Policy Perspectives
Vol. 7/2, Mars 2019

Un hub apolitique de recherche en IA

Dans le discours politique international, l'intelligence artificielle (IA) est souvent réduite à une course technologique. Pourtant, la diplomatie scientifique peut promouvoir le potentiel de ces nouvelles technologies et réduire leurs risques à un stade précoce.

Par Sophie-Charlotte Fischer et Andreas Wenger

L'intelligence artificielle agit comme un moteur de changement global de l'économie, de la société et de l'État. Bien que la recherche en IA remonte aux années 1950, il a fallu attendre ces dernières années pour que le sujet se place au cœur de l'agenda politique international. Les causes de cette évolution sont essentiellement d'ordre technique et économique. Trois facteurs combinés ont permis une nouvelle vague de progrès au cours des 15 dernières années: l'amélioration rapide de la puissance de calcul des ordinateurs, l'augmentation des volumes de données et l'optimisation des approches et des algorithmes d'apprentissage automatique. Les technologies d'IA mises au point par les laboratoires de certaines entreprises et universités trouvent de plus en plus d'applications pratiques. Or, l'expérience montre que l'IA renferme d'immenses possibilités économiques et sociales, mais aussi de grands risques.

Les grandes entreprises technologiques mondiales, en particulier aux États-Unis et en Chine, possèdent aujourd'hui une longueur d'avance en matière de connaissances, de données et d'applications dans le domaine de l'IA. Comparativement, les acteurs sociaux et étatiques accusent un retard. Les entreprises technologiques ont acquis un rôle prépondérant grâce à leurs excellentes conditions pour le développement et la commercialisation de l'IA, notamment de grandes puissances de calcul, des jeux de données diversifiés et d'importants investissements financiers dans la recherche et le développement. En associant ces atouts techniques à des salaires attractifs, des conditions de travail souples et la liberté de chercher des solutions créatives à des problèmes intéressants, ces entreprises se donnent les

Messages clés

- Les acteurs étatiques considèrent de plus en plus l'IA comme une ressource stratégique. Ils exercent une influence accrue sur les processus d'innovation et la diffusion des technologies d'IA.
- Dans un contexte de plus en plus marqué par la rhétorique de la compétitivité, les risques associés à des technologies d'IA immatures augmentent.
- Il faut créer un hub apolitique de recherche en IA qui s'engage pour un développement et une utilisation responsables, intégratifs et pacifiques des nouvelles technologies.
- La Suisse possède un écosystème d'IA dynamique et est bien placée sur le plan politique pour accueillir des initiatives de gouvernance de l'IA.

moyens de recruter les meilleurs chercheurs en IA au niveau mondial. En même temps, la concentration de la recherche et du développement en IA aux mains d'une poignée de grandes organisations nuit de plus en plus à la compétitivité des instituts de recherche publics et des petites entreprises.

Les acteurs étatiques, pour leur part, anticipent les répercussions globales de l'IA (comme technologie «à double usage») sur les rapports de force économiques et militaires à l'échelle internationale. L'IA est de plus en plus considérée et traitée comme une ressource stratégique. Un nombre croissant de pays développent des stratégies nationales d'IA pour ne pas se trouver à la traîne sur cette question. Du fait du grand potentiel économique de l'IA et de l'importance qu'elle revêt pour la sécurité, les États cherchent de plus en plus à agir sur les processus d'innovation et la diffusion des nouvelles technologies. Ils soutiennent la formation de talents, la recherche et le développement dans le domaine de l'IA. Leur stratégie: établir des partenariats avec les grands groupes pour acquérir et utiliser des technologies d'IA à des fins publiques. Parallèlement, les gouvernements tentent de protéger leurs ressources nationales d'IA contre les influences internationales indésirables et d'empêcher le transfert des technologies d'IA à leurs concurrents – bien que l'effet de ces mesures soit controversé. Ainsi, ils déploient des efforts pour préserver leur base technologique et industrielle nationale des investissements étrangers directs et développent des considérations visant à restreindre les exportations de certaines technologies d'IA vers des pays concurrents.

La concentration des ressources d'IA aux mains d'entreprises technologiques privées et la concurrence politique que ce phénomène suscite entre les grandes puissances constituent une grande menace pour les sociétés et les régions moins à la pointe sur la question. Sur le plan économique, elles pourraient devenir de plus en plus vulnérables face aux oligopoles technologiques mondiaux; sur le plan politique, elles s'exposent à une dépendance accrue aux décisions des autres acteurs. Dans le même temps, les risques technologiques et sociaux associés au développement de l'IA menacent de passer au second plan. Dans un contexte de course technologique, l'impression qu'il existe un «avantage du précurseur» dans les domaines stratégiques peut créer une spirale descendante. Si l'avenir d'une technologie transformatrice est déterminé par un nombre d'acteurs de plus en plus restreint dans un temps de plus en plus court, la probabilité d'accidents dus à des technologies d'IA immatures augmente.

Ces dynamiques risquent de creuser les inégalités entre les sociétés et les régions. Or, le processus de développement de nouvelles technologies ne devrait pas être dominé par un petit groupe d'acteurs économiques et politiques. Dans un contexte mondialisé, il est indispensable d'impliquer les acteurs économiques, sociaux et politiques de toutes les régions, afin de permettre à tous d'accéder aux opportunités que renferme l'IA. Il faut créer des canaux de coopération internationale efficaces et prospectifs afin de garantir le caractère transparent, équitable et durable du développement et de l'utilisation de l'IA à moyen et long terme.

Diplomatie scientifique et coopération en matière d'IA

L'alliance de la science et de la diplomatie offre un instrument prometteur pour lutter à un stade précoce contre les risques techniques et sociaux liés à l'IA. La diplomatie scientifique jette des ponts entre les sociétés et permet d'élaborer des stratégies communes pour relever les défis internationaux. Fondés sur les valeurs et les normes méthodologiques de la science, ces échanges établissent un dialogue constructif entre les acteurs économiques, sociaux et étatiques. Les chercheurs se penchent sur des coopérations visant à développer de nouvelles connaissances. Les échanges scientifiques favorisent la compréhension et le tissage de liens personnels et institutionnels durables, par-delà les frontières politiques et idéologiques.

Par le passé, la diplomatie scientifique a déjà apporté une contribution décisive à la compréhension mutuelle dans les relations internationales. Des mégaprojets tels que la Station spatiale internationale ou le CERN (Organisation européenne pour la recherche nucléaire) se sont avérés des mécanismes efficaces pour agir contre le durcissement et la stagnation de la coopération politique internationale.

Le milieu de la recherche en IA est très international et a instauré une culture particulièrement ouverte. Dans ce contexte, la diplomatie scientifique pourrait tout à fait apporter des solutions aux défis mondiaux associés à l'intelligence artificielle. Beaucoup de chercheurs en IA publient leurs résultats sur la plateforme en ligne arXiv afin de les rendre accessibles aux autres utilisateurs intéressés et de permettre ainsi leur approfondissement. Les efforts déployés par les États pour restreindre la recherche transnationale et les échanges scientifiques dans le domaine de l'IA ont d'ailleurs soulevé de vives critiques de la part de la communauté scientifique.

Un hub neutre de recherche en IA

Dans le cadre d'une architecture mondiale de gouvernance de l'IA, nous proposons la création d'un hub apolitique, international et interdisciplinaire de recherche fondamentale en IA qui s'engagerait pour un développement et une utilisation responsables, intégratifs et pacifiques de ces nouvelles technologies. En conjuguant science et diplomatie, cette initiative donnerait naissance à un mécanisme durable de coopération internationale et rendrait le potentiel de l'IA accessible au plus grand nombre à travers le monde. Son lancement montrerait que le développement de l'IA ne suit pas un cours inéluctable, mais qu'il peut s'inscrire dans un cadre international. Une telle initiative renforcerait également la conviction que le développement et l'utilisation de l'IA ne sont pas un jeu à somme nulle. L'idée d'une plateforme mondiale de recherche en IA a déjà été abordée dans différents contextes, notamment à l'occasion

du sommet *AI for Good* ou lors de discussions dans le cadre de l'OCDE. Nous nous proposons ici de décrire concrètement une version possible de ce hub, afin de favoriser le débat public sur ses opportunités, ses défis et sa configuration éventuelle. Nous nous pencherons ensuite sur le rôle que pourrait jouer la Suisse dans la mise en œuvre d'un tel projet.

Pour faire face à certaines tensions qui émergent au niveau international, ce hub international de recherche fondamentale en IA aurait quatre fonctions centrales:

En premier lieu, il s'agirait de créer une plateforme intégrative et compétitive sur le plan mondial. Le hub ferait figure de laboratoire et attirerait les meilleurs talents du secteur, afin de jeter les bases d'applications transparentes de l'IA sur toute la planète et d'établir l'IA comme un bien public mondial. Le cadre de recherche de la plateforme couvrirait un large éventail d'approches méthodologiques. Même si les progrès actuels reposent en grande partie sur des algorithmes d'apprentissage profond (ou *Deep Learning*), il conviendrait d'intégrer d'autres paradigmes et des chercheurs spécialisés dans les disciplines connexes, telles que les neurosciences. Ceux-ci pourraient, en effet, donner des impulsions intéressantes à la recherche fondamentale en IA. L'initiative s'appuierait sur les normes scientifiques et éthiques les plus élevées et offrirait un environnement de travail attractif pour les meilleurs talents.

En deuxième lieu, le hub servirait de centre de recherche, de réflexion et de gestion des risques techniques et sociaux associés à l'IA (risques pour la sécurité, traçabilité des processus décisionnels algorithmiques, etc.). Des spécialistes des sciences humaines et sociales participeraient également à l'étude de ces risques.

En troisième lieu, le hub contribuerait à l'élaboration de normes et de bonnes pratiques concernant l'utilisation de l'IA. Cette communauté de recherche internationale et interdisciplinaire participerait aux futurs processus et forums de gouvernance de l'IA à titre consultatif. On éviterait ainsi qu'un petit groupe d'acteurs puissants détermine les normes en matière d'IA.

En quatrième lieu, l'initiative s'établirait comme un centre de formation et de perfectionnement pour les chercheurs en IA. Les étudiants en master et les doctorants suivant une partie de leur formation au sein du hub pourraient ensuite jouer un rôle de multiplicateurs en poursuivant leur carrière dans d'autres organisations. À travers des partenariats de recherche et de formation, le hub contribuerait à renforcer la communauté du savoir dans le domaine de l'IA. Au centre de l'initiative, des chercheurs de différentes disciplines et nationalités pourraient former des groupes de projets flexibles afin d'utiliser l'IA pour aider à résoudre des problèmes mondiaux, tels que ceux formulés dans les objectifs de développement durable des Nations Unies. Une approche décentralisée serait également envisageable, avec la création d'antennes dans d'autres régions du monde afin de développer de façon dynamique des applications d'IA conçues pour répondre à des problèmes précis en s'appuyant sur l'expertise locale et les groupes sociaux concernés.

Gouvernance du hub d'IA

Ce hub international d'IA doit être ouvert à tous les pays. Il semble toutefois pertinent que le grand projet scientifique soit amorcé par une «coalition de volontaires». En effet, il est important qu'un groupe suprarégional d'États partageant les mêmes vues s'engage dès le départ. L'objectif est d'éviter que l'initiative soit perçue comme émanant d'un groupe sélectif de pays désireux de promouvoir unilatéralement la compétitivité de leur région, ce qui découragerait les autres pays de s'associer au projet. S'il est accepté par la communauté internationale d'experts et s'établit comme un grand carrefour au sein d'un réseau mondial en faveur d'une IA sûre et responsable, on pourra alors considérer que le hub remplit sa mission.

Les résultats obtenus dans le cadre du hub de recherche seront mis à la disposition de ses États membres à titre de bien public. Comme le CERN, les groupes de recherche du hub d'IA doivent pouvoir coopérer avec des laboratoires partenaires dans le monde entier – à condition que les résultats de leurs travaux soient accessibles à l'ensemble des États membres. En outre, tous les États membres bénéficieront d'une infrastructure commune de haute qualité pour la recherche en IA, infrastructure qu'ils ne pourraient pas développer seuls. Les participations financières des membres

Lecture supplémentaire

AI Governance: A Research Agenda *Dafoe, A., Center for the Governance of Artificial Intelligence, Université d'Oxford, 2017.*
Ce rapport propose un nouveau programme de recherche intégral sur la gouvernance de l'IA.

Concrete Problems in AI Safety *Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané, 2016.*
Cet article explore différents problèmes de recherche liés au fonctionnement correct des systèmes de machine learning de pointe.

The CERN Community; A Mechanism for Effective Global Collaboration? *Mark Robinson, Global Policy, vol.10 (1), p. 41 – 51, 2018.*
Cet article examine à travers l'exemple du CERN la raison pour laquelle les mégaprojets scientifiques ont été particulièrement efficaces dans le développement de la collaboration internationale.

doivent être en adéquation avec leurs moyens économiques et peuvent prendre la forme de contributions en nature, ce qui facilitera l'adhésion des pays moins riches.

On peut imaginer différents modèles pour créer un lien entre le hub d'IA et l'ONU. Comme l'AIEA (Agence internationale de l'énergie atomique), le hub pourrait être une organisation technique et scientifique autonome liée à l'ONU par un accord de coopération et rendant régulièrement compte à l'Assemblée générale et au Conseil de sécurité.

Un rôle pour la Suisse

Différents paramètres font que la Suisse est particulièrement bien placée pour promouvoir activement l'idée d'un hub international de recherche en IA, voire pour accueillir cette structure. Siège européen de l'ONU, non-membre de l'UE et pays comptant parmi les plus mondialisés de la planète, la Suisse pourrait jouer un rôle de pont afin de garantir dès le départ l'orientation supranationale du hub. Sa neutralité politique, sa stabilité, son autonomie et son expérience des processus multipartites en font un hôte crédible pour une plateforme mondiale de recherche en IA. L'engagement historique de la Suisse pour que le CERN se consacre exclusivement à des recherches civiles, et non militaires, contribue également à cette perception.

En outre, la Suisse combine avantageusement les conditions économiques et scientifiques essentielles à la mise en œuvre d'un tel projet. Elle possède déjà un écosystème très dynamique dans le domaine de l'IA et dans les disciplines connexes, avec d'excellentes universités techniques telles que l'EPFZ et l'EPFL, un paysage de start-ups extrêmement vivant et des entreprises technologiques leaders sur la scène internationale. En intégrant la Confédération, le monde de la recherche et l'industrie, l'initiative faciliterait le regroupement des compétences nationales et, du fait de leur masse critique, leur interconnexion à l'échelle internationale. Avec sa qualité de vie et sa situation favorable au cœur de l'Europe, la Suisse pourrait constituer une alternative attrayante à la Silicon Valley ou à d'autres pôles prisés par les meilleurs chercheurs mondiaux en IA.

Au niveau international, la Suisse aurait ainsi l'occasion d'asseoir sa position de leader mondial de la recherche et de l'innovation, de moteur de la coopération internationale et de bâtisseur de ponts au carrefour de la paix et de la technologie. Sur le plan national, l'initiative offrirait à la Suisse, pays fortement interconnecté et dépendant du commerce extérieur, la possibilité de participer de façon crédible à l'élaboration de normes mondiales en matière d'IA. En accord avec ses valeurs et ses intérêts, la contribution de la Suisse conjuguerait les connaissances et les compétences d'une administration innovante, d'une industrie agile et d'une recherche de haut niveau. En alliant l'engagement historique de la Suisse pour la paix aux compétences technologiques d'un petit État hautement développé, l'idée d'un hub apolitique, international et interdisciplinaire de recherche en IA montre la voie.

Sophie-Charlotte Fischer est doctorante au Center for Security Studies (CSS) à l'ETH de Zurich.

Andreas Wenger est professeur de politique de sécurité internationale et suisse à l'ETH de Zurich et dirige le Center for Security Studies (CSS).

Policy Perspectives is edited by the Center for Security Studies (CSS) at ETH Zurich. The CSS is a center of competence for Swiss and international security policy. It offers security policy expertise in research, teaching, and consultancy. The CSS promotes understanding of security policy challenges as a contribution to a more peaceful world. Its work is independent, practice-relevant, and based on a sound academic footing.

Editor: John (Jack) Thompson

Comments welcome at PolicyPerspectives@sipo.gess.ethz.ch

Recent editions of **Policy Perspectives**:
- **Trump's Missile Defense: Challenges for Europe** (7/1) by Oliver Thränert
- **The Use of Smart Pressure to Resolve Civil Wars** (6/8) by Dr. Allard Duursma

For more editions, and for a free online subscription, visit www.css.ethz.ch/en/publications/css-policy-perspectives

© 2019 Center for Security Studies (CSS), ETH Zurich www.css.ethz.ch

ISSN: 2296-6471; DOI: 10.3929/ethz-b-000332541