importe quoi d'autre. L'IA, j'espère vous l'avoir montré, est déjà utilisée tout au long du cycle de la recherche, et pour automatiser les flux de travail et diffuser les résultats de la recherche. On peut alors estimer que l'IA n'est qu'un accélérateur : elle peut améliorer les choses plus rapidement, mais elle peut aussi accélérer la survenue de mauvaises choses. Les plus grands problèmes sont la qualité de l'information, la capacité à générer des déchets et à polluer la recherche, et le risque de « capture du code », c'est-à-dire le fait que, pour pouvoir faire de la science, il faille s'abonner à une IA ouverte ou une entreprise ou utiliser les GPU des grands fournisseurs. En tout cas, cela soulève l'enjeu de maintenir la rigueur scientifique. Quel que soit l'outil que nous utilisons, le besoin de transparence, de reproductibilité, de franchise dans nos communications, et la capacité d'examiner les choses et de dire « hé, c'était une erreur, faisons une autre expérience », bref, cet état d'esprit scientifique qui consiste à se corriger les uns les autres, nous devrions le remettre à l'honneur. À l'ère de l'IA pour la science, la rigueur devient encore plus cruciale, et comme ce rapport y insiste, en tant que chercheurs, en tant que scientifiques, et en tant qu'ingénieurs, c'est la chose que nous devrions tous valoriser, que nous devrions conserver, quels que soient les outils que nous utilisons.



Discussion

Qu'en est-il de la qualification ou de l'amélioration des compétences des scientifiques? Est-il facile ou difficile de maîtriser ces nouveaux outils et logiques d'IA?

Paul Groth : L'accessibilité de ces outils augmente considérablement. Et dans les centres de science des données, nous enseignons comment faire de l'apprentissage automatique à nos chercheurs en les formant à l'apprentissage profond, et comme il existe de nombreux outils accessibles, les systèmes qui sont construits autour de ces outils sont de plus en plus accessibles eux-aussi. Il y a cependant aussi beaucoup de questions sur la transparence ou la reproductibilité des résultats. C'est ce que nous avons souligné dans notre rapport, et il y a notamment un véritable enjeu pour les sciences d'avant-garde. Ainsi, le modèle de fondation météorologique que je vous ai montré nécessite des capacités de calculs qui ne sont disponibles que pour Microsoft, tandis que d'autres sont assez faciles à réaliser, comme le système d'automatisation en boucle fermée. Le coût énergétique qui en découle constitue aussi un problème important, en particulier pour l'UE.

J'ai deux questions sur ce qui est habituellement opposé à l'IA: les hallucinations et les exemples contradictoires. Et aussi, qu'en est-il de la compréhension très limitée du fonctionnement du LLM ou du réseau neuronal en général?

Paul Groth : Je pense donc que nous devrions séparer la question des hallucinations de celle de l'explicabilité. L'enjeu est l'introduction d'un transformeur ou d'un modèle d'apprentissage profond au sein du processus de recherche. Mais, par exemple, dans la science des matériaux, l'un des éléments clés est la procédure de vérification, opérant au cœur du processus, à savoir la théorie de la fonctionnelle de la densité. En cas d'erreur, vous pouvez la détecter ou voir ce qui ne va pas. Il en va de même pour le pliage des protéines ou pour les sciences de ce type, qui posent ce genre de problèmes. En revanche, il y a, selon moi, un risque plus important quand l'IA accomplit la recherche de bout en bout, quand on s'en remet au système pour obtenir des réponses. C'est là qu'on commence à voir plus d'hallucinations. Avec des modèles plus avancés, on constate que ce problème s'atténue avec le temps. En ce qui concerne l'explicabilité de ces modèles, je pense qu'il n'est pas possible d'expliquer très bien les très grands

modèles qui font vraiment avancer le domaine de l'IA pour la science et qui sont très difficiles. Toutefois, je m'en remets à la « quasi-justification »: je ne sais pas, alors même que je vous donne une justification en ce moment, quel est le réseau neuronal qui opère dans ma tête - je ne pense pas que l'on puisse vraiment démontrer mon cerveau et regarder les neurones qui fonctionnent - pourtant je vous donne un raisonnement, je vous donne une explication. Je pense que c'est vers cela que nous allons nous diriger: le système pourra, s'il vous donne un résultat, produire également une explication et les sources qu'il a utilisées. C'est une voie plus positive que d'essayer de faire les neurosciences d'un grand modèle de langage.

J'ai deux questions, qui sont d'ailleurs liées: avez-vous comparé l'Europe et la Chine au sujet de l'IA? Ensuite, pouvez-vous faire une prédiction solide de ce qui se passera dans cinq ans?

Paul Groth : La Chine est le deuxième pays derrière les États-Unis en termes d'investissement dans l'IA. Pour la compétitivité mondiale en matière d'IA, les États-Unis sont vraiment en tête, suivis de près par la Chine, et l'Europe est en baisse sur toute la ligne, bien que la France ait progressé dernièrement. Ensuite, je refuse de faire des prévisions à cinq ans, vue la vitesse insensée à laquelle les choses évoluent. C'est si difficile de suivre. Il se peut que nous franchissions un seuil. Il existe une hypothèse appelée « lois d'échelle » selon laquelle plus la capacité de calcul augmente, plus l'intelligence s'accroît. Jusqu'à présent, ces lois d'échelle se sont avérées exactes, mais elles pourraient atteindre une limite ou, au contraire, qu'elles continuent et que nous obtenions des systèmes intelligents très puissants. Ce que je peux dire à propos des cinq prochaines années, c'est que ce qui compte n'est pas vraiment de l'évolution de l'IA, en un certain sens. Les capacités de l'IA dont nous disposons actuellement sont déjà assez étonnantes. Il s'agit plutôt de savoir comment les intégrer dans nos processus. Si vous êtes une société d'ingénierie, un organisme scientifique ou une entreprise, il s'agit de savoir comment intégrer l'IA existante, car il faudra du temps pour que les systèmes sociaux l'utilisent correctement.

Lorsque vous cherchez un nouveau matériau ou une nouvelle molécule, vous pouvez vérifier après coup si cela fonctionne ou non. En revanche, lorsque vous gérez, par exemple, le plasma d'un réacteur à fusion, vous avez besoin d'une IA « de confiance ». Vous devez être sûr que les données que vous avez introduites et la manière dont elles sont traitées sont suffisamment bonnes pour qu'elles ne commettent pas soudainement une erreur le jour où vous les utiliserez. Comment y parvenir ?

Paul Groth : Dans tous ces cas, vous construisez des systèmes, et il est très rare qu'un de ces systèmes soit juste un modèle d'apprentissage automatique ou un LLM ou un modèle de fondation tout seul. Il s'insère toujours à l'intérieur d'un système pour lequel les gens écrivent des programmes qui vérifient des choses. On le constate même avec l'IA en ligne : si vous utilisez l'API de vision artificielle d'OpenAI et que vous essayez, par exemple, de lui poser une question sur le nombre de femmes ou d'hommes qui se trouvent sur une photo ou une vidéo, vous n'obtiendrez généralement pas de réponse. Ce n'est pas parce que l'IA ne le peut pas. C'est parce qu'il y a des contraintes intégrées dans ce système. Vous avez vu des exemples où ces contraintes ont mal tourné, mais elles sont construites pour servir de garde-fous. C'est ainsi qu'il faut voir les choses : comment mettre en place des garde-fous ? comme pour ce réacteur à fusion, qui dispose d'un système de contrôle en temps réel. Il y a donc un système qui vérifie si nous sommes dans les paramètres attendus. Il ne s'agit pas seulement de faire confiance.

Avez-vous une idée de la manière dont ce type d'outils va fondamentalement changer la façon dont nous conduisons la science ? L'ancienne méthode consiste à créer un modèle et à faire des expériences pour vérifier si mon modèle correspond à la réalité, si nous automatisons la découverte, ne risquons-nous pas de perdre le contrôle sur la construction de nouvelles connaissances ?

Paul Groth : Je suis partagé sur ce point. D'une part, la seule façon dont je sais que les choses vont changer, c'est qu'il s'agit d'un outil puissant que tout le monde utilisera parce qu'il accélère le travail. Par exemple, si je dois rédiger un résumé, il se peut que j'utilise cet outil. Si je suis un locuteur dont l'anglais n'est pas la langue maternelle, j'utiliserai cet outil pour m'aider à écrire un meilleur anglais. D'autre part, il y a une idée plus audacieuse, sur laquelle je ne veux pas spéculer, mais des gens y travaillent et elle évolue rapidement, c'est celle d'une automatisation complète de la science : un système d'IA qui émet automatiquement des hypothèses, les expérimente et obtient des résultats. Actuellement, en tant que chercheurs et scientifiques, nous attribuons nos ressources en fonction de nos priorités, c'est-à-dire selon nos hypothèses, parfois nous théorisons. Dans le nouveau système l'outil s'en chargera et nous ne ferons

qu'établir des priorités. Ce serait une révolution pour la science. Il y a beaucoup de paris faits sur cela. Mais je ne sais pas quand cela arrivera : dans quelques années ou dans 15 ans ? Je n'en sais rien. Je fais cours sur les bases de données. J'enseigne donc aux gens comment ils doivent programmer une base de données. Et l'une des choses que je leur fais faire - et mes étudiants me détestent tous pour cela - c'est écrire du code sur du papier avec un stylo. Ils me demandent pourquoi je leur fais faire cela. Ils peuvent obtenir un système qui fera toute l'autosuggestion pour eux, ils peuvent tout consulter en ligne, mais en connaissant ces aspects internes, ils savent comment utiliser ces outils. Cela les oblige en quelque sorte à penser. Cela doit faire partie de notre système éducatif de mon point de vue. D'après notre rapport et les options politiques proposées par ce groupe de travail, l'un des principaux aspects de la question est l'éducation et la formation, et l'accent est mis sur la question de savoir comment maintenir la rigueur, c'est-à-dire notre capacité à vérifier les choses, et comment maintenir le temps pour le faire.

Vous avez parlé du fait que vous pourriez utiliser ces systèmes pour vous aider à rédiger, par exemple, une proposition de projet. Mais que se passe-t-il si ce sont des machines qui l'évaluent ? En d'autres termes, si vous fermez complètement la boucle, est-ce que ce ne sera pas problématique pour innover en dehors de l'espace de paramètres dans lequel vous vous trouvez ?

Paul Groth : Il y a énormément de recherches en cours à ce sujet. Si vous consultez notre rapport, il y a une section intitulée « Considérez toutes les choses que vous pouvez faire avec l'IA pour la science » et ensuite une longue section sur « Quelles sont toutes les implications ? » L'une d'entre elles est exactement celle que vous venez de décrire : je me demande déjà si mon article a été évalué par une IA et si l'article que j'évalue a été écrit par une IA. Il y a donc beaucoup à dire sur la vérification de ce que font les gens et sur la possibilité d'une plus grande transparence sur ce qui se passe. Maintenant, vous posez aussi une question différente, à savoir si nous y perdrons notre créativité. En fait, je ne le pense pas. Je pense que cela donne plus de pouvoir aux gens, mais c'est comme être positif : mon avis n'a aucun fondement. Toutefois toutes ces préoccupations ont été soulevées dans le rapport.

Patrick Maestro : Qu'en est-il de la prévision de ce qu'il ne sera pas possible de faire ? Par exemple, vous mentionnez de nouveaux supraconducteurs, mais s'ils sont à 10° K, cela n'a pas d'intérêt. Pouvons-nous envisager de mettre toutes les données que nous connaissons dans une machine et que celle-ci puisse nous dire : « Il y a 1% de chances de trouver un supraconducteur à température ambiante » ?

Paul Groth : Je ne voudrais pas dire que je ne sais pas. Mais si vous regardez les exemples que nous avons traités, ils étaient dans des environnements difficiles mais bien définis. Comme vous l'avez dit, je pense qu'il faut chercher partout pour trouver la solution. Potentiellement, l'IA peut fournir ces conclusions automatiquement à partir de toute la littérature. Je pense qu'un aspect intéressant dont je n'ai pas parlé est celui des liens entre la capacité à faire ces déductions et des choses comme la simulation et la mécanique. Peut-être que cela nous aiderait à obtenir quelque chose comme : pouvez-vous exécuter la simulation pour moi automatiquement afin de vérifier la faisabilité de ce système?

Comment le groupe a-t-il abordé la question de la pollution des bases de données utilisées pour former les outils d'IA sur la base de l'IA générative ?

Paul Groth : Le rapport a abordé cette question sous l'angle de la pollution de l'information par les systèmes d'IA qui la consomment, mais aussi par les personnes qui la consomment. Vous pouvez vous former à partir d'informations produites par d'autres systèmes d'IA, mais il est vraiment essentiel de savoir d'où proviennent ses informations, d'où l'intérêt pour une technologie qui relève de mon domaine de recherche, à savoir la provenance des données : l'Initiative pour l'authenticité du contenu. Cette technologie utilise la cryptographie pour authentifier la source de toutes les informations, de sorte que l'on peut réellement vérifier d'où viennent ces informations. J'ai l'intuition que nous allons évoluer vers un environnement dans lequel vous aurez des systèmes qui vérifieront si vous avez ce type d'authenticité de contenu. Si ce n'est pas le cas, c'est que l'information provient d'un système d'IA. Un bon article a été publié récemment sur la disponibilité des données pour entraîner de grands modèles de langage. Ce que nous avons vu au cours de l'année et demie écoulée, c'est un resserrement rapide de la part des sites web qui rendent leurs informations disponibles pour l'entraînement.

[Michael Matlosz complète la réponse:] Cette question a été soulevée dès les années 1990 à propos du World Wide Web, qui a été développé comme un service gratuit permettant aux gens d'obtenir des informations. Très vite, on s'est inquiété du fait que les seules bonnes bases de données sur le World Wide Web seraient les services payants. Il se pourrait que nous soyons confrontés à la même situation.

Sam Altman a déclaré qu'il pourrait y avoir, dans les dix prochaines années, une entreprise d'un milliard de dollars créée par un seul homme qui maîtrise très bien toute cette IA et qui est capable de générer une entreprise très puissante à lui tout seul. Pensez-vous que cela puisse également s'appliquer aux équipes de recherche et qu'un petit groupe de recherche chinois, américain ou européen devienne extrêmement productif et atteigne subitement un facteur 10 dans tous les classements grâce à l'IA? Et devrions-nous essayer d'encourager certaines équipes européennes à le faire ?

Paul Groth : Je le crois. Je peux très bien imaginer des équipes très restreintes faisant des choses importantes. Parce qu'il y a de la concurrence, la disponibilité de l'IA devient de moins en moins chère chaque mois. Donc, nous pouvons nous y attendre. La question est de savoir si cela viendra d'en bas ou d'en haut. Est-ce que ce sera un étudiant en doctorat qui le fera? Ou s'agira-t-il de quelque chose que l'UE, à partir de son appareil de recherche, pourra encourager? Quel serait le modèle qui encouragerait le développement de ce type d'équipes? C'est aussi une question de domaines de recherche. C'est de la biologie, de la chimie, de la science des matériaux ou d'une partie de la physique, c'est-à-dire d'un domaine qui aura une idée claire du problème qu'elle veut résoudre et qui utilisera ces outils, qu'émergera ce genre de choses.

Vous avez parlé de la nécessité pour l'Union européenne d'investir des sommes considérables dans la capacité de calcul. Avez-vous des lignes directrices pour aider l'Union européenne à décider de la quantité de ressources à affecter à ce domaine ?

Paul Groth : Nous n'avons pas estimé le niveau réel d'investissement que cela représente. D'autres organismes ont donné des estimations approximatives. En tout état de cause, il s'agit d'une somme considérable. Si vous considérez l'un des modèles de fondation de la dernière génération, il s'agissait de centaines de millions d'euros ; actuellement, les experts estiment que les nouveaux modèles coûtent un milliard ; certains pensent que le prochain investissement se situera dans la fourchette des dizaines de milliards. Il s'agit de ressources qui, dans le système scientifique, ne sont mobilisables que par une très grande entreprise ou l'Union européenne, peut-être un État-nation.

L'Académie des technologies de France travaille sur un avis au sujet des investissements de haut niveau que la France doit engager au cours de la prochaine décennie. Votre recommandation concernant un investissement très important nous donne à réfléchir. Comment a-t-elle été reçue ?

Paul Groth : Un aspect positif de la France, c'est de prendre au sérieux cette question, et l'on en voit le résultat dans son classement (la France est passée du 10^e au 5^e rang des pays investisseurs), et que vous ayez un écosystème de startups émergents très brillants autour de l'IA. C'est peut-être la bonne façon de procéder : créer un écosystème qui demande essentiellement cet ensemble de ressources et obtient ces ressources. Le rapport ayant été examiné par de nombreuses personnes au niveau des institutions européennes, ce que nous constatons c'est que les responsables y voient une confirmation de leurs vues. Cela nous conforte dans l'opinion que cela correspond à la perspective européenne.

[Patrick Maestro apporte un complément de réponse :] Le rapport a été très bien accueilli. Le groupe de travail a d'abord discuté avec le GCSA, le groupe des conseillers scientifiques en chef, qui a formulé quelques remarques supplémentaires, puis le transfert à la Commission européenne s'est très bien passé puisque certaines recommandations sont déjà prises en compte dans le programme du Commissaire européen.

Vous avez une certaine expérience de l'industrie de l'édition scientifique, puisque vous êtes un ancien d'Elsevier, et que vous travaillez maintenant du côté universitaire. Assistons-nous à la fin de l'évaluation par les pairs telle que nous la connaissons ?

Paul Groth : Nous allons nous orienter vers une plus grande transparence dans les évaluations (par les relecteurs), ce qui est aligné avec l'agenda de la science ouverte de l'UE. On va donner la priorité à la capacité de reproduire automatiquement les choses, à la capacité de vérifier automatiquement les choses, et ensuite seulement, à la toute fin, à l'examen par les pairs. Nous ne pouvons plus gaspiller les ressources de l'évaluation par les pairs parce que nous n'avons plus assez de ces personnes. Donc nous aurons des systèmes automatisés qui pourront également prendre en compte des éléments tels que les archives, où l'on examine le statut des personnes ou l'intégration institutionnelle pour faciliter les choses. Cela pourrait exacerber les inégalités, mais ce qui intéresse les chercheurs, c'est de savoir s'il existe des outils d'IA permettant de déterminer quelles sont les publications scientifiques qui sont réellement intéressantes. Or, il en existe. Depuis des années, il y a un débat sur l'inflation du nombre de publications et de très nombreuses personnes, y compris les universitaires eux-mêmes, disent que nous devrions sans doute produire moins d'articles, de meilleure qualité, qui apportent vraiment une contribution. L'IA peut-elle contribuer à améliorer la qualité des publications ? Soyons optimistes : les systèmes d'IA sont très doués pour extraire les principaux résultats, résumer les analyses, donc ils promouvoir la qualité, simplement en mettant en évidence les bonnes productions.

Mots-clés : AlphaFold, apprentissage profond, données, intelligence artificielle (IA), LLM, modèle de fondation, réseaux de neurones, transformeur, transparence

Citation : Paul Groth, Patrick Maestro & Michael Matlosz. (2024). *Intelligence artificielle pour la recherche et la technologie : un regard européen*. Les soirées de l'Académie des technologies. @

Retrouvez les autres parutions de l'Académie des technologies sur notre site academie-technologies.fr

Académie des technologies. Le Ponant, 19 rue Leblanc, 75015 Paris. 01 53 85 44 44

Production du comité des travaux.

Directeur de la publication : Patrick Pélatà

Rédacteur en chef de la série : Hélène Louvel

Auteur : Vincent Bontems

n° ISSN : en attente