Recommandations de bonnes pratiques suite à l'analyse des questions éthiques soulevées par l'utilisation de l'Intelligence Artificielle dans la recherche à l'Inserm.

Henri Atlan, Catherine Bourgain, Herve Chneiweiss, François Eisinger, Catherine Vidal, Frédérique Lesaulnier, Marion Abecassis, Philippe Ravaud, Ghislaine Filliatreau, Ioana Andreescu, et al.

▶ **To cite this version:**

Henri Atlan, Catherine Bourgain, Herve Chneiweiss, François Eisinger, Catherine Vidal, et al.. Recommandations de bonnes pratiques suite à l'analyse des questions éthiques soulevées par l'utilisation de l'Intelligence Artificielle dans la recherche à l'Inserm.. Guide de bonnes pratiques de l'Intelligence Artificielle à l'Inserm / Février 2025, 2025. inserm-04975393

HAL Id: inserm-04975393
https://inserm.hal.science/inserm-04975393v1

Submitted on 4 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

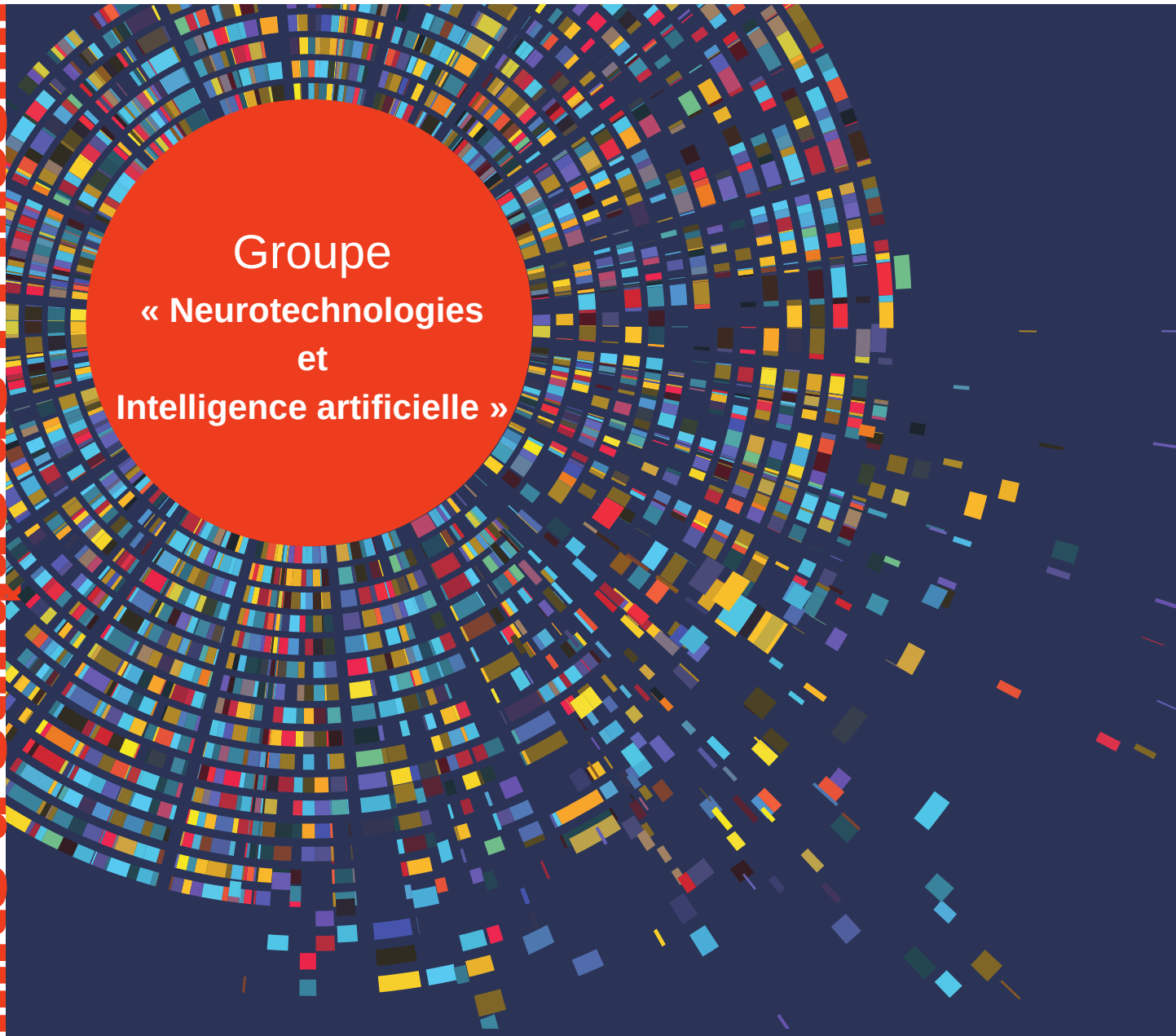
RÉPUBLIQUE FRANÇAISE

Liberté
Égalité
Fraternité

Inserm

La science pour la santé
From science to health

Comité d'éthique de l'Inserm



Groupe
« Neurotechnologies et Intelligence artificielle »

Guide de bonnes pratiques de l'Intelligence Artificielle à l'Inserm

Février 2025

Note conjointe du Comité d'éthique de l'Inserm, du programme LORIER et du Conseil scientifique de l'Inserm

Membres du Groupe de Travail (Juillet 2024 - Novembre 2024)

Pour le Comité d'éthique de l'Inserm : Henri Atlan, Catherine Bourgain, Hervé Chneiweiss, François Eisinger, Catherine Vidal, Frédérique Lesaulnier et avec la contribution juridique de Marion Abecassis.

Pour le programme LORIER : Philippe Ravaud; Ghislaine Filliatreau; Ioana Andreescu

Pour le Conseil Scientifique de l'Inserm : Rodolphe Thiebaut, Isabelle Remy-Jouet, Lotfi Senhadji, Valérie Urback, Murielle Gaudry

Experts extérieurs :

- Anita Burgun, PUPH informatique médicale Institut Imagine, Université Paris Cité, AP-HP & Inserm, Chaire PRAIRIE.
- Jean Charlet, Directeur de recherche AP-HP, LIMICS Sorbonne Université & Inserm U1142.
- Pierre-Antoine Gourraud, PUPH Biologie Cellulaire Nantes Université, Inserm UMR 1064 & Responsable entrepôt de données de santé du CHU de Nantes « Clinique des Données ». Membre de la direction scientifique du Ouest Data Hub.

Recommandations de bonnes pratiques suite à l'analyse des questions éthiques soulevées par l'utilisation de l'Intelligence Artificielle dans la recherche à l'Inserm.

« ... si nous n'éprouvons pas de l'inquiétude devant une technique, c'est qu'elle n'est pas assez révolutionnaire »

« l'ingénieur de demain ne sera pas apprenti-sorcier par négligence ou incompétence mais par dessein »
Jean-Pierre Dupuy -La marque du sacré

« il se pourrait que nous ne soyons plus jamais capables de comprendre les choses que nous sommes cependant capables de faire »
Hannah Arendt citée par Jean-Pierre Dupuy dans La marque du sacré

Résumé des principales recommandations

1- **Divulgation et Transparence** :

1.1 Divulgation naïve : les scientifiques doivent prendre garde à la diffusion non maitrisée de données (perte de la confidentialité des données) comme de la diffusion de données non maitrisées (hallucinations, données non sourcées) quel que soit l'outil numérique utilisé.

1.2 Publication scientifique : les chercheurs doivent indiquer explicitement l'utilisation de systèmes d'intelligence artificielle (SIA) dans leurs travaux (outils, algorithmes, paramètres) et distinguer les contributions obtenues via les SIA de celles qui sont le fruit de leur activité créative. Dans les publications, une section "Utilisation de SIA" doit détailler le rôle des SIA, à l'instar des logiciels utilisés pour les analyses statistiques. Un guide pratique et des références seront élaborés et disponibles sur le portail LORIER.

1.3 Administration de la recherche et fonctions supports : la transparence doit aussi concerner toutes les autres utilisations de SIA à l'Inserm, en particulier dans le secteur des ressources humaines (recrutements, traitement des carrières...).

2- **Attribution et modèles transparents** : les scientifiques doivent publier les détails des modèles de SIA (données d'entraînement, versions) qu'ils créent ou utilisent, et assurer leur archivage à long terme pour les études de réplication. Ils doivent aussi favoriser les modèles permettant de travailler dans le respect des bonnes pratiques éthiques, en particulier la citation des sources de données, et encourager des usages académiques de la science ouverte.

3- **Utilisation de données synthétiques** : les scientifiques de l'Inserm sont encouragés à développer l'utilisation de données synthétiques qui permettraient de limiter les risques concernant la ré-identification des participants à une recherche (anonymat au sens CNIL/G29[1]/EDPB[2]), tout en assurant la vérisimilitude, leur adéquation à un ensemble d'utilisation secondaires et la diversité des données, et en évitant les biais qui pourraient compromettre la recherche. Ce recours aux données synthétiques anonymes mises en qualité

[1] G29 fait référence à l'article 29 de la directive européenne du 24 octobre 1995 sur la protection des données et la libre circulation de celles-ci qui a institué un groupe de travail rassemblant les représentants de chaque autorité indépendante de protection des données nationale.

[2] Le Comité européen de la protection des données (EDPB) est un organisme européen indépendant qui a succédé au G29. Cette organisation réunit les autorités nationales de protection des données des pays de l'Espace économique européen, ainsi que le Contrôleur européen de la protection des données (CEPD). https://www.edpb.europa.eu/edpb_fr.

permet de plus dans certains cas d'enrichir les jeux de données, en particulier dans les domaines où les données réelles sont rares ou sensibles.

4- **Vérification des résultats** des SIA : la responsabilité de l'exactitude des analyses générées à l'aide de SIA incombe à ses utilisateurs et en particulier aux chercheurs, qui doivent en valider la fiabilité et identifier les biais potentiels. Les personnels sont invités à bien tester la reproductibilité et la fiabilité des modèles d'IA en 1) comparant les résultats obtenus avec différents jeux de données et 2) testant les résultats obtenus avec différents algorithmes d'IA. La création d'une cellule nationale et transversale sur les usages du numérique pour les sciences pour la santé à l'Inserm[3] pourrait centraliser les informations de suivi (collecte de détection des biais et solutions trouvées) et aider à la formulation de l'offre de formation pour appuyer cette démarche. Compte-tenu de l'évolution rapide des SIA et des pratiques, une réflexion continue est nécessaire impliquant le Comité d'éthique, le Conseil scientifique et le programme LORIER.

5- **Documentation des données des SIA** : les données générées à l'aide de SIA doivent être clairement identifiées pour éviter toute confusion avec des observations réelles. Les chercheurs doivent garantir la traçabilité des données des SIA utilisées dans les études.

6- **Intégrité et équité** : les chercheurs doivent chercher à anticiper les impacts sociaux des SIA. Ils doivent se former aux règles légales, notamment en matière de protection des données et de respect des droits de propriété. Une vigilance particulière est requise pour les groupes sous-représentés ou historiquement discriminés.

7- **Contrôle, solutions alternatives et engagement public** : une surveillance continue de l'impact sociétal de l'usage des SIA est essentielle. L'emploi des SIA en recherche a un impact environnemental. Ils doivent donc être utilisés avec parcimonie et dans des contextes d'intérêt certain. Les SIA utilisant un jeu de données réduit et moins de paramètres, par exemple les Small Language Models (SLMs), représentent une IA frugale et sobre. Ce développement de SIA responsables peut également contribuer à améliorer la souveraineté numérique. La cellule nationale et transversale sur les usages du numérique pour les sciences pour la santé à l'Inserm pourrait recueillir les retours d'expérience et développer une culture de la transparence sur les usages des SIA.

[3] Lors des vœux 2025 du Pdg au personnel du siège, Didier Samuel a évoqué la création d'une cellule Inserm dédiée aux SIA.

8- **Formation continue** : former tous les personnels, scientifiques et administratifs, au numérique et à l'usage des SIA, ses principes et limites, est crucial pour une utilisation éclairée et responsable.

9- **Passeport des données de santé** : un "passeport de données" documentant l'origine, la qualité et les biais des données permettrait leur réutilisation responsable.

10- **Portail de transparence** : un portail donnerait aux participants aux recherches menées à l'Inserm une visibilité sur l'utilisation de leurs données de santé, assurant transparence, conformité réglementaire et flexibilité pour la recherche.

Pourquoi une telle Note ?

Les systèmes d'intelligence artificielle (SIA) sont devenus d'usage courant. Ils sont un champ de recherches et permettent le développement de la recherche. Un exemple ? Des développeurs de SIA ont reçu en 2024 deux prix Nobel ! Celui de Physique est allé à John J. Hopfield (Princeton University) et Geoffrey Hinton (University of Toronto) *" pour les découvertes fondamentales et les inventions qui permettent l'apprentissage automatique à l'aide de réseaux neuronaux artificiels"*[4]. Celui de Chimie a récompensé Demis Hassabis et John Jumper de l'équipe Google DeepMind qui a conçu le logiciel AlphaFold de prédiction de la structure des protéines[5].

L'utilisation croissante de SIA dans la recherche médicale et en santé, notamment au sein d'institutions telles que l'Inserm, soulève des enjeux majeurs à la fois prometteurs et posant questions. Les SIA servent à l'Inserm évidemment à soutenir la démarche scientifique mais leur usage se développe également au sein de l'administration pouvant à terme impliquer l'aide à la décision en matière de projets, de gestion ou de ressources humaines. Les SIA offrent des avantages indéniables : ils permettent de traiter des quantités massives de données, d'accélérer la découverte de nouveaux mécanismes à l'origine de pathologies, ils peuvent aider à révéler de nouveaux biomarqueurs, aider à prédire la survenue de pathologies et permettre d'envisager une médecine de plus en plus personnalisable. Au quotidien, les SIA deviennent de remarquables outils d'aide à la rédaction de texte en produisant des résumés, le plan d'un document, en proposant des améliorations de style, corrigeant l'orthographe, la grammaire, aidant à la génération de codes informatiques et à leur vérification ou encore comme outil de traduction, tous usages qui permettent de gagner du temps sur des tâches à faible valeur ajoutée. Cependant, cette révolution technologique s'accompagne également d'impacts délétères, par exemple leur coût énergétique et son retentissement écologique, et de nombreuses questions sur les effets de plus ou moins long terme de ces technologies, les incertitudes et risques associés qui concernent notamment la qualité et la confidentialité des données, les biais algorithmiques, l'interprétabilité des résultats, la dépendance aux

[4] "This year's two Nobel Laureates in Physics have used tools from physics to develop methods that are the foundation of today's powerful machine learning. John Hopfield created an associative memory that can store and reconstruct images and other types of patterns in data. Geoffrey Hinton invented a method that can autonomously find properties in data, and so perform tasks such as identifying specific elements in pictures." https://www.nobelprize.org/prizes/physics/2024/press-release/.

[5] https://www.nobelprize.org/prizes/chemistry/.

technologies propriétaires, la fragilisation de pratiques et des infrastructures en cas d'attaques malveillantes, ou encore la perte de souveraineté.

Les personnels de l'Inserm et sa gouvernance doivent ainsi s'organiser pour répondre aux défis éthiques, scientifiques, juridiques et sociaux que posent les SIA : comment permettre la transparence et l'absence de discrimination dans des algorithmes souvent opaques ? Quelle place faut-il (ou non) préserver pour l'expertise humaine face à des systèmes capables de « prédire » puis d'affirmer des diagnostics ou de prescrire des traitements ? Sur le plan de la santé publique, comment contrôler les risques d'accentuation de l'hétérogénéité des prises en charge et le creusement des inégalités d'accès aux soins, possiblement associés à des SIA, en particulier pour les populations sous-représentées dans les bases de données ? Comment utiliser les SIA de façon raisonnée afin de limiter l'impact écologique au juste indispensable ?

Notre capacité à apprendre à utiliser, contrôler et développer les pratiques les plus rigoureuses et intègres de ces technologies sera un facteur crucial pour en minimiser les effets négatifs anticipables et maximiser leurs impacts sur l'innovation biomédicale et l'amélioration de la prise en charge de tous les patients. Ce travail d'arbitrage doit faire l'objet de débats et pouvoir être mené de façon collective et transparente.

À l'Inserm, l'intégration des SIA représente à la fois une opportunité pour faire progresser les connaissances, mais aussi une responsabilité collective, afin de garantir que ces technologies seront utilisées de manière équitable, éthique, efficace, transparente. Une distinction devra en outre être effectuée selon que l'Inserm est utilisateur des systèmes d'IA ou concepteur de ces systèmes.

Comme l'intelligence humaine qui s'exprime sous de multiples formes, et en particulier par la créativité des chercheurs scientifiques, l'intelligence artificielle revêt de nombreux aspects. Si le terme aujourd'hui apparaît partout, son origine est déjà ancienne, puisque c'est en 1956 lors de la conférence de Dartmouth que le terme « intelligence artificielle » est officiellement utilisé pour la première fois par John McCarthy, Marvin Minsky, Nathaniel Rochester et Claude Shannon[6]. La recherche publique est un acteur central des développements des SIA. Aujourd'hui de nombreux outils algorithmiques sont utilisés dans le quotidien très divers de la

[6] https://co-ia.fr/chronologie-histoire/.

plupart des disciplines scientifiques, dont ils ont profondément transformé les pratiques et les perceptions. La montée en puissance récente de SIA dite IA générative et la mise à disposition, sans coût directement apparents pour les utilisateurs (inconscients de l'aspiration de leurs données par ces systèmes et des impacts environnementaux associés), de grands systèmes *Large Language Models* (LLM) comme ChatGPT de la société Open IA, ou DeepSeek-R1 de la société DeepSeek, constituent une source nouvelle de bouleversements qui concernent des dimensions encore plus nombreuses du travail scientifique. L'adoption rapide de SIA générative rend nécessaire une réflexion sur les conditions et les effets sur le travail scientifique, de ces nouvelles technologies, dont les développements vont très probablement, se poursuivre activement dans les années à venir. Des questions relatives à l'intégrité et à l'éthique de la recherche se posent notamment.

Avec le Règlement (UE) 2024/1689 établissant des règles harmonisées concernant l'intelligence artificielle (AI Act), l'Union européenne s'est positionnée en cheffe de file de la régulation des outils d'IA.

L'AI Act est entré en vigueur le 2 août 2024, et la mise en application de ses dispositions est échelonnée dans le temps, en fonction du niveau de risque associé aux SIA concernés. L'AI Act définit un "*système d'IA*" comme "*le processus consistant à générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions, qui peuvent influencer l'environnement physique ou virtuel, et la capacité des systèmes d'IA à inférer des modèles ou des algorithmes, ou les deux, à partir d'entrées ou de données*"[7]. La Commission européenne a publié des lignes directrices sur la définition des Systèmes d'IA le 6 février 2025[8].

Il convient de noter que l'AI Act exclut de son champ d'application les SIA ou les modèles d'IA spécifiquement développés et mis en service uniquement à des fins de recherche et développement scientifiques, ainsi que leurs sorties[9]. En effet, le législateur européen estime que la réglementation « *devrait soutenir l'innovation et respecter la liberté scientifique et ne devrait pas compromettre les activités de recherche et de développement* »[10].

[7] https://www.vie-publique.fr/questions-reponses/292157-intelligence-artificielle-le-cadre-juridique-europeen-en-6-questions.
[8] https://digital-strategy.ec.europa.eu/fr/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application.
[9] Article 2.6 de l'AI Act.
[10] Récital 25 de l'AI Act.

Toutefois, les contours de cette importante exemption de recherche n'ont pas encore été définis avec précision. Certains acteurs, comme la Fédération européenne des associations et industries pharmaceutiques (**EFPIA**), ont d'ores et déjà alerté les autorités européennes sur la nécessité d'apporter des clarifications à cet égard.

Il faut retenir que l'AI Act est conçu comme une réglementation applicable à un « produit », très inspirée de la réglementation européenne relative aux dispositifs médicaux (DM). Cela signifie qu'il y a des balises réglementaires à mettre en place tout au long du cycle de vie du produit et que si le résultat présente un risque, on recherchera la responsabilité des acteurs impliqués dans l'émergence de ce risque tout au long de la chaîne de valeur. Compte tenu de ses activités de recherche, l'INSERM devra donc suivre de près ces précisions à venir, ainsi que les modalités de gouvernance qui seront mises en place pour encadrer ses projets. En tout état de cause, si l'AI Act ne s'applique pas à certains SIA développés et utilisés par l'INSERM exclusivement à des fins de recherche, d'autres cadres réglementaires demeurent pertinents, notamment le RGPD en cas de traitement de données à caractère personnel.

Les techniques qui permettent l'inférence lors de la construction d'un système d'IA comprennent :

- des approches d'apprentissage automatique, qui apprennent à partir des données la manière d'atteindre certains objectifs ;
- des approches fondées sur la logique et les connaissances qui font des inférences à partir des connaissances encodées ou de la représentation symbolique de la tâche à résoudre.

Ces techniques vont au-delà du traitement de données de base en ce que la capacité d'inférence d'un système d'IA lui permet l'apprentissage, le raisonnement[11] ou la modélisation. En effet, Les systèmes d'IA étant conçus pour fonctionner à différents niveaux d'autonomie, ils bénéficient d'un degré variable d'indépendance dans leur action par rapport à une intervention humaine et de capacités à fonctionner sans intervention humaine.

Le règlement européen définit aussi :

- un système d'identification biométrique à distance comme "*un système d'IA destiné à identifier des personnes physiques sans leur participation active, généralement à*

[11] Le raisonnement versus l'application de règles logiques. On utilise toujours le terme « raisonnement » en IA mais il s'agit bien de l'application par des algorithmes de règles logiques.

distance, en comparant les données biométriques d'une personne avec celles qui figurent dans une base de données" ;

- un système d'identification biométrique à distance en temps réel comme "*un système d'identification biométrique à distance dans lequel l'acquisition des données biométriques, la comparaison et l'identification se déroulent sans décalage temporel important et qui comprend non seulement l'identification instantanée, mais aussi avec un léger décalage afin d'éviter tout contournement des règles*".

La recherche étant un des secteurs que l'IA générative peut transformer très vite, la Commission européenne a aussi publié des lignes directrices pour un usage responsable en science, « qui visent à préserver l'intégrité de la recherche ». Destinées aux chercheurs, aux organismes de recherche et aux agences de financement, ces recommandations s'appuient sur les chartes et guides déjà élaborés (COPE, ALLEA , etc). Pour l'écriture des articles, nous pouvons citer les reporting guidelines comme TRIPOD-AI qui sont requis par de nombreux journaux[12]. Tout en encourageant les scientifiques à tirer le meilleur de ces nouveaux outils, il leur est notamment demandé d'éviter de les utiliser dans certaines pratiques comme l'examen d'articles par les pairs et l'évaluation des projets de recherche. Les établissements de recherche se doivent, eux, de fournir ou faciliter la formation à ces outils et de contrôler la manière dont ils sont développés et utilisés en leur sein.

Une institution de recherche comme l'Inserm doit aborder cette question afin de déterminer une politique claire quant à l'usage de SIA tant dans la recherche que l'administration, et permettre à ses personnels de connaître ses préconisations et favoriser les débats. Un tel guide de bonnes pratiques est d'autant plus important que la qualité de la production scientifique utilisant les SIA ira de pair avec la responsabilité des scientifiques qui utiliseront ces systèmes. De fait ici l'éthique rejoint déjà la pratique du droit. Comme l'indique la juriste Juliette Sénéchal dans une analyse récente[13] : *« La rupture technologique découlant de la capacité d' «usage général » de ces nouveaux systèmes d'IA et de l'engouement des utilisateurs qui en a immédiatement découlé, a, en premier lieu, éloigné ces « systèmes » de la notion classique de produit, envisagé comme un simple « outil» technique doté d'un certain nombre de fonctions déterminées, pour se rapprocher de notions plus souples et polyvalentes,*

[12] https://www.tripod-statement.org/wp-content/uploads/2019/12/TRIPOD-Abstract_AI_Paper.pdf.
[13] L'AI Act dans sa version finale – provisoire -, une hydre à trois têtes Juliette Sénéchal, Dalloz Actualité, 11 mars 2023 https://www.dalloz-actualite.fr/flash/l-ai-act-dans-sa-version-finale-provisoire-une-hydre-troistetes.

telles que celles de services numériques et de contenus illicites au sens du Digital Services Act (DSA)[14], et ce pour une double raison. D'un premier côté, ces systèmes d'IA à usage général ont immédiatement été proposés comme des services complémentaires par des fournisseurs de services numériques de moteur de recherche, de réseaux sociaux, ou de places de marché au sens du DSA. D'un second côté, ces systèmes d'IA à usage général, en tant qu'IA génératives, peuvent produire, au titre de leurs résultats en sortie, des deep fakes[15] ou « hypertrucages » qui pourraient, pour certains, être qualifiés de contenus illicites au sens du DSA et qui auront, en outre, potentiellement vocation à circuler massivement sur les services de plateformes en ligne. L'on pense par exemple aux deep fakes pornographiques de la chanteuse américaine Taylor Swift qui ont récemment fait l'objet de 47 millions de vues en quelques heures sur un service numérique de réseau social[16]. » Nous devrons donc être particulièrement attentifs à chercher à prévenir (détecter, signaler) le recours à des pratiques de trucage, plagiat, usurpation… qui conduisent à enfreindre les règles de l'intégrité scientifique. Par ailleurs, l'augmentation du recours à ce type d'outil peut induire une dépendance accrue à des systèmes propriétaires, soumis à des logiques (notamment financières, aspirations de données…) qui peuvent ne pas être compatibles avec le respect des bonnes pratiques scientifiques, comme notamment le consentement, le respect de la propriété intellectuelle, les citations ad hoc, la vérification des résultats.

Dans la présente Note, nous donnons en Annexe 2 un rappel des principes de fonctionnement des grands systèmes d'IA, la Note elle-même soulignant les limites et contraintes avant de proposer un certain nombre de recommandations pour la gouvernance de l'Inserm et ses personnels.

Pour développer des recommandations destinées aux personnels des unités Inserm et au-delà pour la recherche en Santé et en sciences de la vie, le Comité d'éthique a mobilisé un groupe de travail en partenariat avec des animateurs du programme LORIER et du Conseil

[14] Règlement (UE) 2022/2065 du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE, JO L 277 du 27.10.2022, p. 1–102.
[15] AI Act: "deep fake means AI generated or manipulated image, audio or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful".
[16] Le Monde, 26 janvier 2024, Taylor Swift victime de fausses images pornographiques, l'intelligence artificielle en accusation. https://www.lemonde.fr/pixels/article/2024/01/26/taylor-swift-victime-de-fausses-images-pornographiques-lacommunaute-des-swifties-a-la-rescousse_6213215_4408996.html.

Scientifique auquel se sont associés des experts des SIA (liste complète des membres du GT en Annexe 1).

L'utilisation de SIA devra répondre des quatre grands principes de l'intégrité scientifique tel que définis par le Code de Conduite Européen[17]:

- *Fiabilité* en s'assurant de la qualité de la recherche au niveau de sa conception, des méthodes utilisées, de l'analyse des résultats obtenus, et de l'usage des ressources mobilisées pour la mener.
- *Honnêteté* par une recherche transparente, honnête et non-biaisée depuis son commencement, son analyse par les pairs et sa communication.
- *Respect* des collègues scientifiques, des participants de la recherche, de la société comme un tout y compris son héritage culturel et des écosystèmes.
- *Capacité à rendre compte* des différents aspects de la recherche, depuis ses idées princeps jusqu'à la publication, de son organisation, de sa gestion, de la formation des participants, de la supervision et du mentoring, et des impacts de cette recherche.

Notons enfin que plusieurs instituts de recherche comparables à l'Inserm se sont dotés de stratégies d'accompagnement de leurs personnels à l'utilisation éthique et intègre de SIA[18].

Les SIA ne se limitent pas aux LLM mais l'architecture même des LLMs soulève des questions spécifiques

Le développement de l'IA générative grâce aux Large Language Models (LLMs) qui sont développés à partir de l'encodage vectoriel de grandes quantités de textes et les avatars qui en découlent, les agents conversationnels comme ChatGPT, ont mis en lumière une IA facilement accessible, en apparence peu coûteuse, pleine de promesses (et de fantasmes)[19] , ce qui nous conduit à mettre en avant immédiatement plusieurs champs de questionnements :

- les SIA utilisent et surtout réutilisent des données. Cette réutilisation est le crédo des SIA. Les SIA permettraient de réutiliser les cimetières de données qui peuvent et doivent servir - les anglo-saxons parlent de « data graveyards », de les « sortir » et de les exploiter.

[17] https://european-research-area.ec.europa.eu/news/living-guidelines-responsible-use-generative-ai-research-published.
[18] https://www.helmholtz.ai.
[19] Voir aussi la page Inserm maintenue par le LIMICS : https://www.Inserm.fr/dossier/intelligence-artificielle-et-sante/.

C'est un vocable assez marketing[20] et cela oblige à réfléchir à la validité des données quand le contexte d'utilisation change. Ainsi, l'utilisation secondaire des données médico-administratives met en évidence un écart prononcé entre l'intention initiale de leur collecte dans un contexte spécifique et les réutilisations, parfois idéalisées, auxquelles elles sont soumises, que ce soit isolément ou en combinaison, en tenant compte de changements potentiels au fil du temps et des spécificités propres à chaque centre de collecte et/ou de traitement.

- Un LLM fouille de grandes quantités de textes et répond très bien à des questions générales mais la pertinence de sa réponse est souvent limitée pour des questions liées à des expertises précises, médicales ou biologiques. Si le réentraînement avec des données plus spécifiques peut permettre d'affiner les résultats, la qualité des réponses peut rester insatisfaisante dans la mesure où les modalités de raisonnement, l'expérience et les connaissances mobilisées ne sont pas toujours disponibles ni même explicitables sous forme de textes.

- Le ré/entrainement fait en médecine pose avec plus d'acuité qu'ailleurs la question de la véracité des réponses rendues par l'agent : le LLM a été entraîné pour produire des énoncés « vraisemblables » pour un lecteur humain en affichant des combinaisons de signes (par exemple du texte) mais ne peut discerner ce qui est véridique[21] de ce qui ne l'est pas. Dès lors, quelles que soient la taille et la qualité du corpus sur lesquels il s'appuie, on ne peut jamais être sûr de la fiabilité de la réponse produite. Des « hallucinations » du LLM restent possibles.

Pour l'apprentissage comme pour les LLM, si les données ou textes sont liés à des patients, les droits de ces patients doivent être respectés : secret professionnel et confidentialité des données, non utilisation pour la recherche sans information préalable, voire consentement des intéressés, application du RGPD (Règlement général sur la Protection des Données), Digital Act, Règlement européen sur l'intelligence artificielle dit « AI act » et autres dispositions du droit national…. Cette problématique de respect des droits des patients ou des volontaires

[20] https://grapheneai.com/the-ai-analyst-strategic-insights-and.-data/ ou https://medium.com/@renhaosu_12442/why-some-ai-ideas-end-up-in-the-graveyard-lessons-learned-c399854abbf7.

[21] « Discerner le vrai » est pris ici comme la possibilité de vérifier, par des procédures informatiques en particulier, la réalité d'un énoncé et donc d'évaluer son exactitude.

sains, est également présente lorsqu'il s'agit d'un essai clinique ou dans la réutilisation de données originales, collectées dans le cadre de la prise en charge médicale ou de précédentes recherches. Les hôpitaux répondent à cette problématique en organisant la réutilisation des données dans des Entrepôts de Données de Santé (EDS) protégés. Cette réutilisation dans le cadre de la médecine de ville ou dans des collaborations ville-hôpital est en cours de construction :voir la Plateforme des données de santé (PDS), également appelée Health Data Hub (HDH) et le référentiel EDS/demandes d'autorisation auprès de la CNIL pour la mise en place d'EDS de médecine libérale (en cours). Une proposition de stratégie nationale interministérielle sur l'utilisation secondaire des données de santé[22], a été ouverte à la concertation publique en octobre et novembre 2024. L'utilisation des données de santé peut présenter un caractère particulièrement intrusif lorsqu'elles concernent les patients ou les participants à une recherche, mais également lorsqu'elles reflètent les pratiques des professionnels de santé et de recherche qui les produisent. Il est ainsi important de noter qu'il existe plusieurs niveaux de confidentialité, tant pour les patients que pour les professionnels, et que la confidentialité n'est qu'un des nombreux enjeux qui peuvent rendre ces données sensibles.

- La capacité des SIA (symbolique ou numérique) à proposer des résultats (de raisonnement ou de calcul) ne doit pas les rendre indépendants d'une intervention humaine[23]. D'autant plus en médecine où la vie des patients est en jeu (même des dépistages sur images, très efficaces, engagent des responsabilités humaines). Un raisonnement éthique aussi bien déontologique que conséquentialiste appuie cette conclusion en se demandant quel serait le prix à payer pour se passer d'un être humain (gradient de qualité des décisions). Ceci est acté dans la loi « Informatique et liberté » (art. 47) et dans la loi de bioéthique (art. 17)[24] qui, de plus, demande une explicitation précise du fonctionnement pour les utilisateurs et une information explicite du patient. Une obligation d'explicabilité pèse donc sur les concepteurs, afin que les professionnels de santé utilisateurs puissent exercer un contrôle sur les résultats du traitement dans le

[22] Stratégie interministérielle pour construire notre patrimoine national des données de santé 2025-2028 (sante.gouv.fr) et https://esante.gouv.fr/actualites/consultation-publique-sur-lutilisation-secondaire-des-donnees-de-sante.

[23] Certaines personnes envisagent de doter de nombreux dispositifs de responsabilité juridique. Nous souhaitons préciser ici que jusqu'à présent c'est toujours le médecin ou le chercheur qui est responsable. Ceci est sous-entendu dans la loi bioéthique mais pas écrit explicitement.

[24] https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043884384.

respect de la garantie qu'apporte une intervention humaine (CSP, art. L. 4001-3) (voir Annexe 3).

- Enfin, les avancées récentes en IA résultent en grande partie de l'abaissement significatif de coûts d'accès aux données et du coût de mobilisation de calcul à la demande. Cependant, il est crucial de souligner que, comme toute technologie, ces outils ne sont jamais neutres, tant dans leurs phases de conception que dans leurs utilisations, et qu'ils peuvent avoir été entraînés sur des données qui ne reflètent pas les principes d'équité. En effet, chaque étape de leur développement, des algorithmes de base aux applications finales en passant par la mise en qualité des données initiales utilisées – avec souvent un énorme travail humain qui ne peut être neutre –, est imprégnée de choix et de valeurs qui reflètent les biais potentiels des concepteurs et leur inscription dans un contexte social. De plus, il est important de garder en tête qu'une grande partie de ces outils sont développés, à une vitesse extrêmement rapide, par des sociétés privées à but commercial, dans un marché concurrentiel, ce qui interroge la capacité des législateurs à garantir la qualité des services fournis dans l'intérêt du « public » et l'absence de biais portés par des enjeux économiques, politiques ou autres.

Questions éthiques concernant l'utilisation des SIA à l'Inserm

Les SIA transforment déjà et vont continuer à transformer profondément le quotidien des pratiques dans les laboratoires de l'Inserm, comme dans les services administratifs de l'Institut. Nous souhaitons donc ici envisager les conséquences de ces usages, les bénéfices et les risques potentiels, les questions ouvertes afin d'élaborer au chapitre suivant les propositions pour un usage éthique et responsable de SIA.

1. **L'évaluation du coût carbone réel et de l'empreinte écologique de l'utilisation de SIA à l'Inserm** :

- Responsabilité environnementale : le boom des SIA est associé à une explosion de l'empreinte écologique, qui met sérieusement à mal notre capacité à atteindre l'engagement pris par la France et l'Europe de zéro émission nette d'ici à 2050[25]. Le débordement du numérique et de SIA dans toutes nos sphères, induit un malaise grandissant face à un asynchronisme entre l'hyper accélération du présent sous l'effet de la transition numérique et la vitesse de développement de SIA et un blocage de l'avenir du fait du dérèglement climatique et écologique[26]. Il existe une opposition nette entre la transition numérique et plus encore avec le développement de SIA (une requête sur internet mobilisant des données déjà indexées est bien moins couteuse en puissance de calcul que le SIA sollicité pour le même service) et la transition socio-écologique[27]. En effet, la dématérialisation n'induit pas une réduction de l'exploitation des ressources naturelles - l'exploitation des ressources naturelles a triplé depuis les années 70[28]. En France, « le seul secteur des *data centers* pourrait mobiliser la puissance de cinq à sept réacteurs nucléaires d'ici 2030[1] ». A cela s'ajoute l'énergie nécessaire à l'extraction des métaux et des terres rares qui entrent dans la composition des terminaux, induisant des dommages écologiques catastrophiques et durables[29]. De plus, ce développement outrancier mobilise une ressource inestimable : l'eau.

  L'eau est transformée en effluents pollués et écocides par les méthodes d'extractions notamment des terres rares[30]. L'eau employée pour le refroidissement des équipements

[25] L'insoutenable coût écologique du boom de l'IA, Nastasia Hadjadji, juillet 2024 REPORTERRE
[26] Rapport de synthèse du GIEC 2023.
[27] L'impasse collaborative. Pour une véritable économie de la coopération. Éloi Laurent. Octobre 2018 ed. LLL
[28] International Institute for Sustainable Development 2021 DOSSIER #16 L'exploitation durable des ressources naturelles : Le défi de la gouvernance, Jennifer Bansard et Mika Schröder.
[29] Evaluation de l'impact environnemental du numérique en France et analyse prospective. Ademe Arcep, janvier 2022.
[30] Les terres rares : Quels impacts ? EcoInfo pour une informatique éco-responsable.

est aussi en continuelle augmentation avec une hausse de 17% entre 2022 et 2023 pour Google et de 34% entre 2021 et 2022 (soit le volume de plus de 2500 piscines olympiques[31]) pour Microsoft. Une étude récente établit que les systèmes d'IA pourraient consommer en 2027 entre 4,2 et 6,6 milliards de mètres cubes d'eau, soit une consommation légèrement supérieure à celle du Danemark et équivalente à la moitié de celle du Royaume-Uni[32]. Nous sollicitons ces deux géants de SIA par de nombreuses requêtes sans avoir parfaitement conscience des impacts écologiques et par conséquence sur la santé.

- Responsabilité sociale : la préconisation d'une évaluation et d'une régulation des émissions de gaz à effet de serre (GES) induit par l'emploi de SIA au niveau international[8], devrait aboutir à des préconisations de frugalité dans la création de nouveaux algorithmes de SIA et dans leur emploi, associées à une communication sur l'impact écologique qui permettraient aux utilisateurs de faire des choix raisonnés[8]. Ces choix raisonnés des utilisateurs nécessitent aussi d'être parfaitement informés sur l'analyse du cycle de vie de chaque terminal. Si la frugalité est un point d'entrée, il reste néanmoins un moyen insuffisant face à l'ampleur du défi climatique. Il est aujourd'hui impératif d'effectuer une transition vers la sobriété. Un emploi raisonné de SIA dans une visée de bien-être social en prenant en compte que les limites des ressources nécessitent un changement de cap impérieux.

2. **La question de la taille des systèmes d'IA**

Le principe de minimisation devrait nous conduire à nous poser la question « avons-nous toujours besoin d'utiliser les grands systèmes ? ».

Il existe des LLMs utilisant un jeu de données réduit et moins de paramètres et qui représentent à la fois une IA frugale (en termes de données d'apprentissage) et une IA sobre (en termes de consommation d'électricité en utilisation). On parle de modèles de langage petits (SLMs). Ils sont plus faciles à déployer avec des ressources limitées. On peut compenser un échantillon d'apprentissage plus petit et donc des performances théoriquement moins bonnes par le fait qu'on peut le rendre plus performant en le développant sur des

[31] Artificial intelligence technology behind ChatGPT was built in Iowa — with a lot of water. https://apnews.com/.

[32] Impacts de l'intelligence artificielle : risques et opportunités pour l'environnement. CESE Rapporteur.es Fabienne Tatot et Gilles Vermot Desroches Septembre 2024.

données plus spécifiques et que l'on maîtrise. Plusieurs entrepôts de données de santé s'appuient sur ces systèmes.

Cela répond à plusieurs enjeux éthiques et normatifs :

- en matière de souveraineté numérique, l'utilisation de SLMs peut renforcer l'indépendance technologique des institutions et des nations, en réduisant la dépendance aux grandes entreprises technologiques internationales ;
- en matière de respect de la vie privée, les modèles de plus petite taille peuvent être plus faciles à contrôler et à sécuriser, offrant de meilleures garanties pour la protection des données des patients et des participants à la recherche.

3. Les démarches de conception des systèmes d'IA : de façon ascendante ou descendante ?

Si au sein de l'Inserm, beaucoup sont utilisateurs de SIA, de plus en plus d'acteurs sont également des concepteurs de SIA, seuls ou en collaboration avec d'autres instituts de recherche comme l'INRIA, le CNRS ou avec des entreprises privées. Nous identifions comme enjeux éthiques :

° Inclusion et participation : une approche ascendante dans la conception peut encourager une plus grande participation des utilisateurs finaux et des parties prenantes, assurant que les systèmes d'IA répondent aux besoins réels et respectent les valeurs des communautés.

° le risque de technologies imposées : une approche descendante peut imposer des solutions technologiques sans un véritable consensus ou une compréhension des impacts locaux, ce qui peut susciter des résistances ou des utilisations non éthiques. Un soutien des pouvoirs publics à la pluralité des solutions proposées, en évitant les situations de monopole, pourra contribuer à se prémunir contre ces risques.

4. Qualité, conditions d'accès et de qualification des données

Les missions d'un organisme national de recherche public comme l'Inserm nous obligent à considérer :

- Équité et accessibilité : assurer un accès équitable aux données, encourager la production de données, y compris dans des domaines non immédiatement rentables économiquement, en évitant les discriminations basées sur les ressources ou les capacités technologiques.
- Intégrité et fiabilité : maintenir des standards élevés pour la qualité des données et leur qualification afin de garantir des résultats précis et fiables.

- Egalité d’accès : comme pour toute innovation technologique, il est possible que seuls quelques privilégiés en profitent, voire qu'ils exploitent les nouveaux outils d’IA au détriment de ceux qui n'y ont pas accès. L’accès aux données et aux ressources de calcul nécessaires aux SIA soulève ce type de dilemme. Il est essentiel de veiller à ce que l’émergence de ces technologies ne contribue pas à accentuer des inégalités. Prenons l'exemple de l’instauration d’une redevance pour l’accès en vue d’un usage secondaire, à des données déjà générées et financées par la solidarité nationale, comme celles du SNDS (Système National des Données de Santé). Une telle redevance pourrait avantager les grands acteurs disposant de moyens financiers importants, au détriment des plus petits, freinant ainsi l’émergence de nouvelles entreprises ou initiatives dans le secteur. Une vigilance particulière des pouvoirs publics est donc nécessaire pour éviter que des barrières économiques ne renforcent les disparités déjà existantes.
- Science ouverte : les développements des SIA génératifs sont largement tributaires et bénéficiaires du soutien fort des institutions publiques à l’ouverture des données publiques, dont les données de recherche. Il est important que ce mouvement d’ouverture motivé par le souci de partage des savoirs dans la société ne soit pas entravé mais amplifié. Les algorithmes et outils d’IA développés grâce à des données ouvertes doivent rester accessibles, à moindre coût, pour les institutions publiques de recherche.

5. Vigilance concernant le déplacement de l’expertise vers l’exécution, du savoir vers le savoir-faire

Les SIA sont un sujet de recherche à l’Inserm avec des chercheurs qui travaillent les questions théoriques et dont certains vont jusqu’à la mise au point « d’outils » ou de démonstrateurs. Toutefois la majorité des personnels amenés à utiliser les SIA à l’Inserm n’en sont pas les concepteurs. Une attention particulière devra donc être portée à :

- Perte d'expertise : le déplacement excessif de l'expertise humaine vers les SIA peut entraîner une perte de compétences critiques, une dépendance excessive aux systèmes automatisés. Cela peut aussi engendrer des situations de vulnérabilité professionnelle face à un outil dont la maitrise parfaite est une affaire de spécialistes. Les enjeux de formation et d’acquisition d’expertise sont majeurs. Comment acquiert-on- une expertise complémentaire à celle des SIA, si on laisse toujours les SIA « réfléchir et

analyser » à notre place ? Comment accompagner les changements dans les pratiques d'apprentissage ?

- Évaluation critique : il est crucial de maintenir la capacité de questionner et de comprendre les jeux de données, afin d'éviter des biais et des erreurs dans les analyses et les décisions basées sur les SIA. Pour éviter la délégation totale de tâches à des SIA et la mise sous tutelle de l'utilisateur par les SIA, la réglementation contraint les développeurs industriels de SIA à une transparence sur les méthodes qu'ils utilisent, la clarté et l'intelligibilité étant nécessaires au contrôle et à l'explicabilité par l'utilisateur. Concernant un développement en milieu académique, l'Inserm devra définir le cadre précis à utiliser.
- Vérification de la transposabilité de l'algorithme à la recherche en cours : les chercheurs doivent obligatoirement évaluer les modèles qu'ils utilisent et s'assurer de leur caractère Fair, Appropriate, Valid, Effective and Safe (FAVES) dans le cadre de leur travail de recherche. De la même façon que les hôpitaux doivent évaluer les SIA localement, il appartient aux équipes de garantir le bon usage des algorithmes et de le documenter.

6. Place de l'hypothèse scientifique dans la génération et le questionnement des jeux de données

L'intégrité scientifique réclame une attention particulière à :

- Rigueur scientifique : maintenir l'importance de l'hypothèse scientifique et de la méthodologie rigoureuse dans l'utilisation des données et des systèmes d'IA pour éviter des conclusions hâtives, incorrectes ou la reproduction de normes sociales discriminantes.
- Transparence et reproductibilité: assurer la transparence des processus et des modèles utilisés pour que les hypothèses et les résultats puissent être vérifiés et reproduits. La reproductibilité est un des piliers essentiels de l'intégrité scientifique. L'Inserm en a fait un de ses axes prioritaires au sein du programme LORIER.

7. Point de vigilance sur la Propriété Intellectuelle (PI)

Enjeux éthiques :

- Justice et accès : garantir que les innovations en IA sont accessibles et bénéfiques pour tous, et non pas restreintes par des barrières de propriété intellectuelle excessives. La question de l'accessibilité aux codes sources soulève des enjeux multiples et complexes. L'Inserm pourrait émettre des recommandations tendant à favoriser l'Open Science et la Transparence en encourageant la publication des données, modèles et algorithmes dans des formats ouverts, tout en respectant les normes éthiques et de confidentialité. Cette approche faciliterait le partage des avancées en IA tout en renforçant la collaboration avec d'autres chercheurs et instituts. L'Inserm pourrait également recommander d'utiliser des Licences Open Source, de développer et diffuser les outils d'IA sous licences open source, permettant à d'autres chercheurs et organisations de les utiliser, les modifier et les distribuer, tout en évitant les restrictions d'accès imposées par des brevets restrictifs. L'Inserm pourrait également favoriser des partenariats avec des acteurs publics et privés afin de construire des alliances avec d'autres organismes de recherche, entreprises technologiques et autorités publiques pour soutenir le développement d'innovations accessibles, tout en évitant les accords de licence exclusifs qui pourraient limiter l'accès public aux technologies développées. Enfin l'Inserm pourrait établir une politique de brevets équilibrée. En cas de dépôt de brevets pour protéger certaines innovations, l'Inserm pourrait privilégier des licences non exclusives ou des licences de type « patent pool » qui facilitent l'accès de la technologie à plusieurs acteurs en échange d'un faible coût, voire gratuitement pour les organismes publics.
- Reconnaissance : assurer une reconnaissance et une compensation équitables pour les contributions individuelles et collectives dans le développement de technologies d'IA.

Enjeux juridiques :

- Risque de perte du critère de priorité ou du critère « innovant ». Une divulgation de résultats de la recherche, ou d'hypothèses argumentées sur des résultats préliminaires, par injection d'une interrogation dans un LLM (un prompt contenant tout ou partie de la découverte brevetable), menace de faire perdre « l'originalité » ou « l'antériorité » et compromet l'obtention ou la reconnaissance d'un droit de propriété intellectuelle s'agissant de données considérées dès lors comme déjà publiées dans un SIA. En cas d'obtention de résultats trop directement « inspirés » de contenus protégés (code

informatique posté sur un forum…), le risque de contrefaçon, à l'insu de l'utilisateur, est accru.

8. Modes de vie induits par les SIA et conséquences des SIA sur les pratiques de recherche et les façons de prendre soin

Une des dimensions éthiques à considérer est la manière dont la technologie façonne son utilisateur autant voire plus que l'inverse :

- Changements sociétaux : examiner comment l'usage des SIA modifie les modes de vie, les pratiques d'enseignement, de travail et les relations sociales. Evaluer les implications éthiques de ces changements et prévenir des situations de vulnérabilité détectables en amont (ethics-by-design/Réflexion-Anticipation-Délibération).
- Qualité des soins : assurer que les outils d'IA élaborés à l'Inserm améliorent la qualité des soins, par exemple en donnant plus de temps aux soignants pour mieux personnaliser la prise en charge, sans déshumaniser les interactions entre les patients et les professionnels de santé. Il existe en conséquence des enjeux financiers et organisationnels, en particulier sur l'équité de la prise en charge. Si le recrutement d'informaticiens devient nécessaire pour contrôler et conserver la distance critique envers les IA, la vigilance sur les effets indésirables induits pour le soin – comme une diminution corollaire du nombre de soignants- est de mise.
- Prévention et contrôle social : si l'usage des SIA accompagne des pratiques de prévention, ou de diagnostic plus précoce, il faudra mesurer le risque en termes de nouvelles pratiques de surveillance généralisée[33] ainsi que le risque de sur-diagnostic bien connu dans le contexte des dépistages (cf aussi le concept de biopouvoir de Michel Foucault : échange de plus de vie contre plus de contrainte de surveillance).
- Pratiques de la recherche : anticiper les effets d'une capacité à traiter rapidement des quantités de données et de références bibliographiques toujours plus importantes sur la qualité et la quantité des publications scientifiques produites, et sur les modalités d'évaluation de la recherche.
- Pratiques administratives : anticiper les effets de l'utilisation de SIA dans le domaine des ressources humaines voir des recrutements.

[33] David Armstrong The rise of surveillance medicine Sociology of Health and illness Volume17, June 1995 Pages 393-404.

9. Avantages immédiats versus effets potentiellement délétères à plus long terme

Les SIA sont une technologie récente, en pleine évolution, et dont l'impact sur la durée n'est pas évalué :

- Prudence : équilibrer les avantages immédiatement perçus des technologies d'IA avec une évaluation prudente des effets à long terme, pour contrôler la balance bénéfices/risques. L'évaluation des approches basées sur SIA doit intégrer l'efficacité, la tolérance et l'impact social à l'instar de la modification de l'organisation des soins. L'enjeu éthique est de taille au regard de notre volonté de ne pas retarder l'accès à des innovations performantes. Pour compléter le principe de garantie humaine qui figure dans l'AI Act et l'évaluation « by design », des évaluations de plus grande échelle après la mise sur le marché sont recommandées, à l'instar des études de phase 4, en vie réelle, pour les médicaments. La création d'une cellule nationale et transversale sur les usages du numérique pour les sciences pour la santé à l'Insermpermettra également d'élaborer des scénarii d'impact à long terme.
- Responsabilité intergénérationnelle : prendre en compte les impacts à long terme sur les futures générations, en s'assurant que les bénéfices actuels ne compromettent pas leur bien-être.

10. Protection de la vie privée versus intérêt espéré de la recherche : l'intérêt des données synthétiques

- Le RGPD (Règlement Général sur la Protection des Données) a, à juste titre, accentué l'attention portée à la protection de la vie privée. L'exploitation des données personnelles reste possible, à condition de mettre en place des mesures adaptées pour garantir la confidentialité. Dans le domaine de la santé et de la recherche biomédicale, ces enjeux sont encore plus critiques compte tenu du caractère sensible des données traitées. La différence entre « données anonymes » et « données pseudonymes » n'est pas toujours bien comprise:
- La pseudonymisation consiste à retirer les informations directement identifiantes (nom, prénom, etc.) d'un jeu de données, mais celles-ci conservent des éléments indirectement identifiants (comme une combinaison unique de caractéristiques),

rendant possible la réidentification par croisements de données ou décryptage d'une identité masquée par un code conservé dans une table de correspondance[34].

- En revanche, l'anonymisation est un aller simple : un processus irréversible qui empêche l'identification d'une personne. Les autorités européennes de protection des données (ancien G29, EDPB, et la CNIL en France) ont réaffirmé trois critères qui, s'ils sont conjointement réunis, permettent de garantir qu'un jeu de données est véritablement anonyme : individualisation, corrélation et inférence[35]. Notons ici que la réglementation en vigueur a presque 10 ans avec les critères du G29[36].

- C'est dans le contexte de l'anonymisation que la génération de données synthétiques prend tout son sens : recréer des données d'entrainement pour des LLM sans problème de confidentialité, en exploitant la puissance de calcul des modèles mathématiques. Plusieurs méthodes existent. Certaines reposent sur des principes mathématiques, d'autres visent à répondre par construction aux critères réglementaires d'individualisation, de corrélation et d'inférence. Certaines approches utilisent des modèles basés sur les données elles-mêmes, allant jusqu'à créer un modèle local pour chaque observation, tandis que d'autres sont fondées sur des équations purement théoriques. *In fine*, les données synthétiques, sont entièrement simulées mais réalistes. Elles reproduisent les caractéristiques des données originales, permettent des analyses similaires, tout en garantissant qu'il est impossible en pratique de remonter de manière certaine à des individus réels. A titre d'exemple concret, plusieurs CHU sont impliqués dans le projet PARTAGES coordonné par le Health Data Hub, un projet de LLM avec une base commune de comptes rendus synthétiques rédigés par des internes (et d'autres médecins hospitaliers). La base partagée est un fichier de type « commun numérique » que l'on pourra utiliser partout. Un reparamétrage local sera fait avec les bases de documents locaux à chaque CHU et ne pourra être utilisé à l'extérieur. Le recours aux données synthétiques anonymes à chaque fois que cela est possible

[34] Définie à l'article 4 du RGPD, la pseudonymisation est « le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable ».

[35] Pour en savoir plus, voir les lignes directrices. Celles-ci sont en cours de mise à jour mais les trois critères exposés demeurent valides.

[36] Singularité corrélation inférence. (https://www.cnil.fr/fr/le-g29-publie-un-avis-sur-les-techniques-danonymisation ; https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216, en.pdf).

deviendra rapidement une exigence éthique. Elles constituent également un outil privilégié pour favoriser l'ouverture des données dans le cadre de la recherche[37]. Enfin l'utilisation de données synthétiques peut permettre d'améliorer la représentativité d'un jeu de données et la correction de biais qualitatifs et/ou quantitatifs par augmentation.

Il convient d'indiquer ici que des collègues considèrent dans la phase actuelle que les attentes vis à vis des données synthétiques leur apparaissent à la fois très incertaines, excessives, et non appuyées le plus souvent sur des éléments de preuves suffisants. Ces collègues alertent sur le risque d'un développement mercantile inapproprié de ces jeux de données synthétiques.

Ces réserves conduisent à préciser les contextes d'utilisation des données synthétiques :

Il y a d'abord des données synthétiques que l'on peut qualifier d'anonymes, et d'autres pas, dont certaines produites par confidentialité différentielle qui est confortable mathématiquement et a fait en effet l'objet de la création de plusieurs entreprises, dont par exemple Hazy récemment rachetée par SAS. Les usages sont multiples, mais ne serait-ce que celui qui consiste à publier les données permettant de reproduire les analyses d'un article[38] semble justifier un intérêt pour l'Inserm : codes, données anonymes au sens RGPD et comparaison avec les données pseudonymisées disponibles[39].

En second lieu, il faut être attentif à ne pas introduire de confusion : dans les données synthétiques, le critère essentiel est le caractère plus ou moins sensible des données et au premier chef leur potentiel ré-identifiant ou non pour les participants à une recherche. Cette nature ré-identifiante pour des personnes justifie leur régulation légale et exige qu'on les qualifie d'anonymes ou non. De fait, même anonymisées en apparence, elles peuvent rester sensibles. Il s'agira donc de travailler aussi à la meilleure manière de les « désensibiliser ». Comme il existe plusieurs techniques de production de données synthétiques (plus ou moins "data-driven", plus ou moins "model-driven"), il y a toujours des hypothèses scientifiques et techniques dans le

[37] En ce sens, le livre blanc *Données de santé artificielles : analyse et pistes de réflexion*, avril 2024.
[38] https://jnnp.bmj.com/content/92/2/122.
[39] https://github.com/ICAN-aneurysms/RIA-predict/tree/master/notebooks.

processus de production des données et des modélisations[40]. Il convient donc pour chaque cas de rechercher où pourrait exister une faille de sécurité. Il existe également plusieurs techniques de contrôle des usages des données: blockchain, fédération... Les données synthètiques peuvent d'ailleurs elles-mêmes permettre de désensibiliser des données personnelles. Ainsi dans le RHU PRIMUS, les données synthétiques anonymes sont utilisées pour désensibiliser les données des essais cliniques dans la SEP : leur mise en qualité est déterminée par l'usage, ici l'alimentation pour usage intensif d'un outil d'aide à la décision. De surcroît, un modèle fédéré propose un contrôle des usages des données synthétiques par les responsables légaux des données.

11. Biais et hallucinations liées à la langue utilisée ou à la voix

Les résultats des SIA étant basés sur la représentativité des données, ce qui est faiblement représenté est mal traité. Une étude publiée début octobre dans le BMJ[41] met en garde contre l'utilisation des chatbots basés sur des LLM en santé en raison de leurs biais linguistiques. Des chercheurs des universités de Hong Kong, de Singapour et du RMIT au Vietnam ont constaté que le module d'intelligence artificielle (IA) générative GPT-3.5 de ChatGPT (OpenAI) fournit des informations erronées lorsqu'on l'interroge dans une langue peu répandue. Questionné en vietnamien, le chatbot a ainsi interprété à tort les symptômes de la fibrillation auriculaire comme étant ceux de la maladie de Parkinson. Les chercheurs recommandent d'améliorer les capacités de traduction des modèles via le partage de données linguistiques diverses en *open source*. C'est la loi des séries pour OpenAI, qui doit aussi répondre aux critiques touchant son outil de reconnaissance vocale Whisper, accusé d'"hallucinations" fréquentes. Notamment utilisé par la start-up tricolore Nabla, qui conçoit un logiciel de transcription médicale doté d'IA, l'outil invente fréquemment des passages de texte entiers lorsqu'il est confronté à des moments de silence. Ces hallucinations faussent les comptes rendus médicaux et augmentent *in fine* la charge de travail des médecins qui doivent alors vérifier méticuleusement le contenu généré… perdant le temps prétendument gagné. OpenAI déconseille "l'utilisation de Whisper dans les domaines à haut risque".

[40] Guillaudeux et al. Patient-centric synthetic data generation, no reason to risk re-identification in biomedical data analysis, https://www.nature.com/articles/s41746-023-00771-5.

[41] Kwok et al. Health information for all: do large language models bridge or widen the digital divide? *BMJ 2024;387:e080208, doi: https://doi.org/10.1136/bmj-2024-080208 (Published 11 October 2024).*

Les propositions de recommandation[42]

1- Divulgation et attribution transparentes

Les personnels de l'Inserm devraient indiquer clairement l'utilisation des SIA dans le cours de leur recherche, en particulier les outils, algorithmes et paramètres spécifiques utilisés, y compris dans les réponses aux appels à projets et dans la section « matériel et méthodes » des articles. Les sources d'information ou d'idées issues directement de SIA, devraient être indiquées de façon précise, en distinguant bien les contributions revenant aux idées humaines et celles directement proposées par des SIA. Le souci de citer dûment l'expertise humaine et la littérature antérieure doit être maintenu, en particulier lorsque les machines ne fournissent pas de telles citations dans leurs résultats. Pour accompagner les personnels Inserm dans leurs tâches quotidiennes, qu'ils soient chercheurs, personnels techniques ou administratifs, nous préconisons la mise en place d'un parcours de formation adapté à leurs enjeux professionnels.

En pratique : prévoir dans les « Matériels et Méthodes » des articles soumis une partie "Utilisation de SIA" avec comme sous-chapitres les contextes de leur utilisation

- Pour définir la question posée
- Pour définir la méthode choisie et/ou pour répondre à cette interrogation
- Pour faire la recherche bibliographique
- Pour en faire la synthèse
- Pour rédiger tout ou partie de l'article

De même que dans la partie statistique il est mentionné que les calculs ont été réalisés avec tel ou tel logiciel, les IA utilisées seraient indiquées. Préciser également quel est (quels sont) les auteurs humains qui ont supervisé l'éventuel travail de SIA.
Une proposition très développée a été proposée dans un éditorial récent du JAMA[43].

En dialogue avec les concepteurs et développeurs de SIA, les personnels Inserm devraient :

- s'assurer que les concepteurs de SIA utilisés pour leur recherche respectent les règles éthiques et la règlementation en vigueur

[42] Nous indiquons avoir repris certaines des propositions formulées par Blau et al. https ://www.pnas.org, May 29, 2024 en les adaptant à la situation particulière de la recherche en santé.
[43] Flanagin et al. JAMA April 2, 2024 Volume 331, Number 13.

- fournir des détails accessibles au public sur les conditions d'obtention des modèles produits grâce aux SIA, en particulier les données utilisées pour les créer ou les affiner;
- gérer et publier soigneusement les informations sur les modèles et leurs variantes afin de fournir aux autres scientifiques un moyen de citer l'utilisation de modèles particuliers avec spécificité ;
- fournir des archives à long terme des modèles pour permettre des études de réplication;
- Ecarter un résultat lorsque l'attribution correcte du contenu généré ne peut pas être fournie;
- poursuivre les innovations dans les machines d'apprentissage, de raisonnement et de recherche d'informations visant à fournir aux utilisateurs de ces modèles la capacité d'attribuer/reconnaitre les sources et la paternité (authorship) des données utilisées dans le contenu généré par l'intelligence artificielle. Etre transparent avec les évolutions et en expliquer la cause et la nature ;
- permettre des usages gratuits ou à moindre coût pour les institutions de recherche académiques lorsque les développements ont mobilisé des données librement disponibles ou mises à disposition par des acteurs publics.

2. Vérification du contenu et des analyses générées par les SIA

Les personnels de l'Inserm sont responsables de l'exactitude des données, des images et des déductions qu'ils tirent de leur utilisation des modèles d'IA génératives. La responsabilité exige l'utilisation de méthodes appropriées pour valider l'exactitude et la fiabilité des déductions faites par ou avec l'aide de SIA, ainsi qu'une divulgation complète des éléments de preuve pertinents pour ces déductions. Il s'agit notamment de contrôler et de tester les biais dans les algorithmes et les résultats de SIA, dans le but d'identifier et de corriger les biais susceptibles de fausser les résultats de la recherche ou les interprétations.

Les personnels travaillant à l'Inserm doivent connaître et signaler les limites de la capacité des systèmes à confirmer la véracité des données, textes ou images générés par SIA. L'Inserm devrait leur en donner les moyens. La création d'une cellule nationale et transversale sur les usages du numérique pour les sciences pour la santé à l'Inserm permettra de recueillir ces informations et de contribuer à la formation. Cette structure nouvelle pourrait s'intégrer au programme LORIER. Lorsque la vérification de la véracité du contenu généré n'est pas

possible, les résultats du modèle doivent fournir des évaluations claires et bien calibrées de la confiance. Les créateurs de modèles doivent identifier, signaler et corriger de manière proactive les biais des algorithmes d'IA susceptibles de fausser les résultats de la recherche ou leurs interprétations. Soulignons ici qu'il convient de distinguer deux types de biais : ceux des SIA eux-mêmes, liés à la base de données et à son mode de traitement, et ceux liés à l'utilisateur (qui s'exprime dans les questions qu'il pose aux SIA).

3. Documentation des résultats produits par des SIA

Les personnels de l'Inserm doivent marquer (incorporation de filigranes) les données, les déductions et les images générées avec l'aide de SIA ou les données synthétiques avec des informations de provenance sur le rôle de SIA dans leur création, afin qu'elles ne soient pas confondues avec des observations collectées dans le monde réel. Les scientifiques ne doivent pas présenter le contenu généré par SIA comme des observations recueillies dans le monde réel.

Les personnels de l'Inserm en dialogue avec les concepteurs de modèles doivent clairement identifier, annoter et conserver la provenance des données synthétiques utilisées dans leurs procédures de formation et surveiller les problèmes, les préoccupations et les comportements découlant de la réutilisation de contenus générés par ordinateur pour la formation de futurs modèles.

4. L'importance de l'éthique et l'équité

Les personnels de l'Inserm sont appelés à la prudence dans le recours aux SIA, en prenant bien en considération ses limites, ses biais potentiels, ses impacts et à mettre en œuvre des stratégies de vérification pour s'assurer que leur utilisation de SIA produit des résultats scientifiquement valables et socialement bénéfiques, tout en prenant des mesures appropriées pour atténuer le risque de préjudice. Ici encore, la formation de tous les acteurs de la recherche, y compris les étudiants et le personnel administratifs, est essentielle.

Les personnels de l'Inserm doivent respecter les lignes directrices éthiques relatives à l'utilisation de SIA, notamment en ce qui concerne l'attribution claire des sources de données issues de l'observation par rapport à celles générées avec l'aide de SIA, le respect du droit d'auteur et autres droits de la propriété intellectuelle, le respect du droit relatif aux données personnelles et à la protection de la vie privée, le respect des règles de divulgation de

certaines données spécifiques et le respect des règles concernant le consentement à l'accès à ces données, ainsi que la détection et la correction des biais potentiels dans la construction et l'utilisation des systèmes d'IA.

Les personnels de l'Inserm devraient également maintenir une vigilance sur les conséquences sociétales susceptibles d'apparaître au fur et à mesure du développement et du déploiement de SIA, et mettre à jour les pratiques et les règles qui favorisent les utilisations bénéfiques et atténuent la perspective de dommages écologiques et sociaux.

5. Une attention particulière aux données personnelles utilisées avec des SIA ou générées à partir des SIA

Les personnels de l'Inserm sont appelés à utiliser des données personnelles de patients ou participants à des cohortes, uniquement lorsque leur recherche l'impose. Une manière de développer la recherche sans risque d'atteinte aux données personnelles sensibles est d'utiliser des données synthétiques. Trois dimensions des données synthétiques sont particulièrement importantes : la confidentialité, la vérisimilitude et l'augmentation des données.

a. La confidentialité est assurée par la génération simulée de données synthétiques anonymes, permettant ainsi le partage de données sans être soumis aux contraintes des réglementations sur les données personnelles, comme le RGPD et les critères du G29/EDPB en Europe. Le travail de génération de données synthétiques doit se concentrer sur la garantie contre toute forme de ré-identification et sur une documentation rigoureuse de ces efforts. En effet, en cas de petits effectifs, par exemple les maladies rares, même des données simulées mathématiquement peuvent être considérées comme personnelles si elles sont trop proches des données réelles ;

b. La vérisimilitude consiste à produire des données simulées les plus réalistes possibles. Elles reproduisent fidèlement les ensembles de données réels préservant la multiplicité des usages ultérieurs. Des données synthétiques bien élaborées doivent apparaitre plausibles tout en évitant les biais de représentativité. Atteindre une forte vérisimilitude garantit la valeur des ensembles de données synthétiques, faisant de la confidentialité un avantage supplémentaire, plutôt qu'un objectif premier.

c. Enfin, l'augmentation des données est un autre aspect essentiel, plus ancien peut-être, où les données synthétiques permettent un enrichissement de données existantes en générant des données supplémentaires, spécifiques à un contexte particulier.

L'Inserm et les chercheurs sont garants de la confidentialité des données sensibles qu'ils détiennent et ne doivent en aucun cas injecter des données de santé relatives à des patients ou à des participants à des recherches, potentiellement identifiantes, dans un *Chat* sans maîtrise des réutilisations et destinataires (pour demander au SIA de générer des données fictives similaires….). Cette obligation devrait clairement figurer dans la Charte informatique de l'institution.

Les personnels de l'Inserm et la gouvernance de l'Institut sont appelés à promouvoir l'équité en étant vigilants sur les questions pour lesquelles les SIA sont utilisés. Ils sont également appelés à promouvoir un accès équitable aux outils d'IA et aux formations aux usages des SIA. Ces efforts devraient permettre à une communauté diversifiée de chercheurs scientifiques d'exploiter efficacement les SIA et de répondre à la diversité des besoins de recherche en santé, y compris ceux des groupes de personnes ou de patients traditionnellement peu pris en considération ou marginalisés. En outre, il convient d'étudier et d'utiliser des méthodes permettant de solliciter une participation significative du public à l'évaluation de l'équité et de la justice dans l'utilisation des technologies de SIA.
Les SIA ne devraient pas être utilisés sans une surveillance humaine attentive lors des étapes décisionnelles des processus d'évaluation par les pairs ou des décisions relatives à l'avancement professionnel et à l'attribution de financements.
Dans le cadre régi par la gouvernance des données traitées, promouvoir une politique, voire une offre de service en direction des équipes de recherche privilégiant des SIA écologiquement sobres et sécurisés hébergés dans des infrastructures de confiance conformes à la réglementation en vigueur.

6. Contrôle continu, surveillance et engagement public

Les personnels Inserm, ainsi que les représentants du monde universitaire, de l'industrie, du gouvernement et de la société civile, sont appelés à surveiller et évaluer en permanence l'impact des SIA sur les émissions de gaz à effet de serre, sur le processus scientifique, sur les conditions du travail scientifique, sur la qualité des résultats scientifiques publiés et, en toute transparence, à adapter si nécessaire les stratégies pour maintenir l'intégrité. Les technologies de SIA évoluant rapidement, les communautés de chercheurs doivent continuer à examiner et à comprendre les apports, les déficiences et les influences des SIA, s'efforcer d'anticiper et de prévenir les utilisations néfastes et d'exploiter leur potentiel pour relever des défis sociétaux cruciaux. Les spécialistes des SIA doivent en même temps s'efforcer d'améliorer l'efficacité

des SIA pour les sciences, notamment en s'attaquant aux problèmes de véracité, d'attribution, d'explication et de transparence des données d'apprentissage et des modèles générés. Des efforts devraient être entrepris dans les différents domaines scientifiques pour poursuivre l'étude permanente de l'état et de la dynamique de l'utilisation de SIA dans la recherche et mettre au point de nouvelles procédures pour intégrer les points de vue et les expériences des personnes concernées (patient.es, professionnel.les, usager.es, public plus large) au fur et à mesure que les SIA sont développés, appliqués et réglementés. L'existence de ces démarches participatives et leurs résultats devraient être largement diffusés.
Dans le cadre de la mise en place de la cellule nationale et transversale sur les usages du numérique pour les sciences pour la santé à l'Inserm, nous recommandons à la gouvernance de l'Institut la mise en place d'un portail web permettant le signalement de dérives ou d'inadaptations.

Les personnels Inserm sont appelés à évaluer de façon collective (dans les labos, les communautés de pratiques…) la transposabilité des SIA à leur recherche, les impacts positifs ou problématiques de ces outils sur leurs pratiques individuelles et collectives, notamment pour ce qui est des logiques de valorisation/publication de la recherche, de l'impact sur leurs expertises et ses transmissions. La cellule nationale et transversale sur les usages du numérique pour les sciences pour la santé à l'Inserm sera en charge de collecter et partager au sein de la communauté scientifique l'expertise nouvelle ainsi constituée.

Etant donné l'impact de l'usage des SIA sur l'environnement, une recherche orientée vers la sobriété (IA frugale) est en élément stratégique majeur à considérer dans le développement et l'utilisation de SIA à l'Inserm.

7. La formation des personnels à l'Inserm

Il est capital de s'assurer de la formation régulière du personnel de l'Inserm en tant qu'utilisateur, développeur, chercheur en IA. Cela couvre les SIA en tant qu'objets de recherche ou en qualité d'outils pour écrire/conduire les projets de recherche. Il s'agit de former aux principes, aux limites et aux enjeux des SIA tant pour les étudiants que les personnels titulaires.

8. Création d'un passeport de la donnée de santé

La qualité et la pertinence des données est centrale pour l'efficacité des systèmes d'IA utilisés en recherche biomédicale et en santé. Il est donc essentiel de pouvoir disposer des métadonnées associées. Des travaux commencent à rapporter les conséquences éthiques de l'utilisation abusive ou inappropriée de données en contexte IA. De la même manière, il convient de s'interroger sur les questions posées par la réutilisation de données qui sont collectées pour un usage spécifique puis employées dans un autre contexte, et d'en explorer les dimensions scientifiques, éthiques et légales. En effet les limites de réutilisation, qui reposent notamment sur une maitrise des différences entre le contexte de recueil des données et celui de leur réutilisation sont difficiles à identifier pour un humain, et le sont plus encore pour une IA.

C'est pourquoi nous recommandons la généralisation de l'association à toute base de données, de métadonnées permettant d'en décrire les caractéristiques, en particulier l'origine et la méthodologie d'obtention des données. Il s'agirait de créer un passeport des données, qui constituerait un cadre éthique et technique visant à documenter l'origine, la qualité, les biais potentiels et les transformations opérées au fil de la vie de la donnée (normalisation, filtrage, ..) pour en assurer une réutilisation appropriée et responsable. Ce passeport servirait de support à l'octroi de visas des données, qui agiraient comme un filtre à chaque réutilisation. Le visa permettrait non seulement de valider l'adéquation des données à de nouveaux contextes d'utilisation, mais aussi de tracer leur parcours à chaque réutilisation. Il inclurait des informations sur les modifications, transformations ou validations appliquées, sur les problèmes ou succès rencontrés, et des indications sur les utilisations futures possibles.

Ce dispositif fournirait :

- plus de transparence et de traçabilité : le passeport et le visa des données garantissent la traçabilité complète de l'origine et des usages des données. Cela renforce la responsabilité dans la manière dont les données sont traitées, évitant ainsi une réutilisation aveugle qui pourrait entraîner des biais ou des erreurs ;
- une meilleure prévention des biais : évaluer et documenter les biais potentiels dès la collecte initiale (population, contexte, granularité) aide à éviter leur amplification lors de la réutilisation, un aspect éthique crucial surtout dans les contextes médicaux ;
- une meilleure conformité légale et éthique : le visa pourrait aussi inclure des informations sur les contraintes juridiques et éthiques liées aux données

(consentement, RGPD), garantissant que les données ne sont pas réutilisées de manière illégale ou contraire à l'éthique ;

- une meilleure protection vis-à-vis des réutilisations abusives : en limitant l'utilisation et la réutilisation des données à des contextes compatibles, le passeport et le visa limiteraient l'élaboration de résultats faux et conclusions erronées qui pourraient nuire, par exemple, à des populations mal représentées dans le jeu de données.

En résumé, la création des passeports et visas des données peuvent améliorer les utilisations et les réutilisations de données assistées par SIA en contribuant à les rendre plus transparentes et plus responsables, et en minimisant ainsi les risques de biais, de mauvaises interprétations et de non-conformité qui, en recherche biomédicale, sont porteurs de conséquences éthiques potentiellement dramatiques.

9 - Création d'un portail de transparence de la donnée à l'Inserm

Créer un portail de transparence pour un organisme de recherche biomédicale comme l'Inserm pourrait être une avancée majeure pour instaurer la confiance avec les donneurs de données de santé à visée de recherche biomédicale, tout en répondant aux exigences de réglementation et en favorisant une recherche agile:

1. Transparence

Le portail serait avant tout un outil de transparence envers les participants aux études et les fournisseurs de données de santé, appelés « donneurs » ci-après. Il permettrait de :

- Informer clairement les producteurs des données et les donneurs sur l'utilisation de leurs données, les objectifs de la recherche et les retombées potentielles.
- Présenter les projets de recherche en cours de manière accessible, avec des résumés compréhensibles et des mises à jour régulières sur l'avancement des travaux.
- Donner des informations sur la gestion des données (durée de conservation, types d'analyse, niveau de confidentialité, etc.), en précisant les mesures de sécurité mises en place pour protéger la vie privée des donneurs.
- Offrir une traçabilité complète des données en permettant aux individus et organismes ayant consenti à l' (ré) utilisation de leurs données pour la recherche de voir quand et comment leurs données sont utilisées, ainsi que les résultats et publications issus de ces travaux.

2. Outil de Réglementation

Le portail agirait comme un outil de conformité réglementaire en intégrant des éléments essentiels liés aux lois et réglementations sur les données personnelles, notamment le RGPD. Cela inclut :

- Faciliter le recueil et la gestion du consentement éclairé, avec un suivi en ligne des autorisations et une option de rétractation simple et rapide.
- Garantir l'accès aux informations légales : le portail afficherait des informations relatives aux droits des donneurs (droits d'accès, de rectification, d'opposition, etc.) et aux procédures de gestion des réclamations.
- Assurer une transparence juridique sur les partenariats et les collaborations, en informant les donneurs des entités impliquées et des transferts éventuels de données, notamment en dehors de l'UE s'ils ont fait l'objet d'un accord spécifique.

3. Instrument d'Agilité de la Recherche

En parallèle, ce portail de transparence serait un levier d'agilité pour la recherche en permettant une gestion plus réactive et collaborative des données de santé, notamment grâce à :

- Des mises à jour rapides des protocoles de recherche et de la réglementation, que les donneurs pourraient suivre en temps réel. En cas d'évolution d'un projet, les participants pourraient être informés et reconfirmés en quelques clics.
- Une communication ouverte et interactive avec les participants : le portail pourrait intégrer des fonctionnalités permettant aux chercheurs de solliciter l'avis ou le consentement des donneurs pour de nouvelles études ou l'utilisation élargie de leurs données.
- Un accès facilité aux données pour les chercheurs dans le respect des normes de sécurité, afin d'accélérer la mise en place de projets sans délais administratifs complexes, tout en respectant les obligations réglementaires.

ANNEXE 1

Membres du Groupe de Travail (Juillet 2024 - Novembre 2024)

Pour le Comité d'éthique de l'Inserm : Henri Atlan, Catherine Bourgain, Hervé Chneiweiss, François Eisinger, Catherine Vidal, Frédérique Lesaulnier et avec la contribution juridique de Marion Abecassis.

Pour le programme LORIER : Philippe Ravaud; Ghislaine Filliatreau; Ioana Andreescu

Pour le Conseil Scientifique de l'Inserm : Rodolphe Thiebaut, Isabelle Remy-Jouet, Lotfi Senhadji, Valérie Urback, Murielle Gaudry

Experts extérieurs :

- Anita Burgun, PUPH informatique médicale Institut Imagine, Université Paris Cité, AP-HP & Inserm, Chaire PRAIRIE.
- Jean Charlet, Directeur de recherche AP-HP, LIMICS Sorbonne Université & Inserm U1142.
- Pierre-Antoine Gourraud, PUPH Biologie Cellulaire Nantes Université, Inserm UMR 1064 & Responsable entrepôt de données de santé du CHU de Nantes « Clinique des Données ». Membre de la direction scientifique du Ouest Data Hub.

ANNEXE 2

Rappels sur les différents types d'Intelligence Artificielle

Si les modèles de langage de grande taille (Large Language Models, LLMs) mobilisent aujourd'hui toute l'attention en raison du développement très rapide de leur utilisation, il convient de rappeler qu'ils ne sont qu'une sous-catégorie des réseaux de neurones profonds (Deep Learning), plus spécifiquement des Transformers. Ils méritent une mention spéciale en raison de leur complexité et de leur large utilisation dans le traitement du langage naturel. Ils s'intègrent à un champ plus vaste de modèles[44] :

1. Apprentissage supervisé : Le système reçoit la bonne réponse et il lui est demandé de la retrouver à partir d'autres jeux de données
   - Réseaux de neurones : Utilisés pour des tâches variées comme la reconnaissance d'image et la reconnaissance vocale. Exemples : MLP (Perceptron multicouche), CNN (Réseaux de neurones convolutifs).
   - Machines à vecteurs de support (SVM) : Utilisés pour la classification et la régression.
   - Arbres de décision et forêts aléatoires : Utilisés pour la classification et la régression, en agrégeant plusieurs arbres de décision pour améliorer la précision.

2. Apprentissage non supervisé : L'apprentissage non supervisé utilise des algorithmes d'apprentissage automatique pour analyser et regrouper des jeux de données non étiquetés. Ces algorithmes découvrent des modèles cachés ou des groupements de données sans nécessiter d'intervention humaine.
   - Algorithmes de clustering : Utilisés pour identifier des structures cachées dans les données. Exemples : K-means, DBSCAN.
   - Réseaux de neurones auto-encodeurs : Utilisés pour la réduction de dimensionnalité et la détection d'anomalies.
   - Analyse en composantes principales (PCA) : Utilisée pour la réduction de dimensionnalité et l'identification de tendances dans les données.

[44] Section élaborée à partir de ce que ChatGPT 4.0 répond lorsque la requête (prompt) lui demande (juillet 2024) de nous éclairer sur les principaux modèles d'intelligence artificielle. A noter qu'il omet la section 8 de la présente annexe.

3. Apprentissage par renforcement : à chaque réponse de la machine, il est attribué une valeur « vrai » ou « faux »
    - Q-learning : Un algorithme basé sur des tableaux pour trouver des politiques optimales.
    - Deep Q-Networks (DQN) : Une approche utilisant des réseaux de neurones pour approximativement apprendre les valeurs Q.
    - Algorithmes de politiques : Comme l'algorithme de politique proximal (PPO) et l'algorithme de détermination de politique profonde (DDPG).

4. Apprentissage autosupervisé : Réseaux de neurones profonds (Deep Learning)
    - Réseaux de neurones convolutifs (CNN) : Principalement utilisés pour la reconnaissance d'image.
    - Réseaux de neurones récurrents (RNN) : Utilisés pour le traitement du langage naturel et les séries temporelles. Exemples : LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit).
    - Réseaux antagonistes génératifs (GAN) : Utilisés pour générer de nouvelles données semblables aux données d'entraînement, comme les images et le texte.
    - Transformers : Utilisés pour le traitement du langage naturel et autres tâches. Exemples : BERT, GPT, GPT-3, GPT-4 (Large Language Models, LLMs). Les LLMs sont des modèles de langage de grande taille, entraînés sur des corpus de texte massifs, capables de comprendre et de générer un langage naturel de manière très sophistiquée.

5. Modèles probabilistes
    - Machines de Boltzmann restreintes (RBM) : Utilisées pour la réduction de dimensionnalité, la classification, la régression et la modélisation des données.
    - Modèles graphiques bayésiens : Utilisés pour la modélisation probabiliste des données complexes. Exemples : Réseaux bayésiens, Modèles de Markov cachés.

6. Algorithmes évolutifs
    - Algorithmes génétiques : Utilisés pour l'optimisation et la recherche de solutions dans de grands espaces de recherche.

7. Modèles de langage de grande taille (LLMs)

- Transformers : Les LLMs tels que BERT, GPT-3 et GPT-4 sont basés sur l'architecture Transformer. Ils sont capables de comprendre et de générer du texte humain de manière très avancée. Ces modèles ont révolutionné le traitement du langage naturel (TALN) et sont utilisés dans diverses applications allant de la traduction automatique à la génération de texte.
- Les LLMs se distinguent par leur capacité à traiter et à générer du langage naturel avec une précision et une fluidité sans précédent. Ils sont entraînés sur des ensembles de données extrêmement larges et utilisent des centaines de milliards de paramètres, ce qui leur permet de capturer une vaste gamme de connaissances et de contextes linguistiques.

8. Modèles formels

- Ontologies : ces modèles organisent les connaissances et les données d'un domaine d'activité sous forme de structure hiérarchique. Ils sont, entre autres, la base de systèmes d'aide à la décision dont les raisonnements sont explicites et peuvent être combinés à d'autres modèles (LLM ou autres) pour développer des IA « hybrides ».
- Graphes de connaissances (KG) : ces modèles fondés sur des données et/ou des ontologies permettent de structurer l'information de manière à rendre explicites les relations entre les différentes données, facilitant ainsi leur utilisation dans divers domaines, notamment la recherche d'information, l'intelligence artificielle, et l'analyse de données.

ANNEXE 3

Article L4001-3 du Code de la Santé

Création LOI n°2021-1017 du 2 août 2021 - art. 17

I.- Le professionnel de santé qui décide d'utiliser, pour un acte de prévention, de diagnostic ou de soin, un dispositif médical comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives s'assure que la personne concernée en a été informée et qu'elle est, le cas échéant, avertie de l'interprétation qui en résulte.

II.- Les professionnels de santé concernés sont informés du recours à ce traitement de données. Les données du patient utilisées dans ce traitement et les résultats qui en sont issus leur sont accessibles.

III.- Les concepteurs d'un traitement algorithmique mentionné au I s'assurent de l'explicabilité de son fonctionnement pour les utilisateurs.

IV.- Un arrêté du ministre chargé de la santé établit, après avis de la Haute Autorité de santé et de la Commission nationale de l'informatique et des libertés, la nature des dispositifs médicaux mentionnés au I et leurs modalités d'utilisation.

ANNEXE 4

Un modèle de "langage petit" (SLM) est un modèle de NLP avec un nombre relativement restreint de paramètres par rapport aux LLMs. Ils sont conçus pour être légers, rapides et plus faciles à déployer sur des appareils avec des ressources limitées. Ils peuvent être ajustés et personnalisés pour des tâches spécifiques avec moins de données et moins de puissance de calcul.

https://www.olvani.com/magazine/actualites-digitales/intelligence-artificielle-explication-des-differents-modeles-de-langages-llm-slm--article.

Exemples de SLMs

1. DistilBERT
   - Capacité : Environ 66 millions de paramètres, soit 40% de moins que BERT.
   - Utilisation : Tâches de compréhension du langage avec des ressources limitées.
   - Ressources : Disponible sur Hugging Face.

2. TinyBERT
   - Capacité : Encore plus petit que DistilBERT, optimisé pour les appareils mobiles.
   - Utilisation : Applications mobiles, IoT, où les ressources sont limitées.
   - Ressources : Disponible sur Hugging Face.

3. ALBERT (A Lite BERT)
   - Capacité : Moins de paramètres que BERT, mais avec des techniques d'optimisation pour maintenir la performance.
   - Utilisation : Tâches NLP avec des ressources moindres sans sacrifier la performance.
   - Ressources : Disponible sur Hugging Face.