

LES CRAINTES LIÉES À L'IA DANS LES RÉGIMES D'INTERDICTION DES ARMES CHIMIQUES ET BIOLOGIQUES

L'intelligence artificielle (IA) a le potentiel de révolutionner les activités scientifiques, économiques et sociales dans divers secteurs. Cependant, les répercussions négatives que les applications à double usage de l'IA et leurs résultats pourraient entraîner suscitent des inquiétudes croissantes. Ces « craintes liées à l'IA » sont à la fois réelles et imaginaires, enracinées dans ce qui est possible aujourd'hui et ce qui pourrait l'être à l'avenir.

Des inquiétudes concernant les potentielles applications à double usage de l'IA ont également été formulées dans le contexte des régimes d'interdiction des armes chimiques et biologiques, où la capacité de l'IA à permettre la conception, la mise au point, l'utilisation et la détection des armes chimiques et biologiques commence à être prise en compte.

L'idée que l'IA pourrait faciliter la mise au point d'agents supertoxiques ou permettre à des États et d'autres acteurs de mettre au point et d'utiliser des armes chimiques et biologiques nourrit l'imagination du grand public.

Dans cette note d'information, nous présentons quatre nouvelles craintes liées à l'IA concernant les armes chimiques et biologiques, en nous appuyant sur les principes éthiques existants pour une IA responsable afin de réfléchir à la manière dont nos normes et nos valeurs peuvent aider à relever ces défis.

En identifiant les caractéristiques de l'IA que ces principes cherchent à atténuer, nous pouvons mieux visualiser comment elles peuvent créer des défis particuliers dans le cadre des régimes d'interdiction des armes chimiques et biologiques. Il s'agit d'une approche préliminaire qui peut contribuer au travail de compréhension de la nature de ces nouveaux défis.

Points clés

- L'IA n'est pas une entité autonome, mais l'un des éléments d'un système plus vaste qui, associé à d'autres technologies, améliore le traitement des données et la prise de décision.
- Si l'IA peut créer de nouveaux défis et de nouvelles craintes, nombre d'entre eux ne sont pas fondamentalement différents de ceux associés à d'autres technologies ou pratiques. L'IA peut amplifier et modifier les défis existants et doit être considérée dans le contexte plus large des systèmes dans lesquels elle est utilisée.
- L'atténuation des conséquences potentiellement néfastes de l'IA ne nécessite pas de nouveaux modèles de gouvernance. En cernant les potentielles craintes liées à l'IA et en y apportant une réponse, elle peut en revanche contribuer à la modification et à l'élargissement des cadres existants.



IA + (TECHNOLOGIE)

L'intégration de l'IA à d'autres technologies pose des défis importants aux régimes d'interdiction des armes chimiques et biologiques. Les scénarios qui répondent à cette crainte tournent autour de la manière dont l'IA pourrait accroître ou amplifier les risques liés aux technologies à double usage existantes. Par exemple, l'utilisation de l'IA pour cibler des véhicules autonomes afin de libérer des agents antiémeutes pourrait renforcer l'ambiguïté quant à l'utilisation sûre et légale des agents antiémeutes dans le cadre de la convention sur les armes chimiques (CAC). Par ailleurs, l'association de l'IA à la robotique avancée pourrait permettre à des acteurs disposant de ressources humaines limitées de concevoir et de synthétiser des produits chimiques ou biologiques à distance. Dans ce cas, les principes éthiques pourraient se conformer davantage aux valeurs d'une IA « sûre et sécurisée » et « axée sur le facteur humain », l'infrastructure, la gouvernance et l'évaluation des intentions pouvant contribuer à clarifier les mesures de gouvernance potentielles dans le cadre de l'utilisation de l'IA.



POSSIBILITÉS DE CONTOURNEMENT

L'IA pourrait aider à surmonter les limites, les obstacles et les contrôles traditionnels qui entravent la mise au point, le stockage et l'utilisation des armes chimiques et biologiques. Par exemple, l'IA pourrait favoriser le contournement des contrôles à l'importation et à l'exportation par des acteurs malveillants en recensant ou en modélisant des précurseurs non répertoriés destinés à la synthèse chimique. En outre, l'utilisation de l'IA peut réduire le nombre de personnes nécessaires à la mise au point ou à l'utilisation de ces armes, ce qui permet de réduire les coûts et d'éviter d'éventuels « dilemmes moraux ». L'IA pourrait également être utilisée pour concevoir des agents plus difficiles à détecter et contre lesquels il pourrait être plus compliqué de se défendre. Dans de telles situations, nous pourrions nous attendre à ce que les valeurs d'une IA « contrôlable » et « responsable » renvoient à la nécessité de se pencher sur la gouvernance dans le développement de l'IA.



L'IA DEVRA



Explicable et responsable

« Capacité d'expliquer comment et pourquoi des résultats particuliers ont été obtenus. Les données d'entrée, la structure de conception et les systèmes d'exploitation, le fonctionnement global et les résultats doivent permettre de rendre des comptes. »



Contrôlable

« Capacité à contrôler les données d'entrée, le fonctionnement et les résultats, de manière à réorienter, modifier, neutraliser ou arrêter le fonctionnement. Les résultats doivent également pouvoir être contrôlés (par exemple, prévention de la diffusion automatique). »

AIT ÊTRE...



Axée sur le facteur humain

« Les humains sont en fin de compte responsables de l'ensemble du cycle de conception et d'exploitation de l'IA, ainsi que de ses résultats, et le fait de comprendre où et comment les humains sont impliqués peut contribuer à l'évaluation des intentions. »



Sûre et sécurisée

« L'IA nécessite une sécurité cybernétique et infrastructurelle pour la protéger des acteurs malveillants. En interne, elle nécessite des solutions de secours, la possibilité de l'arrêter ou de passer outre, ainsi que des mécanismes d'examen pour s'assurer que les résultats ne présentent pas de danger pour les humains. »

LÉGITIMATION PROGRESSIVE

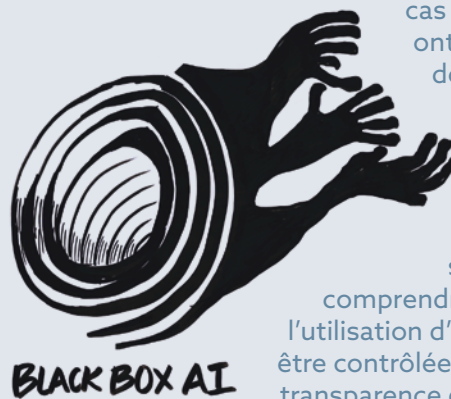
L'érosion insidieuse des interdictions légales et des tabous sociaux par le biais de certains axes de recherche et de développement n'est pas une préoccupation nouvelle. Toutefois, le potentiel de l'IA à faciliter, accélérer et créer de nouvelles voies de recherche qui dépassent légèrement çà et là les frontières de la légitimité peut offrir aux acteurs malveillants de nouvelles possibilités de mettre au point des agents et des technologies susceptibles de remettre en cause les traités sur les armes chimiques et biologiques. Par exemple, l'exacerbation des hypothèses existantes par l'IA peut produire une utilisation involontaire ou ambiguë des agents antiémeutes qui repousse les limites juridiques.



En outre, la recherche militaire sur les processus biologiques visant à obtenir des résultats de type « guerre sans mort » pourrait être ravivée par les possibilités offertes par les technologies de l'IA. Les principes d'une IA « responsable » et « axée sur le facteur humain » peuvent donner un aperçu de la manière dont les différents acteurs peuvent être gouvernés (ou peuvent gouverner) pour maintenir les limites juridiques et sociales à la lumière des nouvelles capacités.

LA BOÎTE NOIRE

La nature opaque des données d'entrée, des méthodes et des résultats de l'IA crée des contraintes de transparence supplémentaires pour les développeurs et les utilisateurs. L'intangibilité, l'opacité et le manque d'explicabilité des systèmes d'IA se répercuteront sur les défis à relever dans de nombreux domaines et les caractériseront. Par exemple, lors de la vérification d'allégations d'utilisation, il peut être difficile d'établir la responsabilité si un outil d'IA standard est utilisé sans que l'on connaisse exactement la provenance de l'outil, les données d'apprentissage et les méthodes de déduction. Par ailleurs, l'utilisation d'outils d'IA pour mettre au point et produire des armes chimiques et biologiques potentielles peut limiter l'attribution dans les cas où de multiples acteurs ont accès aux outils et aux données nécessaires pour ce faire. Dans ce cas, les principes d'une IA « explicable », « responsable » et « contrôlable » sont essentiels pour comprendre comment la création et l'utilisation d'algorithmes d'IA peuvent être contrôlées grâce à une plus grande transparence et à un meilleur partage des informations.



limiter l'attribution dans les cas où de multiples acteurs ont accès aux outils et aux données nécessaires pour ce faire. Dans ce cas, les principes d'une IA « explicable », « responsable » et « contrôlable » sont essentiels pour comprendre comment la création et l'utilisation d'algorithmes d'IA peuvent être contrôlées grâce à une plus grande transparence et à un meilleur partage des informations.

Conséquences

Les quatre craintes liées à l'IA montrent que l'intégration des technologies de l'IA ne produit pas nécessairement de nouvelles catégories de menaces ou de risques, mais qu'elles s'intègrent dans les catégories existantes. Il est important de noter que des scénarios particuliers s'inscrivent dans plusieurs catégories de craintes en fonction de la manière dont l'IA interagit avec les technologies et les modes de gouvernance existants. Ce constat exige que les craintes suscitées par l'IA soient ancrées dans les contextes et les défis que les acteurs concernés par les régimes d'interdiction des armes chimiques et biologiques connaissent déjà. Dans ces contextes, l'IA aura probablement pour effet :

- d'accélérer les processus de recherche et de développement ;
- d'ouvrir de nouvelles voies de recherche potentielles ;
- de réduire la transparence ;
- de compliquer la circulation, la pertinence et les conséquences de l'information.

L'étape suivante consiste alors à évaluer si les formes de gouvernance existantes sont en mesure, grâce aux capacités existantes ou à des modifications, d'atténuer l'impact de l'IA sur les défis existants et émergents.

Les quatre principes éthiques pour une IA responsable nous orientent vers le type de questions que nous devrions poser pour visualiser, caractériser et minimiser les conséquences négatives des technologies de l'IA.

Pour de plus amples informations :

Ces travaux de recherche ont été menés par le programme Harvard-Sussex du service de recherche en matière de politique scientifique de l'école de commerce de l'université de Sussex (University of Sussex Business School). Le projet de recherche a été gracieusement financé par le Counter Proliferation and Arms Control Centre du ministère britannique des Affaires étrangères et du Commonwealth. Nous remercions Shaunna McIvor et Boaz Chan qui ont apporté une aide précieuse au projet.

Chercheur principal : Joshua R. Moon
Adresse électronique : J.R.Moon@sussex.ac.uk

Quelle gouvernance pour l'IA dans le contexte des armes chimiques et biologiques ?

1. Mettre en contexte le développement et les applications de l'IA afin de mieux comprendre les défis en matière de politique.
2. Déterminer l'environnement social de l'IA en identifiant les acteurs impliqués dans le développement et l'objectif de l'utilisation.
3. Examiner les craintes liées à l'IA en tenant compte de la manière dont elles peuvent entraver ou affaiblir les initiatives actuelles pour interdire les armes chimiques et biologiques.
4. Évaluer les acteurs et les modèles de gouvernance existants pour repérer les possibilités et les lacunes afin de répondre à ces craintes liées à l'IA et, si possible, d'y remédier.
5. Se concentrer sur les outils et les mécanismes existants pour répondre aux craintes liées à l'IA et propres à chaque défi.

Prochaines étapes pour une meilleure politique en matière d'IA :

- L'élaboration de scénarios potentiels plus détaillés concernant l'impact de l'IA sur les interdictions relatives aux armes chimiques et biologiques, en fournissant notamment des précisions sur l'utilisation des technologies et les rôles des différents acteurs dans ces scénarios, permettra de mieux comprendre comment les risques peuvent être atténués.
- La compréhension de l'impact et des conséquences de l'IA en contexte et dans la pratique permettra d'approfondir les connaissances sur les modalités de gouvernance de l'IA en vue de préserver ses conséquences positives sur la société.
- La cartographie et l'évaluation de la gouvernance existante pour les régimes d'interdiction des armes chimiques et biologiques, en s'appuyant notamment sur des scénarios détaillés, permettront aux décideurs politiques de mieux comprendre l'interaction de l'IA avec les différents réseaux de prévention et la manière dont ces réseaux peuvent être adaptés pour prendre en compte ce nouveau domaine de la science et de la technologie.