

SciencesPo

CHAIRE SANTÉ

**COMPTE-RENDU
SEANCE NUMERO 8
LES ENJEUX ETHIQUES
ET LA GARANTIE
HUMAINE DE L'IA
FEVRIER 2023**

Séance 8 : Les enjeux éthiques et la garantie humaine de l'IA

Intervenants :

- *Animation* : **Daniel Benamouzig**, titulaire de la Chaire santé, directeur de recherche au CNRS/CSO
- **Pierre Simon**, médecin, président fondateur et ancien président de la Société française de Télémedecine
- **Nicolas Desrumaux**, responsable juridique et délégué à la Protection des Données du Digital Medical Hub (DMH) à l'AP-HP, Assistance Publique - Hôpitaux de Paris
- **Hélène Marin**, directrice de projet Garantie Humaine Ethik-IA : un cas concret de méthodologie de garantie humaine mise en place pour la solution "Oralien" de Dental Monitoring
- **Catherine Tessier**, chercheuse et référente intégrité scientifique et éthique de la recherche à l'ONERA - membre du Comité national pilote d'éthique du numérique

Présentation de la séance

L'IA introduit de nouveaux principes d'innovation en santé et met en débat des principes acquis, comme ceux du consentement des personnes. Dans un univers où les enjeux économiques et souverains risquent de perturber des principes éthiques protecteurs pour les personnes (principes d'autonomie, de justice ou de non-discrimination par exemple, etc.), d'autres principes doivent-ils être introduits ? (comme ceux de loyauté ou de réflexivité introduits par la CNIL par exemple). Quelles formes d'organisation collective sont susceptibles de garantir la mise en œuvre et le respect de ces principes, et plus largement d'assurer une forme de transparence et de confiance sur des objets dont les dimensions techniques sont à la fois complexes et très évolutives ? La Chaire santé de Sciences Po a pris une part active aux travaux qui ont conduit à la formulation du principe de garantie humaine et à sa reconnaissance. Il vise une information préalable du patient sur le recours au numérique dans le cadre d'un recueil obligatoire de son consentement, ainsi que la mise en œuvre d'une supervision humaine du numérique et de l'IA en santé. Le principe de Garantie Humaine de l'IA et du numérique en santé a été reconnu dans les avis 129 et 130 du CCNE et dans l'article 11 du projet de loi bioéthique. L'idée est d'appliquer les principes de régulation du numérique et de l'intelligence artificielle en amont et en aval de l'algorithme, à des points critiques identifiés en vue d'un dialogue partagé entre les professionnels, les patients et les concepteurs d'innovation, plutôt qu'à chaque étape, ce qui aurait pour effet de bloquer l'innovation. L'objectif consiste à s'assurer « au fil de l'eau » que l'application ou l'algorithme reste sur un développement opérationnel à la fois efficace médicalement et responsable éthiquement. Quels peuvent être les avantages et les limites du principe de garantie humaine ? Comment peut-il être mis en œuvre de manière opérationnelle, notamment à travers un « collège de garantie humaine » ? Quelles sont les possibilités d'extension de ce modèle, peut-il être étendu au-delà de la France (Commission Européenne, OMS) ?

1. Pierre Simon, médecin, président fondateur et ancien président de la Société française de Télémedecine

La télésanté désigne les pratiques professionnelles qui utilisent des services numériques de la e-santé. Comment peut-on articuler éthique médicale et éthique du numérique dans ces nouveaux outils ?

On distingue 4 grands principes éthiques :

- La bienfaisance
- La non-malfaisance
- La justice
- L'autonomie

1.1 La bienfaisance en télésanté clinique

Pour l'éthique médicale

- La pratique de la télésanté prend en compte les attentes d'une personne, elle nécessite l'information et le recueil du consentement
- En cas d'illectronisme (majoritaire chez les plus âgés), la personne doit être assistée. Des réponses sont déjà apportées, par exemple l'assistance possible d'une infirmière ou d'un pharmacien lors d'une téléconsultation, mais elles ne sont pas toujours connues des patients et des professionnels de santé
- La pratique de la télésanté doit être pertinente (décret du 3 juin 2021)

Pour l'éthique du numérique

- Pour les pratiques professionnelles de télésanté : la solution technologique doit être la plus bénéfique possible pour l'utilisateur, par exemple des solutions agiles, ergonomiques et sécurisées
- Pour les données de santé : il faut mettre en place un stockage des données et un usage de ces données qui aient un rapport coût-bénéfice positif pour l'utilisateur, par exemple le nouveau dossier médical partagé (DMP) installé dans Mon Espace Santé est un service numérique gratuit pour le citoyen avec un bénéfice indiscutable pour rendre le patient acteur de sa propre santé.
- Pour les algorithmes : l'algorithme doit déboucher sur une action bénéfique, utile, et délivrer une information claire et compréhensible. C'est aujourd'hui la faiblesse des algorithmes qui ne peuvent pas toujours être expliqués aux usagers, alors que l'information donnée doit être claire et compréhensible. C'est la raison pour laquelle, l'interprétation d'un résultat d'algorithme doit être garantie par le professionnel de santé. Pour l'instant, l'algorithme ne peut assurer aucune responsabilité. Cette dernière est intermédiée par le médecin (principe de garantie humaine)

1.2 La non-malfaisance en télésanté clinique

Pour l'éthique médicale

- L'information du patient décrit des bénéfices supérieurs aux risques. Un acte de télésanté qui aurait plus de risques que de bénéfices violerait le principe de non-malfaisance.
- La personne avec un déclin cognitif doit être respectée. Par exemple, pour la téléconsultation dans les EPHAD, il est nécessaire de mener la démarche d'information et de recueil de consentement auprès de la famille ou de la personne de confiance du patient.

- La téléconsultation ou le télésoin complète la consultation ou le soin en présentiel. L'avenant 9 de la convention médicale de septembre 2021 encadre les pratiques à distance : elles ne doivent pas dépasser 20% de l'activité d'un médecin ou d'un professionnel de santé non médical (la moyenne actuelle est autour de 3%)
- L'environnement d'une téléconsultation ou d'un télésoin ne doit pas créer une relation humaine dégradée. Dans la relation avec le patient, il est par exemple important de prendre en compte la communication non-verbale (pas de téléconsultation sans l'image du patient, c'est-à-dire la nécessité d'utiliser la Visio)

Pour l'éthique du numérique

- Pour les pratiques professionnelles : il ne faut pas fournir ou recommander des outils rendant la pratique dégradée. Par exemple, pendant la pandémie, de nombreux médecins ont vécu la téléconsultation par téléphone comme une forme dégradée de la consultation médicale.
- Pour les données de santé : leur stockage et leur usage ne doivent pas nuire à l'utilisateur. Par exemple, l'usage de solutions grand public des GAFAM (WhatsApp, Skype, Messenger, etc.) expose à une violation de la confidentialité des données personnelles de santé.
- Pour les algorithmes : l'algorithme doit éviter à l'utilisateur de souffrir et d'avoir un préjudice. C'est la raison pour laquelle la garantie humaine de l'interprétation d'un algorithme est devenue une obligation légale dans la loi Bioéthique.

1.3 Le principe d'autonomie en télésanté clinique

Pour l'éthique médicale

- Il est obligatoire d'informer un patient et de recueillir son consentement. C'est une obligation légale inscrite au Code de la santé publique depuis la loi Kouchner du 4 mars 2002.
- Cette information doit présenter à l'utilisateur les bénéfices et les risques d'un acte de télésanté

Pour l'éthique du numérique

- Pour les pratiques professionnelles : des solutions numériques doivent préserver la capacité du patient à penser, décider et agir librement (autonomie de jugement). Ainsi toute solution numérique (en particulier les applis) qui altérerait l'autonomie du patient à décider violerait ce principe éthique d'autonomie.
- Pour les données de santé : l'utilisateur est le principal décideur vis-à-vis de ses données de santé. C'est ce qui a été inscrit dans le décret du 3 mars 2021 créant l'Espace numérique en santé.
- Pour les algorithmes : l'algorithme doit permettre d'accroître le contrôle de l'utilisateur sur sa vie et son environnement. L'algorithme ne doit pas être un moyen d'asservir un utilisateur.

1.4 Le principe de justice en télésanté clinique

Pour l'éthique médicale

- La télésanté doit reposer sur la solidarité des citoyens et l'équité dans l'accès aux soins. Il faut en particulier veiller à la meilleure répartition possible des moyens humains, technologiques et financiers. C'est la mission donnée aux gouvernants par les citoyens.
- Il faut considérer l'illectronisme dans les soins à distance. Le développement de solutions numériques et la réduction progressive des échanges humains en présentiel ont fait émerger cet illectronisme qui est l'impossibilité d'une personne de naviguer facilement sur le Web. C'est particulièrement vrai chez les personnes âgées de 70 ans et plus où 70% sont touchés d'illectronisme.

- La télésanté doit contribuer à améliorer l'accès aux soins dans les déserts médicaux, à la condition que ces déserts médicaux ne soient pas associés à des déserts numériques.

Pour l'éthique du numérique

- Pour les pratiques professionnelles : il faut garantir l'égalité d'accès à l'offre numérique pour que chaque citoyen puisse bénéficier des solutions numériques. C'est en particulier le rôle des opérateurs privés du numérique à qui l'Etat demande d'assurer le maillage complet du territoire en bandes passantes.
- Pour les données de santé : leur stockage et leur usage doivent permettre une amélioration soutenable des soins. Les données de santé ne doivent être stockées que pour une finalité d'amélioration de la vie des citoyens, comme le rappelle le RGPD.
- Pour les algorithmes : il faut utiliser un algorithme qui ne génère aucun biais, aucune discrimination, aucune forme d'exclusion. Un algorithme doit être construit à partir de données représentatives d'une population.

Questions :

N'y a-t-il pas parfois une tension entre les libertés fondamentales et le suivi individuel des patients ?

Les connaissances médicales permettent une amélioration de la qualité de vie des individus et, si l'information est bien faite, cela crée de l'adhésion aux dispositifs médicaux utilisés pour la surveillance à distance des patients atteints de maladies chroniques. L'information doit montrer les bénéfices et les risques. Si l'on prend le cas des expérimentations sur les dialyses à distance, qui nécessitent une surveillance attentive, celle-ci est acceptée car le dispositif améliore la qualité de vie du patient.

2. Nicolas Desrumaux, responsable juridique et délégué à la Protection des Données du Digital Medical Hub (DMH) à l'AP-HP, Assistance Publique - Hôpitaux de Paris

La « garantie humaine », évoquée dans le projet de règlement sur l'intelligence artificielle de la Commission européenne, suppose une information préalable du patient sur le recours au numérique ainsi que la mise en œuvre d'une supervision humaine du numérique et de l'IA en santé. Dans cette optique, la bride réglementaire et normative semble assez lâche.

Pour définir l'IA, le projet de règlement renvoie à un « système d'intelligence artificielle ». La Commission européenne le définit comme un logiciel développé à partir d'une ou plusieurs techniques listées en annexe et distingue 3 catégories de systèmes d'IA :

- Les systèmes autoapprenants : le machine-learning et autres systèmes qui ont la capacité de se reconfigurer eux-mêmes
- Les systèmes logiques : auparavant appelés « systèmes experts », ils sont déterministes et suivent une trame logique
- Les systèmes statistiques, notamment les estimations bayésiennes

La définition donnée à l'intelligence artificielle est donc assez large. 3 niveaux de risques sont distingués :

- Certains types d'IA sont interdits par le règlement
- Certaines pratiques estimées à haut risque sont strictement encadrées par des principes de conformité qui doivent être suivis par les créateurs de l'IA
- Les IA à faibles risques pour les personnes sont enfin simplement soumises à des principes de transparence

La majorité des articles de ce projet concernent les IA à haut risque, dont les applications en santé font partie. Il s'agit donc d'une approche par les risques et non pas d'une réglementation sectorielle. Les objectifs poursuivis par ce projet sont la sûreté, la sécurité juridique pour faciliter investissements et l'innovation dans le domaine, ou encore la lutte contre les biais, bien que ce dernier point soit finalement peu détaillé dans le règlement.

Ce règlement ne se suffit pas à lui-même : une fois adopté, il s'insère dans un tissu réglementaire, comme le RGPD, le DSA (Digital Services Act), mais aussi les législations nationales (comme la loi dite « Informatique et libertés »). On note également des interactions juridiques avec d'autres textes, comme la Charte des droits fondamentaux de l'Union européenne.

Dans ce règlement, une part d'auto-régulation est laissée aux fournisseurs d'IA, par exemple pour ce qui concerne la responsabilité humaine algorithmique (article 14 du projet du règlement).

Nous centrerons ici notre analyse autour de deux axes :

- Loin d'être un règlement homogène et cohérent, ce projet se présente sous la forme d'un patchwork réglementaire complexe et problématique sous certains aspects
- On analysera les controverses sur l'application de ce règlement qui émergent avant même son adoption

2.1 Le projet de réglementation : un patchwork normatif ?

Dès 2018, les institutions européennes s'emparent du sujet de l'IA, à la recherche de consensus entre les instances européennes, les États et les industriels. Mais ce consensus n'est aujourd'hui toujours pas atteint et le règlement comporte des étrangetés et imprécisions.

Quel est le niveau de règle le plus pertinent ?

Ces règles sont supposées informer les comportements, et posent des principes généraux qui laissent une souplesse dans le choix des moyens, et certains principes ne sont pas suffisamment explicités.

On note une prise en compte des différentes facettes de l'éthique mais pas véritablement de cadrage : il manque un consensus sur le contenu de cette éthique de l'IA, et l'incantation à l'éthique ne suffit pas à donner des règles précises.

Certains principes doivent être questionnés, comme celui de « loyauté » qui figure à l'article 6 de la loi « Informatique et libertés » : si ce terme peut être compris pour un humain, il est plus difficilement transposable pour une IA. Cela amène à une vigilance sur les finalités qui ne sont pas explicitées au moment de la collecte des données.

Quelles vont-être les règles applicables en pratique ? Pour les IA à haut risque, deux conditions sont nécessaires :

- Une composante sécurité mentionnée à l'annexe 2 du règlement
- La protection des consommateurs

Le règlement se contente de dire en filigrane qu'il faut traiter l'IA en santé comme un dispositif médical, ce qui renvoie à tout un corpus de règles. Des progrès d'écriture sont nécessaires.

On note des cas d'orthogonalité, de conflits de règles très nets qui ne sont pas résolus par les débats entre le Parlement européen, le Conseil des ministres de l'UE et la Commission européenne. Par exemple, l'article 22 du RGPD stipule le droit de ne pas faire l'objet d'une décision automatisée, alors que l'IA aide à la décision médicale.

2.2 De nombreuses contestations sur l'application de ce futur règlement

Ce projet est présenté comme le point d'acmé du respect des règles éthiques et de leur application aux IA, avec un discours promotionnel autour de ce règlement qui le rend très attendu par les professionnels de santé, les professionnels juridiques mais aussi par les industriels.

Il a le mérite de donner une définition de ce qu'on appelle « IA », un logiciel développé au moyen d'une ou des techniques citées précédemment et qui peut « pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit ». L'écriture reste ici maladroite, floue, et on remarque un manque d'emprise sur le réel : l'IA n'est pas un simple logiciel, c'est un programme susceptible d'auto-reconfiguration. La définition retenue ne prend pas en compte toute l'évolutivité de l'IA.

Le règlement espère que les opérateurs et partenaires industriels vont s'emparer de son contenu de manière proactive en mettant en place par eux-mêmes un suivi et un accompagnement : la démarche reste ici très incantatoire.

On peut également redouter que ce règlement amène à accorder une confiance trop précipitée dans les IA : il manque une vision d'ensemble, ce qui laisse craindre un avancement en ordre dispersé et des décisions disparates.

Les IA à haut risque devront avoir un marquage CE : ce point s'applique aussi aux acteurs extra-européens, et le marquage devra être renouvelé en cas de changement substantiel de l'algorithme, ce qui pourrait mener à un engorgement du côté de la certification.

Questions :

Comment expliquer le décalage entre ce texte juridique et la réalité ?

Il y a ici une volonté de produire un règlement sur toutes les IA : ce périmètre est peut-être trop large et porte sur un objet dont on ne connaît pas la finitude, ce qui rend impossible un texte simple applicable à tout le périmètre UE. On remarque également une volonté d'avancer rapidement sur ce sujet, avec une conscience du caractère imparfait de ce projet. Il y a malgré tout besoin de légiférer, et ce règlement a le mérite de donner des définitions et les bases d'un cadre.

3. Hélène Marin, directrice de projet Garantie Humaine Ethik-IA : un cas concret de méthodologie de garantie humaine mise en place pour la solution "Oralien" de Dental Monitoring

Présentation de la solution « ORALIEN » :

- Il s'agit d'une solution de télésurveillance déployée au sein d'EHPAD, pour des patients qui avaient difficilement accès à la santé bucco-dentaire
- La solution est utilisée par le personnel soignant des EPHAD
- Des écarteurs sont mis à disposition de chaque patient et le dispositif réalise une vidéo de la bouche du patient. Cette vidéo est transposée en panoramique dentaire analysée par de l'IA, qui émet des recommandations
- Les recommandations pour chaque patient sont accessibles au personnel soignant via un tableau de bord sur le site sécurisé « Oralien »
- Deux suivis annuels via télésurveillance sont prévus pour chaque résident

3.1 Les principes généraux du collège de garantie humaine mis en place

- Ethik-IA et Vatieur ont réalisé un audit juridique et éthique initial pour comprendre la solution et proposer un modèle adapté à l'algorithme.
- Le collège se réunit 3 fois par an, avec 30 dossiers maximum révisés par session, et émet des propositions d'actions correctives

- Un bilan du fonctionnement du dispositif garantie humaine est réalisé chaque année

3.2 Le fonctionnement du collège

- 30 dossiers déjà traités par l'IA sont sélectionnés aléatoirement, auxquels sont ajoutés les éventuels dossiers qui seraient remontés à la suite d'éléments indésirables (le cas ne s'est pas encore présenté)
- Un export de ces dossiers « virginisés » (sans trace d'intervention de l'IA pour ne pas influencer la décision) est réalisé dans un environnement sécurisé HDS
- 4 chirurgiens-dentistes réviseurs mandatés par l'UFSBD et sans lien d'intérêt avec la société analysent les dossiers avec les mêmes données que l'IA
- Le processus inclut la validation d'une grille de report des constats et implique des représentants des usagers

3.3 La méthodologie initiale suivie par le collège

- Pour chaque dossier traité, sont comparés les éléments fournis par l'IA et ceux fournis par les deux réviseurs
- Au départ, l'analyse des écarts entre l'IA et les réviseurs était exclusivement centrée sur l'item « recommandation de la consultation d'un dentiste » : il s'agissait d'éviter toute perte de chance pour le patient
- Le collège analysait en intégralité 2 dossiers ayant des écarts entre IA et réviseurs sur cet item

3.4 Les évolutions de la méthodologie

- L'item « hygiène » a été ajouté au processus de vérification
- Tous les dossiers ayant des écarts entre l'IA et les réviseurs sur ces items sont analysés dans leur intégralité

3.5 Les actions correctives et le pilotage continu par la qualité

- Le premier collège avait permis de mettre en discussion la prise en compte de l'âge dans les décisions de l'IA
- Le deuxième collège a recommandé de compléter la formulation de l'instruction « Vérifier l'adaptation de la prothèse » en y ajoutant la mention « avec un chirurgien-dentiste » afin de gagner en clarté
- Lors du troisième collège de garantie humaine, une reprise visco-gel provisoire sur une prothèse n'a pas déclenché une consultation dentaire pour l'IA car ce cas n'était pas connu par la solution
- La méthodologie suivie est en cours de standardisation et de labellisation

3.6 Le bilan annuel du fonctionnement du collège de garantie humaine sous l'égide de l'UFSBD

- Le bilan annuel détaille l'évolution de la méthodologie des collèges, les actions correctives globales, les avis sur les process, les actions et le suivi de l'algorithme, et les avis sur le modèle de garantie humaine en cours

À la suite du bilan annuel, des propositions ont été formulées :

- Passer de 3 à 2 collèges par an en augmentant la durée de chaque collège
- Analyser une fois par le critère « hygiène » plutôt qu'à chaque collège : on constate en effet peu d'écart entre l'IA et les chirurgiens-dentistes
- Ajouter de nouveaux critères déterminés lors d'un point méthodologique avec l'UFSBD (caries et gencives)

4. Catherine Tessier, chercheuse et référente intégrité scientifique et éthique de la recherche à l'ONERA - membre du Comité national pilote d'éthique du numérique

L'IA est un domaine scientifique qui fait souvent l'objet de personnifications et d'anthropomorphisme. On parle ainsi d'une IA « digne de confiance », de « prise de décision », d'« apprentissage », ce qui crée des fantasmes, des attentes, des craintes, des confusions et des erreurs. Il faut utiliser un vocabulaire adéquat, qui ne personnifie pas, qui ne projette pas d'aspects moraux sur les machines, et qui ne met pas sur le même plan analyse humaine et calculs numériques.

Dans les textes internationaux, l'« intelligence artificielle » est en général bien définie en préambule. Mais souvent, dans le corps de ces textes, l'IA désigne exclusivement l'apprentissage machine, sans que cela soit explicitement dit et sans précisions au sujet des techniques qui sont employées. Certains chercheurs, comme Luc Steels, pensent qu'il faudrait cesser de parler d'IA et plutôt expliciter les techniques employées, les finalités des systèmes, et les contextes et propriétés d'utilisation.

Tensions et paradoxes de l'éthique de l'IA

Les textes internationaux évoquent très peu le fait que les principes et exigences ne peuvent être simultanément satisfaits, alors que c'est justement un des fondements de la réflexion éthique. Par exemple :

- La transparence, l'explicabilité, la prédictibilité vs la sécurité
- La précision vs la protection de la vie privée
- La précision vs la préservation de l'environnement

On note par ailleurs un certain nombre de paradoxes : les « systèmes d'intelligence artificielle » doivent respecter des principes et en même temps ils constituent une menace pour ces mêmes principes, ou bien ils sont conçus pour satisfaire un objectif qu'ils contribuent à menacer. Par exemple, certains systèmes visent à réduire les inégalités mais peuvent également les exacerber (cas de l'illectronisme).

L'exemple du « contrôle humain »

L'avis 129 du CCNE (Comité consultatif national d'éthique pour les sciences de la vie et de la santé) prévoit une « *garantie humaine du numérique en santé, c'est-à-dire la garantie d'une supervision humaine de toute utilisation du numérique en santé* ».

Qu'entend-on par « garantir » ? S'agit-il de certifier, d'attester, de démontrer, d'assurer ? Comment garantir ? Quel type de supervision est attendue ? À qui garantit-on ? Au patient, au médecin ? Le terme même de « garantie » est ambigu, et en aucun cas on ne peut garantir que l'humain ou l'ensemble machine-humain fournira un résultat pertinent.

Un consensus international sur le principe de « contrôle humain » des « systèmes d'intelligence artificielle »

Si l'on reprend l'article 14 du projet de réglementation européen qui décrit le « contrôle humain », certains éléments semblent difficiles à rendre opérationnels :

- « *Appréhender totalement les capacités et les limites du système d'IA* » ;
- « *Avoir conscience d'une éventuelle tendance à se fier automatiquement ou excessivement aux résultats produits par un système d'IA (« biais d'automatisation »)* » ;
- « *Être en mesure d'interpréter correctement les résultats du système d'IA* » ;
- « *Être en mesure de décider, dans une situation particulière, de ne pas utiliser le système d'IA ou de négliger, passer outre ou inverser le résultat fourni par ce système* » : comment un médecin justifiera-t-il de passer outre la recommandation d'un système numérique conçu pour la fournir une aide ? S'il est attaqué, comment démontrera-t-il la pertinence de sa décision ?

Les paradoxes du contrôle humain

Pourquoi et pour quoi automatise-t-on des fonctions cognitives ?

- Le problème n'est pas appréhendable humainement (calculs, combinatoire)
- Le contexte applicatif est dangereux ou hostile
- L'automatisation est plus économique
- L'automatisation est plus « performante »
- L'automatisation est plus sûre (par exemple en palliant l'erreur humaine)

Un contrôle humain des fonctions automatisées est exigé alors même qu'un système numérique est mis en place pour pallier les limites humaines.

En résumé : la machine, l'humain et leur interaction

La machine est conçue pour répondre à un cahier des charges et être complémentaire de l'être humain. Elle pourra détecter des éléments que ne voit pas l'humain, mais il s'agira parfois d'artefacts. Un certain nombre de questionnements doivent être soulevés :

- Quels sont les modèles d'interprétation, sur la base de quoi la machine calcule-t-elle sa décision ?
- Quelles sont les données d'entraînement si le système est à base d'apprentissage machine ?
- Quels sont les modèles d'incertitude ?
- Comment est gérée la distinction de situations « proches » mais menant à des décisions différentes ?

L'humain : il dispose des capacités d'invention, de jugement. Il peut différer et déléguer la décision, et demander des informations supplémentaires.

On oublie souvent qu'il doit pouvoir continuer à exercer même s'il n'a pas ou plus l'aide du système numérique, ce qui pose la question des savoir-faire en conditions numériques dégradées.

Par ailleurs, l'humain est faillible : dire qu'on fait vérifier le résultat du logiciel par un humain n'est pas une garantie.

L'interaction entre humain et machine : le résultat de la machine influence le raisonnement humain (biais d'automatisation), il s'agit d'une altération de l'autonomie d'appréciation. L'interaction entre l'humain et la machine pose de nombreuses questions : comment garantir la pertinence de l'intervention humaine ? Comment justifier des appréciations contradictoires ? Comment procéder à la qualification du système humain – machine ?