

MOTS CLÉS

Garantie humaine
 Proof of care
 Intelligence artificielle
 Numérique
 Innovation
 Évaluation

réflexion

MANAGEMENT

Garantie humaine de l'IA et "proof of care" Des méthodologies émergentes complémentaires

Dans notre numéro de juin-juillet étaient présentés les cadres généraux respectifs des méthodologies de *proof of care* (POC) et de garantie humaine de l'intelligence artificielle (IA) ⁽¹⁾. Le présent article apporte un regard croisé sur la complémentarité de ces méthodologies émergentes de management de l'innovation numérique au sein des établissements de santé. Ces démarches en sont désormais parvenues au stade applicatif. Le collègue de garantie humaine peut ainsi être considéré comme un cas d'usage du POC dans le domaine de l'IA en santé. L'évaluation globale des résultats obtenus par les premières expériences pilotes est en cours.

La méthodologie de *proof of care* – POC –, construite par la Chaire de Philosophie à l'hôpital, peut se définir comme un outil de management de projet, dans le sens où il permet aux acteurs en présence de s'approprier l'objet par étapes successives. Dérivé des méthodologies de preuves de concept – *proof of concept* –, le POC constitue une approche globale de l'innovation en santé. Il revendique une démarche résolument expérimentale par laquelle une innovation, dont le besoin d'usage aura été identifié, sera mise au banc d'essai avec les professionnels des établissements de santé et, idéalement, les patients et leurs représentants. Ce faisant, le POC représente une véritable approche de co-construction de l'innovation en santé, s'appuyant tout à la fois sur une dimension participative et sur une logique de capitalisation des résultats dans un objectif de répliquabilité. Le *proof of care* a ainsi vocation à aborder l'ensemble des domaines de l'innovation en santé par une implication des acteurs de terrain, comme dans le cas de l'innovation aux urgences. Il emprunte également

des éléments méthodologiques aux approches de design d'usage ⁽²⁾.

Dans cette optique, la garantie humaine de l'IA peut être présentée comme une déclinaison de cette logique. Ce parallèle est d'autant plus intéressant que la garantie humaine a émergé d'abord de la technique, en partant du besoin identifié d'une nécessité de supervision de l'innovation numérique. Sur le plan des principes éthiques, la garantie humaine de l'intelligence artificielle renvoie, en effet, à la nécessité de ne pas abandonner toute autonomie d'action ou de décision humaine dans un contexte de diffusion de plus en plus rapide de l'IA, en santé et dans d'autres domaines. La terminologie n'a naturellement pas vocation à être figée et ce principe général peut aussi se trouver dans les notions – plus larges – de « supervision humaine ⁽³⁾ » ou – plus resserrées – de « contrôle humain » ou de « décision humaine ». De manière transversale, ce principe recouvre le besoin de ne pas laisser les algorithmes d'intelligence artificielle opérer indépendamment de tout regard humain. Cet intérêt de la supervision humaine se décline également au stade de la conception de ces solutions d'intelligence artificielle.

Marine BACONNET

Doctorante en sciences de l'information et de la communication
 Conservatoire national des arts et métiers

David GRUSON

Membre Chaire Santé Sciences Po Paris
 Co-fondateur Ethik-IA
 Chercheur associé,
 Chaire de Philosophie à l'hôpital

Le principe de garantie humaine de l'IA et du numérique en santé, introduit et porté par Ethik-IA⁽⁴⁾ depuis 2017, puis dans le cadre de la Chaire de Philosophie à l'hôpital a été reconnu dans les avis 129 et 130 du Comité consultatif national d'éthique (CCNE) et dans l'article 11 du projet de loi bioéthique⁽⁵⁾, devenu l'article 17 dans la loi d'août 2021. Cet article 17 de la loi de bioéthique entrée en vigueur en août et le principe de garantie humaine de l'IA qui en découle imposent à tous les utilisateurs et concepteurs de solutions d'IA en santé :

» la mise en œuvre d'une information préalable du patient sur le recours à l'IA dans sa prise en charge ;

» le déploiement d'une supervision humaine de la solution d'IA « en vie réelle » dans des conditions de traçabilité mises en œuvre sous le contrôle et la supervision de la Commission nationale de l'informatique et des libertés (Cnil) et de la Haute Autorité de santé (HAS). Cette supervision peut ainsi, comme le recommande le CCNE, s'exercer dans le cadre de « collèges de garantie humaine ». Nous le verrons, cette méthodologie peut être relue dans un référentiel commun avec la démarche *proof of care*.

Il est à relever que le principe de garantie humaine a reçu, au cours des années 2020-2021, en plus de la révision bioéthique, des concrétisations dans trois autres cadres significatifs :

» la grille d'autoévaluation des dispositifs médicaux intégrant de l'IA préalablement à leur admission au remboursement publié en octobre 2020 par la HAS ;

» les recommandations de l'OMS sur l'éthique et la gouvernance de l'intelligence artificielle en santé du 28 juin 2021 ;

» le livre blanc sur l'IA publié par la Commission européenne le 19 février 2020.

Le principe de garantie humaine de l'IA (*human oversight*) est également introduit à l'article 14 du projet de règlement sur l'intelligence artificielle de la Commission européenne diffusé le 21 avril 2021. On retrouve dans cet article 14 du projet de règlement européen les deux axes essentiels de l'article 17 et des méthodologies construites par Ethik-IA depuis 2017 dans le champ de la santé⁽⁶⁾ :

» l'information des utilisateurs de la solution d'IA ;

» la supervision humaine de l'IA dans sa phase de conception et, dans une logique d'amélioration continue de la qualité, dans son utilisation « en vie réelle ».

Dans la reconnaissance de la garantie humaine de l'IA, nous retrouvons donc l'idée de participation des professionnels et des représentants des patients à la construction de l'innovation et à sa supervision sur le terrain.

Deux différences notables entre les deux approches sont cependant à relever :

» la garantie humaine part de la diffusion de l'intelligence artificielle elle-même et reste en quelque sorte « branchée » sur cette innovation. Elle recevra une portée applicative pour tous les cas d'usage de l'IA – y compris hors santé dans le cadre du règlement européen – mais restera arrimée à cette technologie alors que la méthodologie *proof of care*, centrée, elle, sur la santé, pourra appréhender d'autres formes d'innovation dans ce domaine sectoriel ;

» la reconnaissance du principe de garantie humaine dans le cadre de la loi de bioéthique et du projet de règlement européen sur l'IA lui confère une portée opposable dont ne dispose pas *a priori* l'approche *proof of care*. L'absence de garantie humaine de l'IA pourra donc être reconnue comme irrégulière au plan juridique voire, dans le cadre européen, assortie de sanctions financières lourdes, équivalentes à celles du Règlement général sur la protection des données (RGPD).

Les collèges de garantie humaine de l'IA : une méthodologie de *proof of care* opposable ?

La garantie humaine de l'IA peut ainsi être relue comme un cas juridiquement opposable de *proof of care* par lequel professionnels de santé et patients vont être embarqués dans la co-construction de la supervision de l'intelligence artificielle. Si la garantie humaine de l'IA est donc, par principe, opposable au plan du droit, on retrouve, dans son déploiement méthodologique, la plasticité fonctionnelle qu'emprunte plus globalement le *proof of care*.

Recommandés par l'avis 129 du CCNE, les « collèges de garantie humaine » peuvent associer experts du numérique, médecins, professionnels paramédicaux et représentants des usagers dans le but d'assurer la supervision humaine d'un algorithme. Si ces collèges ne sont donc que recommandés et non exigés, leur existence et leur mise en fonctionnement permettront de vérifier les conditions d'une garantie humaine effective, traçable et proportionnée qui découlent des cadres juridiques français et européen.

Leur vocation est en effet d'assurer *in situ* une supervision de l'efficacité et du caractère éthique des dispositifs algorithmiques, à la fois au stade de leur conception et de leur application. Les dossiers à auditer pourraient être définis à partir d'événements indésirables constatés, de critères prédéterminés et/ou d'une sélection aléatoire.

Cette notion de plasticité et de pilotage qualité dans un cadre juridique opposable se retrouve assez clairement dans les orientations du guide méthodologique récemment publié

NOTES

(1) C. Jobin, « Preuves de concept et capacités de conception collective », *Gestions hospitalières*, juin/juillet 2022, n°617, p. 335-340.

D. Gruson, « Le principe de garantie humaine de l'IA en santé », *Gestions hospitalières*, n°617, juin/juillet 2022, p. 383-385.

(2) C. Jobin, A. Fenoglio, C. Fleury-Perkins, S. Hooge, P. Le Masson, « Prendre soin des urgences hospitalières : un exemple de *proof of care* », 88^e congrès de l'Acfas - Colloque intersections du design 2021 : prendre soin par le design.

(3) La Commission européenne utilise l'anglicisme *human oversight* pour recouvrir cette approche large.

(4) Ethik IA se veut une initiative citoyenne et académique ayant pour but de « proposer une série d'outils et de notes de cadrage pour garantir un regard humain sur les algorithmes en santé ».

(5) Le principe de garantie humaine figure en tant que tel dans l'exposé des motifs et l'étude d'impact du projet de loi. L'article 11 a été adopté par l'Assemblée nationale et le Sénat dans des termes rédactionnels différents mais reprenant les mêmes principes : information préalable du patient sur le recours au numérique dans le cadre du recueil – obligatoire – de son consentement et mise en œuvre d'une supervision humaine du numérique et de l'IA en santé.

(6) Voir notamment sur l'origine de la garantie humaine de l'IA : <https://www.ticsante.com/story?ID=3995> et D. Gruson, *La Machine, le Médecin et Moi*, éditions de l'Observatoire, décembre 2018. Les avancées de la reconnaissance de la garantie humaine de l'IA ont également donné lieu à une communication récapitulative d'Hélène Marin et Marine Baconnet lors du séminaire doctoral de la Chaire de Philosophie à l'hôpital de juin 2022.

par la délégation numérique du ministère de la Santé et qui endosse la méthodologie des collèges de garantie humaine⁽⁷⁾. En pratique, le collège de garantie humaine devra s'assurer, à partir des cas revus, que l'IA reste efficace médicalement et responsable éthiquement au fil de sa transformation de *machine learning*. En cas d'écart non souhaitable, des mesures correctrices seraient proposées aux utilisateurs ou aux concepteurs d'IA.

Le projet pilote de collège de garantie humaine sous l'égide de l'UFSBD

L'Union française pour la santé bucco-dentaire (UFSBD) a confié à Ethik-IA la réalisation d'un audit juridique et éthique du fonctionnement de la solution Oralien (solution d'IA de reconnaissance d'image mise en œuvre par la start-up française Dental Monitoring dans le domaine dentaire, s'appliquant à 48 Ehpad dans le cadre d'un protocole de financements innovants prévu par l'article 51 de la LFSS), afin de s'assurer de la conformité de celle-ci avec les règles en matière de protection des données à caractère personnel et tenant compte des obligations spécifiques qui s'imposent en matière de données dites sensibles concernant la santé.

Cette analyse a été opérée au regard de la législation française et européenne actuelle et des cinq clés de régulation émises par Ethik-IA, ainsi que du texte de l'article 17 de la loi bioéthique. La mise en œuvre de cet audit, et la proposition du système de garantie humaine qui en est issue, ressemble donc à une approche de design d'usage au sens de la méthodologie *proof of care*. L'objectif a en effet consisté, à partir d'une analyse technique et d'une itération avec les acteurs, à co-construire le dispositif de garantie humaine à appliquer en le proportionnant, d'une part, au ratio bénéfices/risques de la solution d'IA et, d'autre part, à la complexité et notamment au caractère auto-apprenant du système algorithmique embarqué. La mise en œuvre du collège de garantie humaine associé à l'expérimentation UFSBD constitue le premier cas effectivement mis en œuvre d'un principe désormais entré dans la législation française et ayant fait l'objet d'une reconnaissance au plan international.

Afin d'être en adéquation avec le futur cadre législatif, le dispositif met en œuvre une supervision humaine autonome de Dental Monitoring prenant la forme d'un collège de garantie humaine. Ayant pour rôle de formuler un deuxième avis humain sur une sélection de dossiers aléatoires, il est positionné auprès de l'UFSBD et se réunit sous son égide.

En pratique, ce collège mobilise une équipe de quatre experts réviseurs chirurgiens-dentistes, indépendants, de Dental Monitoring choisis par l'UFSBD. Il est complété par des représentants des établissements concernés par l'expérimentation, des représentants des patients ainsi que Dental Monitoring (qui avait manifesté à l'occasion des phases préparatoires son accord sur l'engagement d'un tel dispositif).

Le collège se réunit tous les trimestres ; à chaque rencontre, il étudie 30 dossiers identifiés soit de manière aléatoire, soit à partir d'événements indésirables remontés par les acteurs de terrain de l'expérimentation. Pour chacun de ces dossiers, une grille d'analyse élaborée par Ethik-IA en collaboration

avec l'UFSBD devra être renseignée. Elle a pour finalité d'objectiver l'adéquation de la proposition faite par l'IA à celle qui serait faite par le collège.

Ce collège peut être sollicité en sus à la demande d'un établissement, d'un professionnel, d'un patient ou de son entourage pour avoir un deuxième avis. Cet avis sera alors susceptible d'être reconnu comme une téléexpertise de garantie humaine de l'IA, valorisable économiquement au sens de l'avenant n°6 de la convention médicale de 2018. La pratique de la garantie humaine de l'IA a conduit à distinguer des actions correctives de plusieurs catégories :

- » actions portant sur le dispositif algorithmique lui-même ;
- » actions portant sur les pratiques humaines ;
- » actions portant sur le contexte organisationnel d'usage du dispositif algorithmique.

Cette expérimentation pilote est menée en articulation étroite avec la HAS et le Syndicat national de l'industrie de technologies médicales (Snitem) dans une optique d'extrapolation de cette méthodologie de vérification de la conformité garantie humaine pour tous les utilisateurs d'IA. Un processus de normalisation a également été engagé avec l'Afnor, l'association experte en documentation normative et en certification ; une démarche de labellisation est, par ailleurs, en cours avec la plateforme Digital Medical Hub de l'AP-HP.

Perspectives de déploiement des collèges de garantie humaine

Depuis le projet UFSBD, Ethik-IA a développé une dizaine de pilotes de méthodologie de garantie humaine de l'IA en santé dans des domaines aussi divers que les essais cliniques *in silico* avec le secteur pharmaceutique, l'IA en génomique, l'IA de pharmacovigilance ou la détection de pathologies par

NOTE

(7) https://esante.gouv.fr/sites/default/files/media_entity/documents/ethic_by_design_guide_vf.pdf

Recommandations de bonnes pratiques pour intégrer l'éthique dès le développement des solutions d'intelligence artificielle en santé : mise en œuvre de « l'éthique by design ». Présentation des travaux du GT3 de la cellule éthique du numérique en santé de la Délégation ministérielle au numérique en santé, avril 2022.

traitements massifs de données dans des programmes de responsabilité populationnelle en santé.

L'intégration d'Ethik-IA au sein de ParisSanté Campus permettra de diffuser la solution auprès des nombreux acteurs de l'IA présents sur site et ainsi de positionner le premier hub international de déploiement de la conformité en garantie humaine de l'IA.

La démarche est, par ailleurs, en phase d'extrapolation à d'autres domaines d'application de l'IA (éducation, logement, legaltech, communications électroniques et postale).

Un exemple de démarche pilote à l'échelle d'un établissement : le GH Saint-Joseph

Une démarche pilote de structuration de la garantie humaine à l'échelle d'un établissement de santé a été initiée pour le groupe hospitalier Saint-Joseph, à Paris. Plusieurs cas d'usage entrent dans le périmètre de cette démarche, dont l'IA en imagerie. La radiologie constitue, en effet, un domaine particulièrement propice à l'utilisation et au développement de l'IA en santé : les fractures non repérées sur les radiographies sont les causes les plus fréquentes de divergences diagnostiques entre les interprétations par des non-radiologues et par des résidents en radiologie⁽⁸⁾.

Dans le prolongement de cette idée, Ethik-IA développe actuellement avec

le GH Paris Saint-Joseph une démarche pilote de garantie humaine, ayant pour objectif de mettre en place une supervision humaine des systèmes d'IA utilisés au sein d'un établissement dans une logique d'amélioration continue de la qualité. Trois solutions ont été sélectionnées dans le cadre de cette démarche afin d'analyser un dispositif d'IA en radiologie permettant de donner une appréciation sur la présence de fractures et une localisation de celles-ci. L'objectif est de structurer les modalités de fonctionnement d'un collège de garantie humaine à l'échelle de l'établissement, intégrant la supervision de ce dispositif en radiologie ainsi que de deux autres cas d'usage.

Nous retrouvons donc ici les éléments caractéristiques de la démarche *proof of care* :

- » une co-construction *ab initio* avec les acteurs de terrain ;
- » la mise en jeu d'une approche de design d'usage pour déterminer le système de garantie humaine à appliquer pour chaque solution d'IA ;
- » une boucle de rétroaction, en vie réelle, pour réajuster la démarche en fonction des résultats constatés.

Conclusion

Garantie humaine de l'IA et *proof of care* empruntent nombre de caractéristiques communes comme les méthodologies d'incubation et de co-construction de l'innovation en santé avec les professionnels et les représentants des patients. La principale distinction porte sur l'opposabilité juridique acquise par la garantie humaine, condition *sine qua non* devant désormais être validée par tous les systèmes d'IA en santé et dans les autres secteurs d'application. De ce point de vue, l'évaluation globale des résultats obtenus par ces collèges de garantie humaine, en phase de diffusion rapide dans les établissements de santé, pourra aussi s'effectuer à l'aune de la capacité remarquable de l'innovation en santé de porter ses approches sur une échelle plus large afin d'irriguer globalement l'émergence d'une véritable culture du design de l'innovation. ●

NOTE

(8) A. Guermazi, C. Tannoury, A. J. Kompel, A. M. Murakami, A. Ducarouge, A. Gillibert, X. Li *et al.*, "Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence", *Radiology*, 2022, 302(3), 627-636.

» ZOOM La Chaire de Philosophie à l'hôpital

Dirigée par la philosophe et psychanalyste Cynthia Fleury, cette chaire hospitalo-académique est liée au Conservatoire national des arts et métiers (Cnam) et au GHU Paris Psychiatrie & Neurosciences. À travers un dispositif recherche et enseignement, de formation et diplomation, d'expérimentation et déploiement, cette chaire aspire à inventer la fonction soignante en partage et l'alliance efficiente des humanités et de la santé. Ses thématiques de recherche s'articulent autour de cinq pôles : Philosophie clinique et savoirs expérientiels/Santé connectée et intelligence artificielle/Design capacitaire/Résilience et clinique du développement/Nature et patrimoine en santé. La chaire abrite par ailleurs un

espace doctoral composé de douze doctorants.

Les prochains articles publiés par *Gestions hospitalières* porteront, entre autres sujets, sur les facteurs humains au bloc opératoire pédiatrique, le développement des sciences humaines et sociales (SHS) dans la formation des soignants, les arts visuels à l'hôpital comme soin, les humanités au service de l'attractivité de la psychiatrie en contexte de désertification médicale, l'amélioration de la communication médecin/malade en préopératoire par le biais d'interfaces digitales..., toujours sous un angle philosophique, anthropologique et sociologique, avec une approche théorique couplée à un retour d'expérience. www.chaire-philosophie.fr

