

Intelligence Artificielle et Santé : Dépasser le cadre traditionnel de l'éthique du numérique

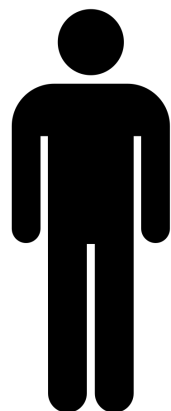
Brigitte Séroussi

Professeure d'Informatique Biomédicale,
Sorbonne Université

Chercheuse au Limics,
Laboratoire de recherche en informatique pour la
santé

Praticien hospitalier,
Assistance publique – Hôpitaux de Paris

Le numérique en santé, c'est quoi ?



Dossier Patient Informatisé
Système d'Information
Hospitalier



Dossier Patient Informatisé
Logiciel métier de ville



Dossier Résident Informatisé
Système d'Information Médico
Social



mon
ESPACE
SANTÉ



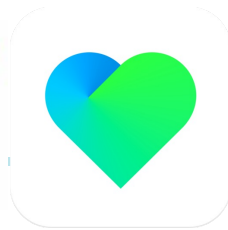
Dossier Patient Informatisé
Système d'Information
Hospitalier



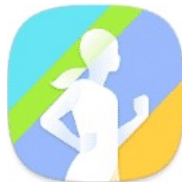
Dossier Patient Informatisé
Logiciel métier de ville



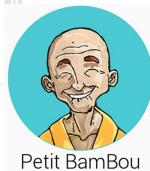
Dossier Résident Informatisé
Système d'Information Médico
Social



Dossier Patient Informatisé
Système d'Information
Hospitalier



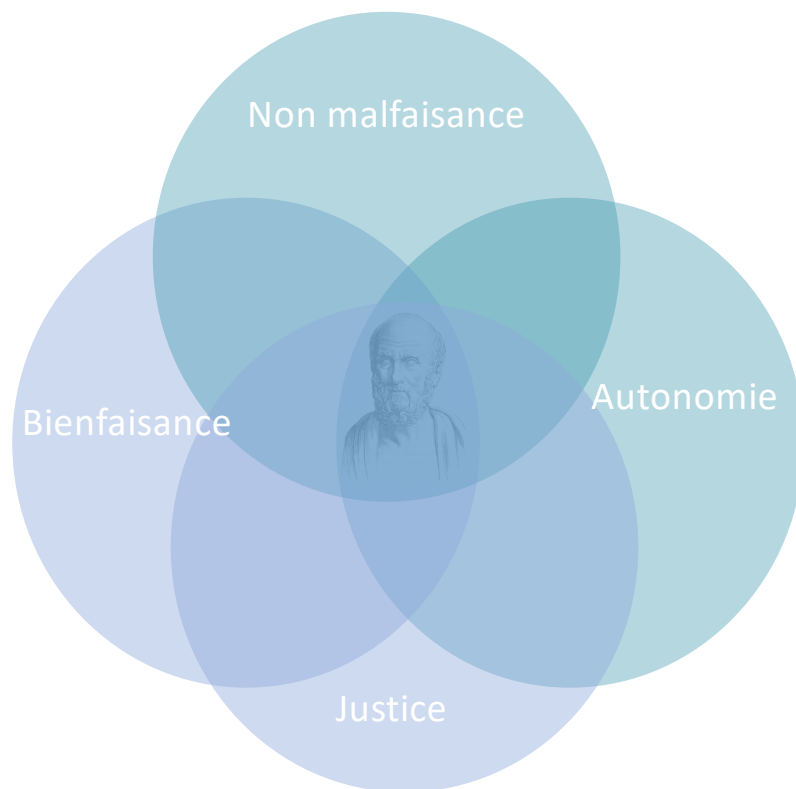
Dossier Patient Informatisé
Logiciel métier de ville



Dossier Résident Informatisé
Système d'Information Médico
Social

**Et l'éthique du numérique en
santé, c'est quoi ?**

L'Éthique médicale selon Hippocrate



SERMENT D'HIPPOCRATE

Au moment d'être admis(e) à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.

Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu(e) à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs.

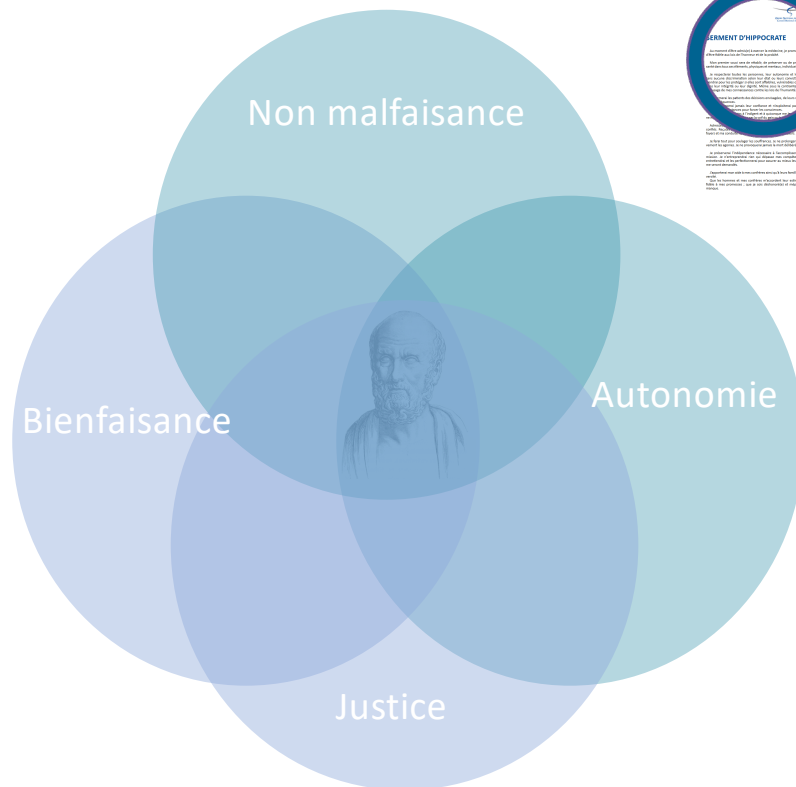
Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré(e) et méprisé(e) si j'y manque.

L'Éthique médicale selon Hippocrate



« Je **respecterai** toutes les personnes, leur autonomie et leur volonté, sans **aucune discrimination** selon leur état ou leurs convictions.

J'interviendrai pour les **protéger** si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.

Je ne tromperai jamais leur **confiance** et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

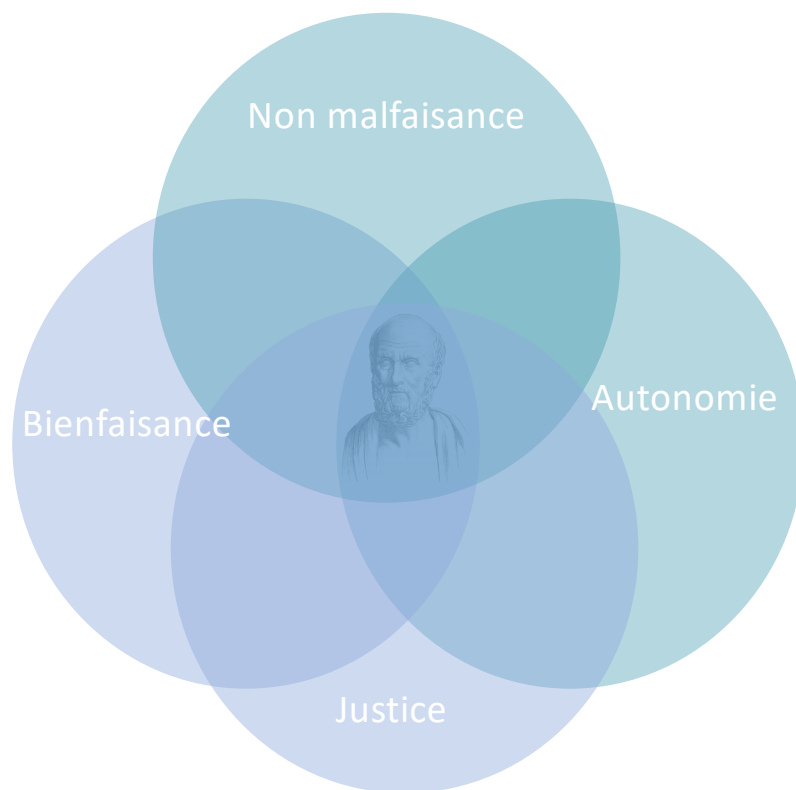
Admis(e) dans l'intimité des personnes, je tairai **les secrets** qui me seront confiés. Reçu(e) à l'intérieur des maisons, je respecterai les secrets des foyers... »

L'Éthique médicale selon Hippocrate



L'Éthique du numérique

Facile à utiliser



Accessible
à tous



Au service
de
l'utilisateur

Durable
(Eco-responsable)



Les outils et services numériques en santé répondent aux besoins des utilisateurs, leur utilisation leur procure un bénéfice.

Les outils et services numériques en santé ne sont pas toxiques, ne font pas de mal, ne génèrent ni stress ni colère aux utilisateurs, ne leur causent aucun préjudice.

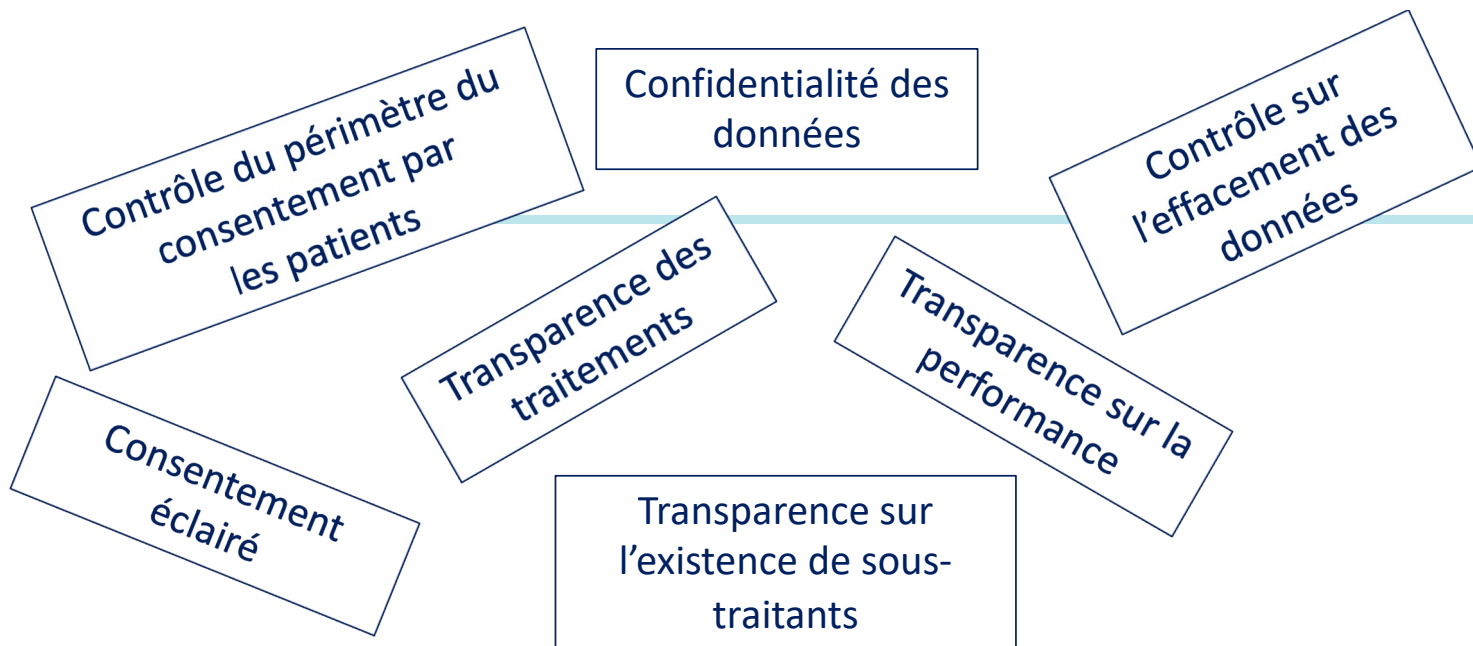


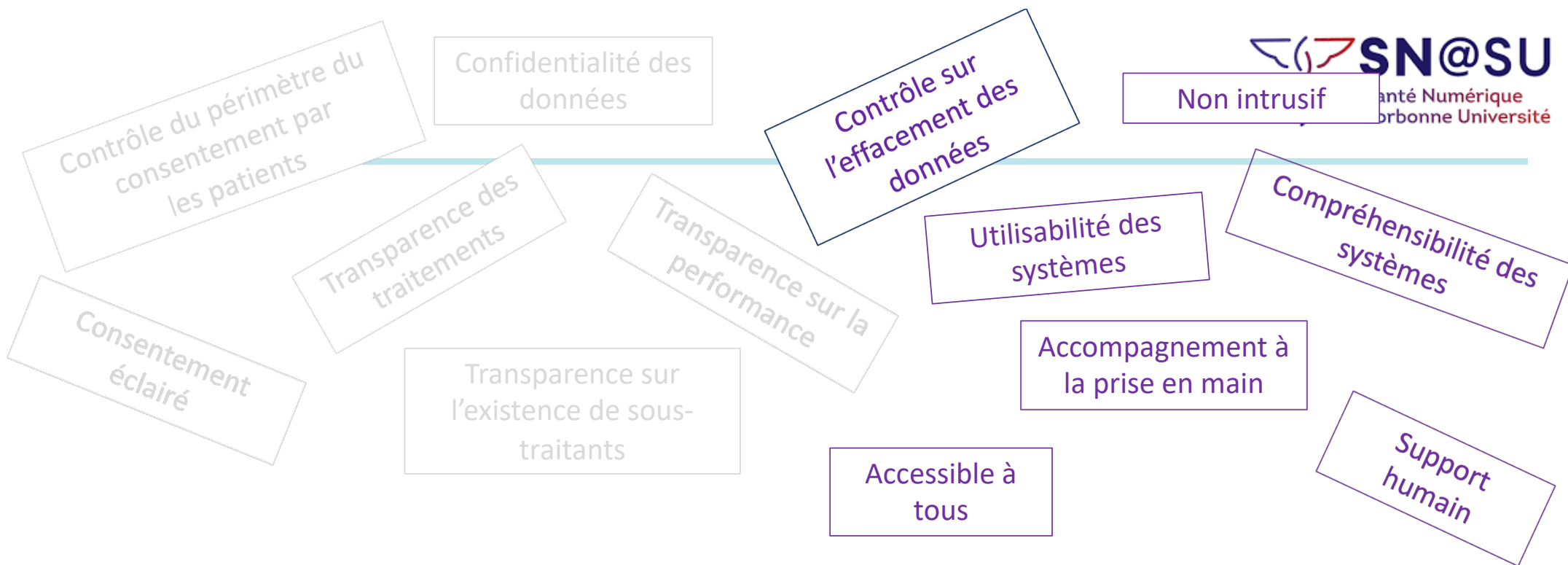
Les outils et services numériques permettent aux utilisateurs d'être justes en proposant une égalité de traitement des personnes. Les charges et les avantages doivent être répartis équitablement.

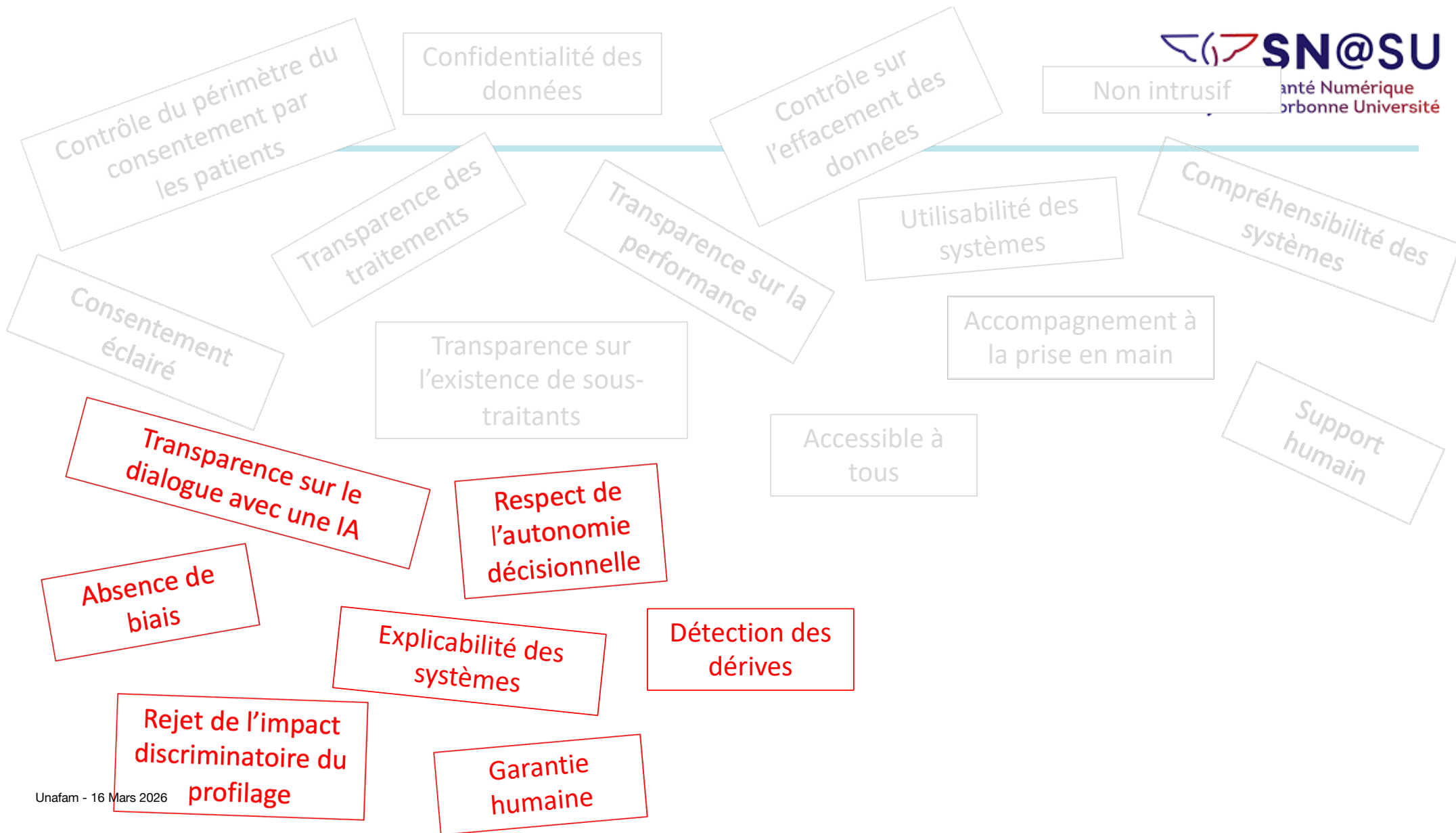
Les utilisateurs des outils et services numériques (qu'ils soient patients ou professionnels de santé) doivent conserver leur autonomie de pensée, d'intention et d'action

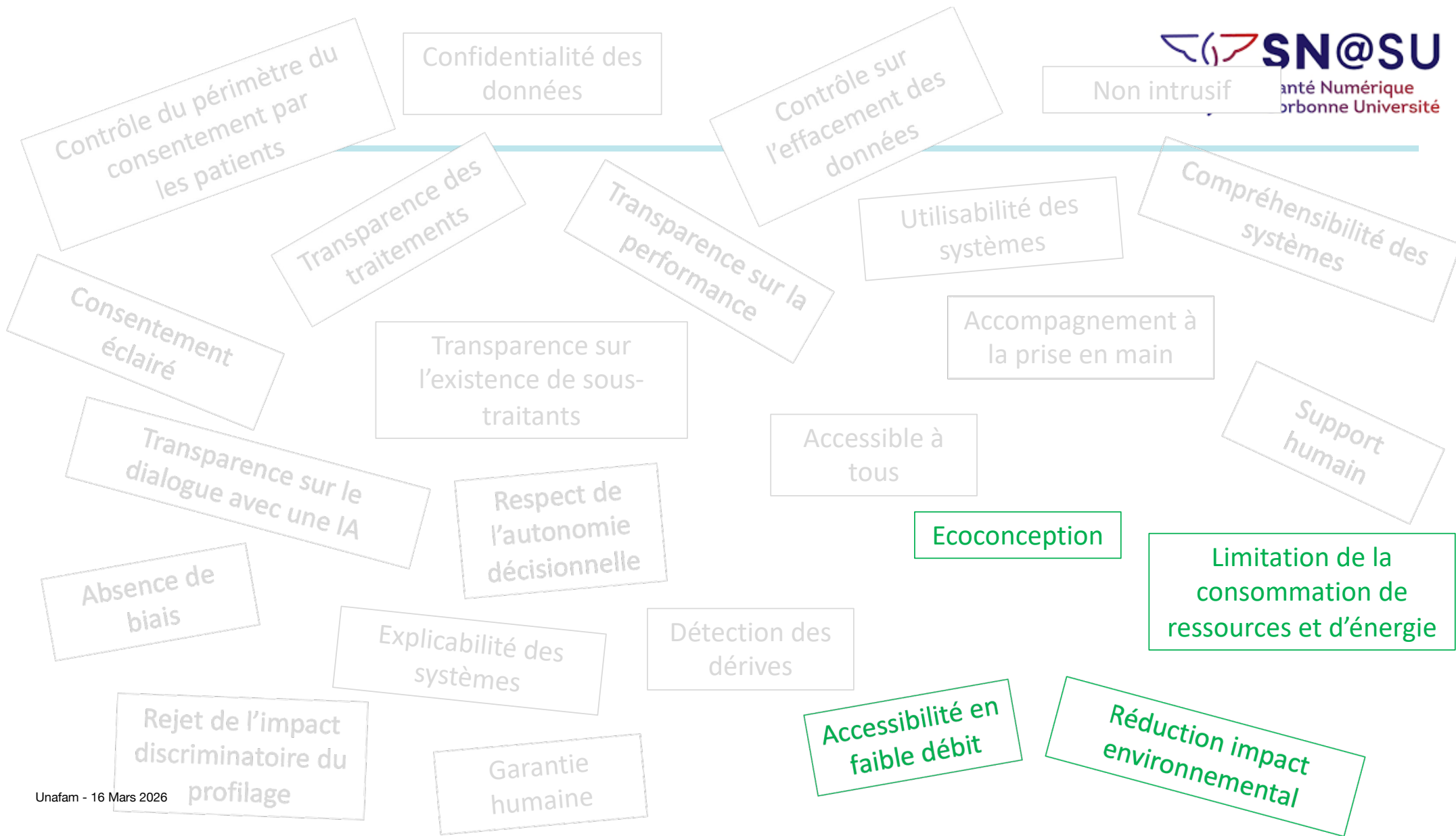


Le développement et les usages du numérique en santé ne doivent pas participer à la détérioration de la santé humaine et privilégier la sobriété numérique, l'écoconception, la gestion de la fin de vie des matériels dans un principe d'économie circulaire (ré-usage / recyclage).









S'il n'est pas éthique de déployer un numérique en santé qui n'est pas éthique

S'il n'est pas éthique de déployer un numérique en santé qui n'est pas éthique

Il n'est pas éthique de ne **pas** déployer un numérique en santé qui **serait éthique**

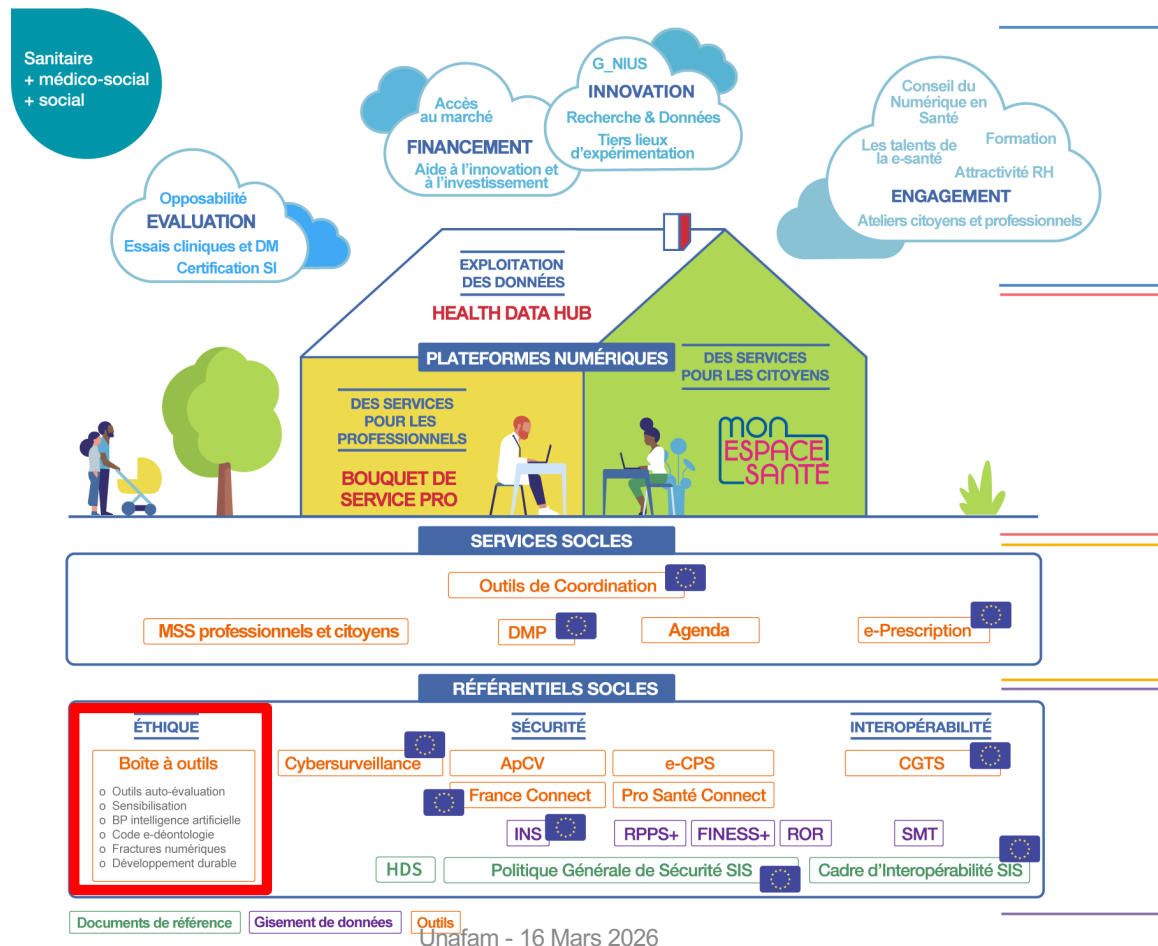
S'il n'est pas éthique de donner les moyens de santé numérique en santé qui n'est pas éthique

Il n'est pas éthique de déployer un numérique en santé qui **serait éthique**

L'éthique du numérique en santé

Action 4 de la FDR 2019-2022

Cellule éthique du numérique en santé



ORIENTATION 5
SOUTENIR L'INNOVATION ET FAVORISER L'ENGAGEMENT DES ACTEURS



ORIENTATION 4
DÉPLOYER AU NIVEAU NATIONAL DES PLATEFORMES NUMÉRIQUES DE SANTÉ

ORIENTATION 3
ACCÉLÉRER LE DÉPLOIEMENT DES SERVICES NUMÉRIQUES SOCLES



ORIENTATION 2
INTENSIFIER LA SÉCURITÉ ET L'INTEROPÉRABILITÉ DU NUMÉRIQUE EN SANTÉ



Action ré-affirmée dans la 2^e FDR

Mettre le numérique au service de la santé

Ces quatre dernières années (2019-2022), la France a rattrapé son retard dans la e-santé !

Des défis de longue date comme l'identité Nationale de Santé ont enfin vu le jour. Pilier de notre souveraineté, Mon espace santé, le carnet de santé numérique des Français est désormais ouvert pour plus de 90 % de la population. Grâce au Ségur Numérique, les patients et les personnes accompagnées récupèrent systématiquement une copie de leurs documents de santé : 20 fois plus de documents ont été envoyés aux citoyens en 2022 par rapport à 2021. L'entrée dans le droit commun de la télésurveillance et la prise en charge anticipée des dispositifs médicaux numériques sont des avancées majeures, scrutées à l'international. **Impulsée par la précédente feuille de route, une dynamique collective et ambitieuse désormais est lancée.**

Ces efforts collectifs et les premiers résultats obtenus ont fait l'objet d'un bilan, intitulé FAIT(S), publié en août 2022. **La recette de l'État plateforme a fonctionné** : l'État s'engage dans un rôle de régulateur et d'opérateur des référentiels et services socles et les acteurs privés déploient les services utiles au quotidien des soignants et des patients et développent les innovations dont notre système de santé a besoin.

Après toutes ces avancées, **nous devons consolider nos acquis** et nous assurer que les usages numériques se développent sur le terrain afin de répondre aux besoins des citoyens et des acteurs de santé, des secteurs médico-social, social et sanitaire. Maintenant que les fondations essentielles existent, nous pouvons davantage nous appuyer sur le numérique pour faire face aux défis majeurs que rencontre notre système de santé. Nous devons **mettre le numérique au service de la santé**, en contribuant humblement à répondre aux enjeux prioritaires de la stratégie nationale de santé et des plans régionaux de santé, et notamment la prévention, l'amélioration de la qualité de la prise en charge et l'accès à la santé.

Pour y arriver, cette nouvelle feuille de route s'enracine dans des valeurs cardinales stables.



Heïa Ghariani et Raphaël Beaufret, co-responsables du numérique en santé

L'éthique.

Le numérique en santé doit être massivement inclusif et solidaire. Chacun doit pouvoir utiliser les outils et services numériques pour être acteur de sa santé et de son parcours de vie, accéder à ses données et en contrôler les accès. Le numérique en santé doit être solidaire et ne laisser personne de côté. Le numérique ne doit pas accroître les inégalités en matière de santé mais bien renforcer l'accès aux soins pour tous. Les innovations ne doivent pas être réservées à une frange de la population mais bien être accessibles à ceux qui en ont le plus besoin. L'éthique n'est pas un frein au déploiement du numérique mais un garant de sa qualité et de son humanisme dans les usages.

La souveraineté.

Il est impératif de faire les choix technologiques qui nous laisseront maîtres de notre destin dans la durée. En définissant un cadre de régulation, en France et en Europe, en adéquation avec nos valeurs, nous pourrions garantir notre indépendance face aux intérêts particuliers et étrangers.

La durabilité.

Face aux immenses défis climatiques et écologiques auxquels l'humanité doit faire face, les services et dispositifs numériques doivent être évalués en termes d'impact environnemental, à la fois sur leurs coûts de construction et de maintien en service mais aussi sur les pratiques qu'ils modifient (ex : téléconsultation, etc.).

Ces valeurs, la France doit également les incarner à l'échelle européenne. **L'Europe sera demain le niveau pertinent des échanges, de la régulation et du marché du numérique en santé.** C'est pour cela que la France poursuivra son investissement dans les travaux européens (présidence française du Conseil de l'Union Européenne en 2022, co-présidence du e-Health Network en 2023, etc.) pour y peser et construire l'avenir, notamment en matière de souveraineté et d'hébergement des données et dans le cadre du futur règlement sur l'espace européen des données de santé (EHDS).

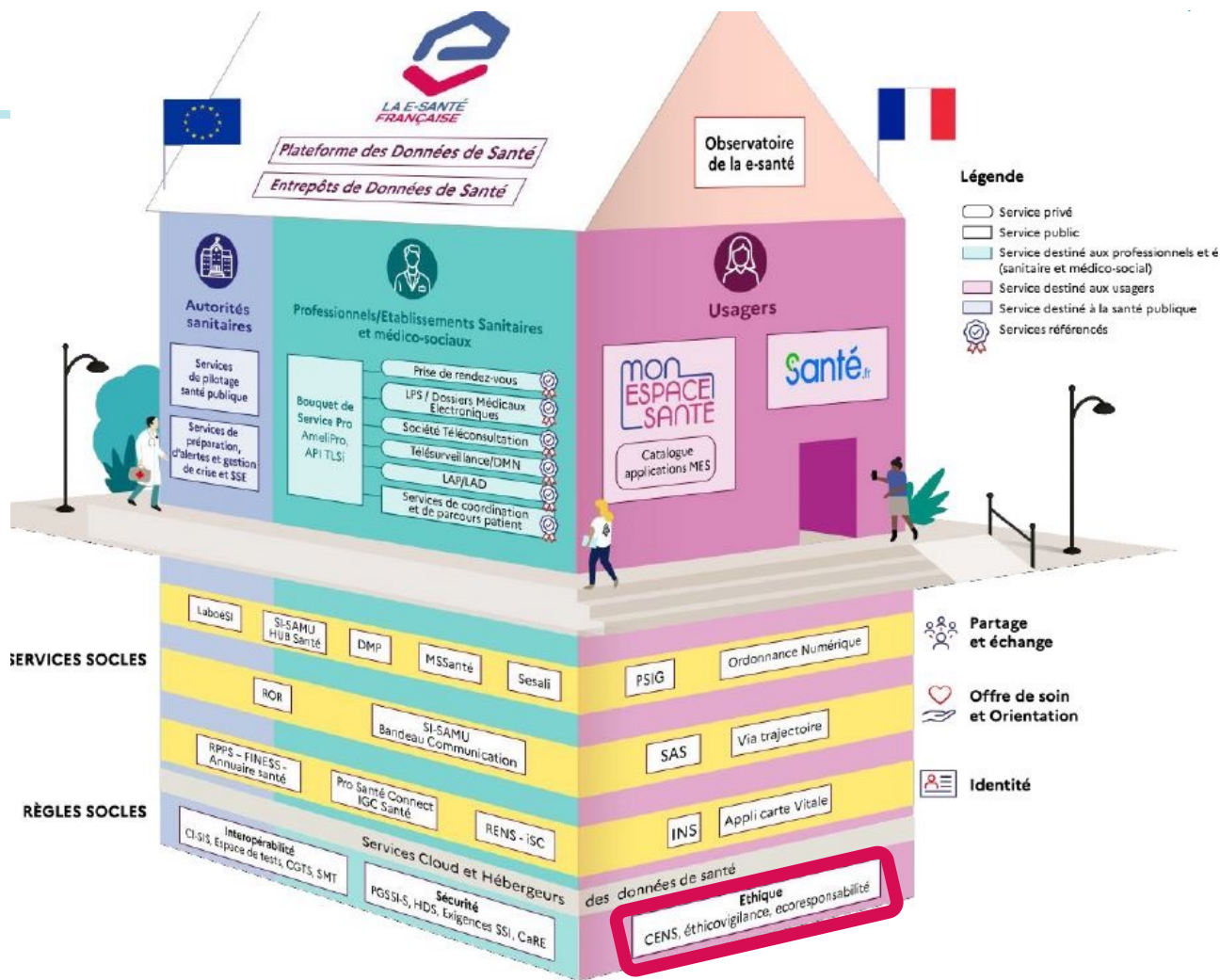
L'éthique

Le numérique en santé doit être massivement inclusif et solidaire. Chacun doit pouvoir utiliser les outils et services numériques pour être acteur de sa santé et de son parcours de vie, accéder à ses données et en contrôler les accès. Le numérique en santé doit être solidaire et ne laisser personne de côté. Le numérique ne doit pas accroître les inégalités en matière de santé mais bien renforcer l'accès aux soins pour tous. Les innovations ne doivent pas être réservées à une frange de la population mais bien être accessibles à ceux qui en ont le plus besoin. L'éthique n'est pas un frein au déploiement du numérique mais un garant de sa qualité et de son humanisme dans les usages.



L'éthique

Le numérique en santé doit être massivement inclusif et solidaire. Chacun doit pouvoir utiliser les outils et services numériques pour être acteur de sa santé et de son parcours de vie, accéder à ses données et en contrôler les accès. Le numérique en santé doit être solidaire et ne laisser personne de côté. Le numérique ne doit pas accroître les inégalités en matière de santé mais bien renforcer l'accès aux soins pour tous. Les innovations ne doivent pas être réservées à une frange de la population mais bien être accessibles à ceux qui en ont le plus besoin. L'éthique n'est pas un frein au déploiement du numérique mais un garant de sa qualité et de son humanisme dans les usages.



GT1 : films
d'animation

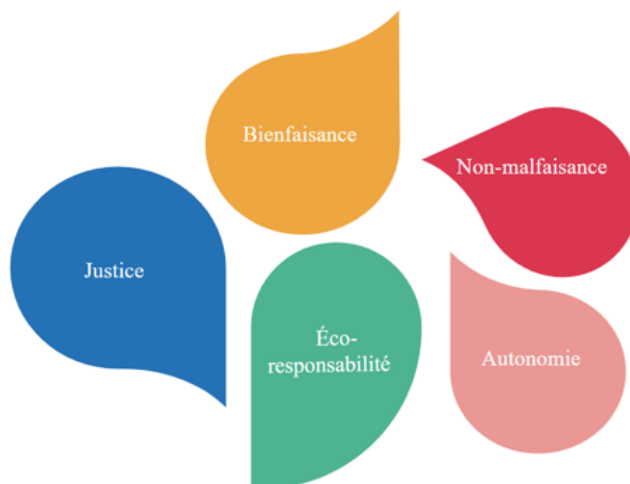
GT2 : éthique
des SIH

GT3 : éthique
de l'IA

GT4 : code de
e-déontologie

GT5 : journées
régionales

GT6 :
écoresponsabilité



GT7 :éthique
des LGC

GT8 : éthique
de la télésanté

GT11 : éthique
des SI ESMS

GT12 : éthique
des applis

GT13 : éthique
de la TLC

GT14 : éthique
des DMN

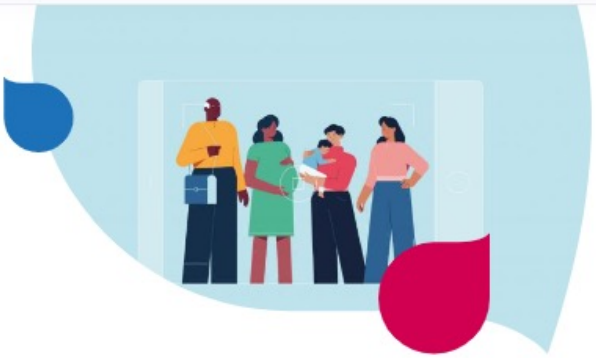
GT15 : éthique
du jumeau
numérique

*GT16 : éthique
de l'IA pour la
démocratie
sanitaire*

Accueil > Nos offres de produits et services > Le cadre de l'éthique du numérique en santé (CENS)

Le cadre de l'éthique du numérique en santé (CENS)

Pour un numérique en santé humaniste, citoyen, inclusif et écoresponsable



L'éthique du numérique en santé repose sur des principes moraux partagés par tous qui guident le développement et l'utilisation des technologies numériques en santé vers un numérique préservant l'intérêt de l'utilisateur.

Co-construit avec les acteurs de l'écosystème et le souhait de couvrir la richesse de l'ensemble des points de vue, le Cadre de l'éthique du numérique en santé (CENS) répond à la volonté exprimée dans le [rapport Pon-Courty](#) d'inscrire l'accélération du virage numérique en santé dans un cadre de valeurs éthiques.

Ainsi, le CENS a vocation à traduire de façon opérationnelle et pragmatique des principes éthiques et humanistes dans l'objectif de garantir la confiance des citoyens et des soignants.

L'éthique du numérique en santé est une discipline vivante qui doit s'adapter en continu aux nouveaux outils technologiques, aux nouvelles pratiques de soins et aux nouvelles réglementations.

L'éthique du numérique en santé en un clic

- [Définition](#)
- [Corpus documentaire](#)
- [Construction et mise à jour](#)
- [Contexte réglementaire](#)
- [Dates clés](#)
- [L'éthique en chiffres](#)



Une définition de l'éthique du numérique en santé

L'éthique du numérique en santé se positionne à l'intersection de l'éthique clinique telle que décrite dans le serment d'Hippocrate et de l'éthique du numérique, le numérique étant considéré comme un outil au service des utilisateurs.

L'éthique clinique est ancrée sur 4 principes aujourd'hui bien connus : la bienfaisance, la non-malfaisance, le respect de l'autonomie du sujet, et la justice ou l'équité.

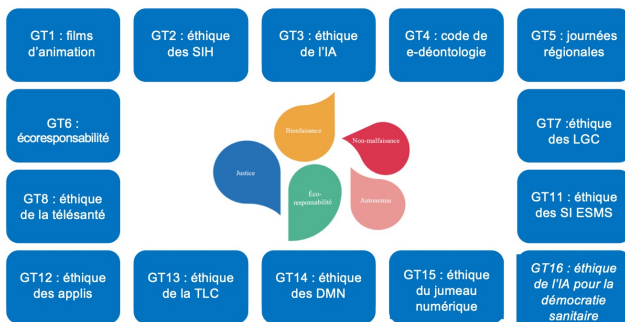
Ainsi l'éthique du numérique est plurifactorielle, et si on continue la métaphore de l'outil, on attend d'un outil qu'il soit facile à utiliser, accessible à tous, au service de son

Unafam - 16 Mars 2026

Besoin d'être guidé ?
Trouvez l'information ou la démarche correspondant à votre situation

<https://esante.gouv.fr/produits-services/referentiel-ethique>

L'IA c'est quoi ?



Donne moi une définition courte de l'IA dans le domaine de la santé ?

L'intelligence artificielle (IA) en santé désigne l'ensemble des technologies capables d'analyser des données médicales pour **aider au diagnostic, au traitement, à la prévention et à la gestion des soins**, en soutenant les professionnels de santé dans leur prise de décision.



L'IA c'est quoi ?

En quoi l'IA générative est incluse dans la première définition
"L'intelligence artificielle (IA) en santé désigne l'ensemble des
technologies capables d'analyser des données médicales pour aider
au diagnostic, au traitement, à la prévention et à la gestion des soins,
en soutenant les professionnels de santé dans leur prise de décision."

Excellente question 🙌

L'**IA générative** est une **sous-catégorie** de l'intelligence artificielle. Elle entre donc dans ta première définition, mais avec une **fonction spécifique** :

- La définition générale parle d'IA qui **analyse des données médicales pour aider à la décision**.
- L'IA générative, elle, **ne se limite pas à analyser**, elle **produit de nouveaux contenus** à partir de ce qu'elle a appris.

👉 Autrement dit, **l'IA générative étend le champ de l'IA en santé** :

- elle peut **générer des comptes rendus médicaux, des résumés de consultations, des données synthétiques pour la recherche, ou encore des supports de formation** ;
- mais ces productions reposent toujours sur l'analyse et la compréhension de données médicales — ce qui la rattache pleinement à la définition générale de l'IA en santé.

Souhaitez-vous que je reformule la première définition pour qu'elle **inclue explicitement l'IA générative** sans l'alourdir ?

Les bénéfices de l'IA en santé

Libérer du temps aux soignants

“On est toujours centré sur l'écran, on écoute moins le patient. Ces tâches prennent beaucoup de temps médical.”

“Les antécédents sont repris avec précision, les codes CIM10 sont intégrés. Je peux aussi générer une recommandation patient ou une lettre d'adressage en un clic.”

“On estime gagner 3 minutes sur une consultation de 15 minutes. Sur 20 à 30 patients par jour, c'est entre 1 et 1h30 de récupérées.”

« Mais ce qui compte, c'est ce qu'on fait de ce temps. »

Unafam - 16 Mars 2026



Nabla lance un assistant pour les médecins basé sur ChatGPT 🤖

La start-up française spécialisée dans les outils de communication dans le secteur de la santé permet de générer automatiquement des comptes rendus à la suite d'une consultation. Elle vient également de créer une filiale aux Etats-Unis pour y commercialiser son produit.



Les bénéfices de l'IA en santé

Aide à la lecture d'images

Cas des mammographies de dépistage

Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Performance of a Breast Cancer Detection AI Algorithm Using the Personal Performance in Mammographic Screening Scheme

Yan Chen, PhD • Adnan G. Taib, BMBS • Iain T. Darker, PhD • Jonathan J. James, FRCR

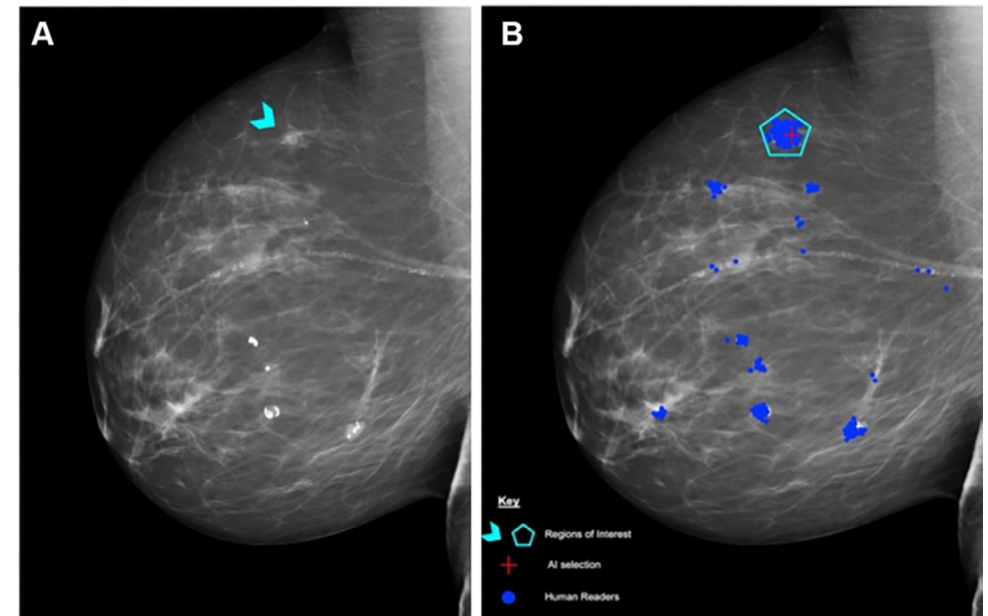
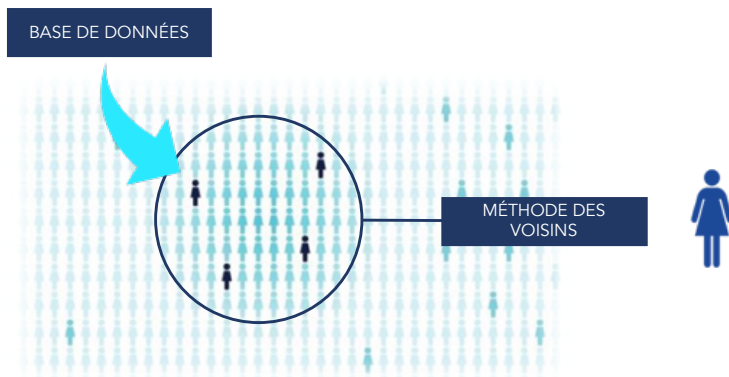


Figure 2: (A) Right mediolateral oblique unadulterated mammogram shows an 8-mm ill-defined mass (arrowhead), which, after biopsy, was determined to be a histologic grade 2 ductal carcinoma of no special type. (B) Mammogram shows findings by human readers (blue areas) and the Lunit INSIGHT MMG artificial intelligence (AI) algorithm (red cross). Each blue dot is a mark placed by an individual human reader on a perceived abnormality when the Personal Performance in Mammographic Screening (PERFORMS) case was read. A region of interest (pentagon) has been annotated by the PERFORMS scheme organizers and their expert radiology panel. AI has correctly marked the region of interest in the right breast for recall. Source: PERFORMS via Yan Chen.

Les bénéfices de l'IA en santé

Risque personnalisé de cancer du sein

Calcul du risque personnalisé de cancer du sein



ESTIMER LE RISQUE DE CANCER DU SEIN

Dossier n°99999999 Jane @

1. ÂGE DE LA PATIENTE :

40 - 44 45 - 49 50 - 54 55 - 59
60 - 64 65 - 69 70 - 74

2. MÈRE, SOEUR OU FILLE AYANT EU UN CANCER DU SEIN :

0 1 2 ou +

3. ANTÉCÉDENTS DE BIOPSIE MAMMAIRE :

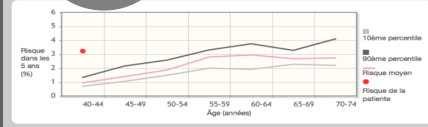
NON OUI

4. ESTIMATION DE LA DENSITÉ MAMMAIRE :

A B C D

RISQUE DE DÉVELOPPER UN CANCER DU SEIN DANS LES 5 ANS
3,2%

CATÉGORIE DE RISQUE DE VOTRE PATIENTE
TRÈS ÉLEVÉE



PROTOCOLE PERSONNALISÉ DE DÉPISTAGE

RETOUR À L'INDEX

MammoRisk®


<https://predilife.com/home.php/>

Les bénéfices de l'IA en santé

Prise en charge thérapeutique



Reset

Predict n'est pas conçu pour être utilisé dans tous les cas. [Cliquez ici pour plus de détails.](#)
Si vous n'êtes pas sûr des renseignements à compléter, cliquez sur le bouton  pour plus d'information.

DCIS
seulement ou
LCIS
seulement?


 Oui Non

Age au
diagnostic


 - 59 +

L'âge doit être entre 25 et 85 ans

Menopausée?

 Oui Non Inconnu


Statut ER

 Positif Négatif

Statut HER2

 Positif Négatif Inconnu

Statut Ki-67

 Positif Négatif Inconnu

Positif signifie supérieur à 10%

Taille de la
tumeur
invasive (mm)

 - 23 +

S'il y avait plus d'une tumeur, entrez la taille de la plus grande tumeur. Si un traitement d'induction (néo-adjuvant) a été employé, entrez la taille avant le traitement néo-adjuvant.

Grade tumoral

 1 2 3

Déteecté par

 Dépistage Symptomes Inconnu

Ganglion positif

 - 3 +


Micrométastases
seulement

 Oui Non Inconnu

Activé lorsque les ganglions positifs sont à 1.

Options de traitement

Traitement
hormonal

 Non 5 ans 10 ans

Traitement hormonal
disponible lorsque le statut ER est positif

Chimiothérapie

 Aucune 2ème gén 3ème gén

Trastuzumab

 Non Oui

Disponible avec la chimiothérapie lorsque le statut HER2 est positif

Bisphosphonates

 Non Oui

Disponible pour les femmes ménopausées

Résultats

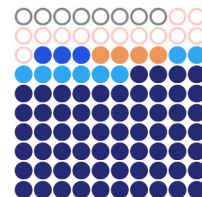
All treatments have side effects. Weigh up the benefits shown with the side effects in [this website](#).

Tableau Courbes Graphique Texte Schéma

Sélectionnez le nombre d'années depuis la chirurgie que vous souhaitez considérer:

5 10 15

Ce tableau montre le nombre de femmes qui survivent au moins 10 ans après la chirurgie.



- 8 morts dû à une autre cause
- 13 morts reliés au cancer
- 3 gain(s) en personnes survivantes dus aux bisphosphonates
- 4 gain(s) en personnes survivantes dus à la chimiothérapie
- 8 gain(s) en personnes survivantes dus de l'hormonothérapie
- 64 personnes survivantes avec chirurgie seule

predict
breast cancer

<https://breast.predict.com/>

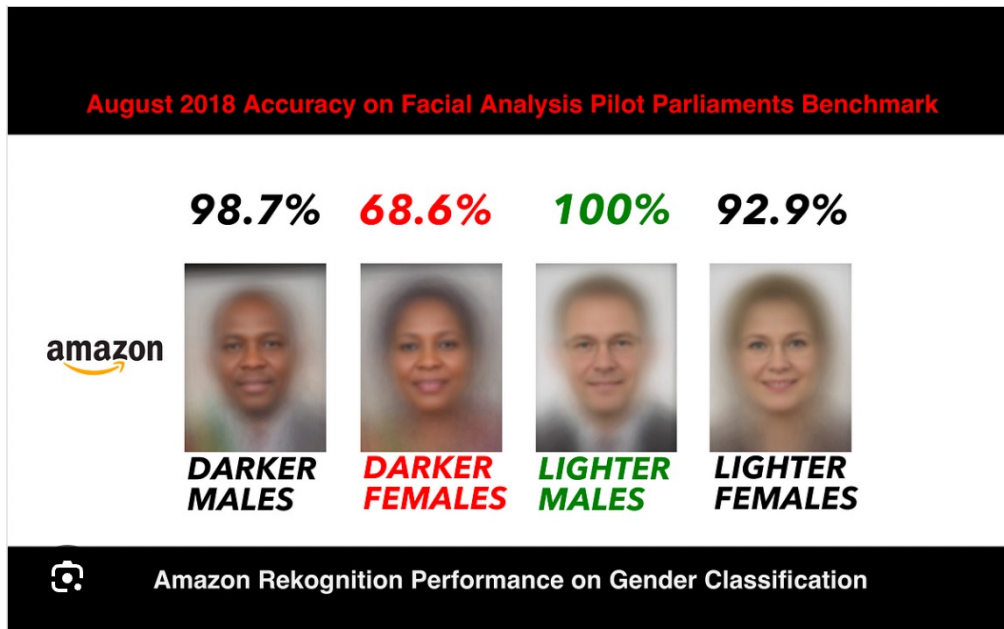
Les risques de l'IA



RISKS OF AI



Les biais de l'IA



AI, facial recognition too white, too male: More minorities needed

[Consulter >](#)

Les biais discriminatoires de l'IA en santé

THEMA
RADIOLOGIE

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

IRSNI DOSIMÉTRIE
INSTITUT DE RADIOPROTECTION
ET DE SÛRETÉ NUCLÉAIRE

Mammographie : les imprécisions d'un algorithme d'IA selon l'âge ou la race de la patiente

MARDI 21 MAI 2024 Soyez le premier à réagir

Dans une étude portant sur près de 5 000 mammographies de dépistage interprétées par un algorithme d'IA approuvé par la Food & Drug Administration (FDA), les caractéristiques des patientes telles que la race et l'âge ont influencé des résultats faussement positifs. Les résultats de cette étude ont été publiés dans la Revue Radiology.

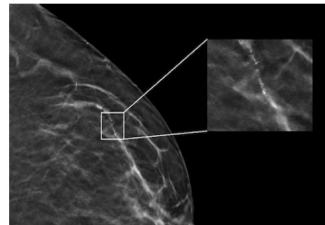
RSNA®
Radiological Society
of North America

Les performances des solutions d'aide à l'interprétation des images médicales construites autour d'algorithmes d'Intelligence Artificielle (IA) suscitent, maintenant que nous avons un peu de recul, certaines interrogations.

Les performances des algorithmes d'IA mise en cause dans le dépistage du cancer du sein

Une étude publiée dans la Revue Radiology et conduite par le Dr Derek L.

Nguyen, professeur adjoint à la Duke University de Durham (Caroline du Nord - USA), identifie des dysfonctionnements ponctuels qu'il est nécessaire de prendre en compte. « L'IA est devenue une ressource permettant aux radiologues d'améliorer leur efficacité et leur précision dans la lecture des mammographies de dépistage tout en atténuant l'épuisement des lecteurs, annonce en préambule le Dr Nguyen. Cependant, l'influence des caractéristiques des patients sur les performances de l'IA n'a pas été bien étudiée. »



Des faux positifs plus nombreux chez les patientes de race noire ou d'âge plus avancé

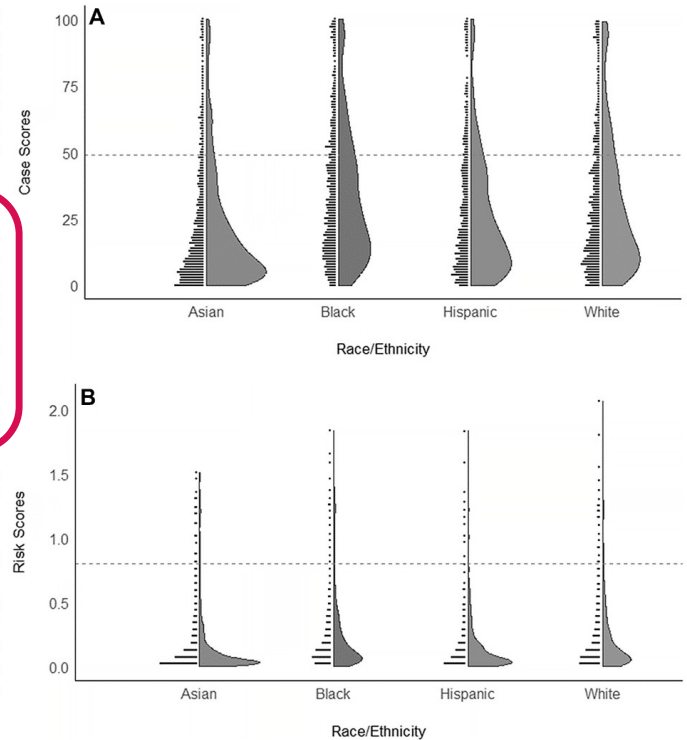
Les scores de cas faussement positifs étaient significativement plus probables chez les patientes noires et plus âgées (71 à 80 ans) et moins probables chez les patientes asiatiques et les patientes plus jeunes (41 à 50 ans) par rapport aux patientes blanches et aux femmes âgées de 51 à 60 ans.

« Cette étude est importante car elle met en évidence que tout logiciel d'IA acheté par un établissement de santé peut ne pas fonctionner de la même manière selon l'âge, la race, l'origine ethnique et la densité mammaire des patients, commente le Dr Nguyen. À l'avenir, je pense que les mises à niveau des logiciels d'IA devraient se concentrer sur la garantie de la diversité démographique. »

Mieux comprendre le type de population exploré avant d'acquiescer un algorithme d'IA

Le Dr. Nguyen exhorte les centres de radiologie à mieux comprendre la population de patients qu'ils explorent avant d'acheter un algorithme d'IA pour l'interprétation des mammographies de dépistage et à interroger les fournisseurs sur la formation de leurs algorithmes.

« Avoir une connaissance de base des données démographiques de votre établissement et interroger le fournisseur sur la diversité ethnique et d'âge de ses données de formation vous aidera à comprendre les limites auxquelles vous serez confronté dans la pratique clinique », conclut-il.



Les risques de l'IA



RISKS OF AI



L'éthique de l'IA

Recommandations de bonnes pratiques pour une éthique *by design* en IA



Accueil > Espace presse > Intelligence artificielle : la Délégation ministérielle au numérique en santé publie ses recommandations de bonnes pratiques pour une éthique « by design » des solutions d'IA

COMMUNIQUÉ DE PRESSE

Intelligence artificielle : la Délégation ministérielle au numérique en santé publie ses recommandations de bonnes pratiques pour une éthique « by design » des solutions d'IA

Publié le 31 mai 2022



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

ActuIA

ARTIFICIELLE THÈMES • SECTEURS • ANNUAIRE • AGENDA TUTORIELS ÉTUDES

La DNS publie ses recommandations pour une éthique de l'IA « by design » pour la santé



La Délégation ministérielle au Numérique en Santé (DNS) a publié en avril dernier ses *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 »*. Retour sur ce rapport.

La DNS publie ses recommandations pour une éthique de l'IA « by design »

ACTUIA, 13/07/2022

Partagé par : Beesens TEAM



Unafam - 16 Mars 2026

dans le secteur de la e-santé.

Enfin, tracer et documenter cette procédure permettra non seulement d'apporter la preuve des efforts déployés pour assurer un traitement éthique et licite des données de santé, mais également d'améliorer la solution.

5 ÉTAPES POUR CONSTRUIRE UNE SOLUTION D'IA EN MATIÈRE DE SANTÉ

- 01 CADRAGE**
Préparation et validation la finalité de la solution développée d'un point de vue éthique.
- 02 COLLECTE DES DONNÉES**
Réflexion sur les différentes mesures mises en œuvre pour assurer une collecte légale et sécurisée des données.
- 03 PRÉ-TRAITEMENT DES DONNÉES**
Élimination des données fausses, qualification des données conservées et élaboration de solutions pour pallier aux données manquantes.
- 04 CONSTRUCTION DE L'ALGORITHME**
Consolidation et documentation du choix d'apprentissage.
- 05 ÉVALUATION DE L'ALGORITHME**
Test de l'algorithme et évaluation de son comportement du regard des principes d'éthique et de sécurité.



Les enjeux d'une éthique « by design » en matière d'IA

L'élaboration d'une solution d'intelligence Artificielle éthique « by design » implique qu'au stade de la construction de l'algorithme la question de la transparence de la solution soit abordée.

Cadrage

Définition de la finalité de la solution d'IA

Validation éthique de la finalité

Caractérisation des principes de gouvernance et de responsabilité

Collecte des données

Mise en œuvre des mesures pour assurer :

- Le consentement éclairé des patients à la réutilisation de leurs données au-delà de la finalité première du recueil (conformité RGPD, article 5 et 7)
- La proportionnalité des données collectées par rapport à la finalité du traitement servant l'élaboration de la solution d'IA (conformité RGPD, article 5, c)
- La non réidentification directe des données (agrégées, pseudonymisées, anonymisées) (conformité RGPD, article 32)
- La qualité des données (lutte contre les biais cognitifs)
- La représentativité de la population d'analyse / population cible / prévention des discriminations (lutte contre les biais de sélection)

Mise en œuvre des mesures de sécurité visant à assurer :

- Le transfert sécurisé des données (source unique, sources multiples, chaînage, intégrité)
- La qualité de l'hébergement des données, serveurs localisés en France / Europe (Cloud privé/public, centralisé/distribué) (conformité HDS)
- La cyber-sécurité à l'état de l'art

Mise en œuvre de mesures pour garantir la non-réutilisation non éthique des données (par ex. en cas de fusion de la société / modifications législatives (pouvant aller jusqu'à la destruction automatique des données))

Pré-traitement des données

Mise en œuvre des mesures de :

- Traitement des données manquantes (réduction des biais)
 - Rééquilibrage des populations minoritaires (réduction des biais)
- Séparation des données (deux jeux totalement distincts, un échantillon pour l'apprentissage et un échantillon pour l'évaluation) et représentativité des deux jeux de données par rapport à la population cible et la finalité du traitement)

Construction de l'algorithme

Choix de l'algorithme d'apprentissage en adéquation avec la finalité

Définition :

- De la politique qualité de l'algorithme
- Des mesures de transparence mises en œuvre
- De la politique de traçabilité de la démarche de construction de l'algorithme
- De la politique d'explicabilité des résultats explicables, du processus d'auditabilité des résultats non explicables

Implémentation des fonctionnalités et mécanismes visant à assurer :

- L'identification et élimination des biais
- La correction des erreurs
- La traçabilité des traitements (rendre les codes sources publics avec protection par dépôt (APP))
- L'adaptabilité
- L'intégration des évolutions réglementaires
- L'intégration des avancées médicales
- La maintenance et le versionning

Définition des indicateurs de dérive du système

Evaluation de l'algorithme et préparation de la mise en production

Mise en œuvre des principes d'évaluation externe :

- Technique (bugs), clinique (gold standard, score de précision)
- De l'utilisabilité (PS, patients, usagers)
- De la non-discrimination/équité
- De la robustesse/reproductibilité

Mise en œuvre des procédures en cas de cyber-attaques (analyse d'impact sur la sécurité du système d'IA)

Mise en œuvre des mesures pour assurer l'information (juste et égalitaire) des utilisateurs (PS, patients) relative à :

- La finalité, gouvernance, responsabilité
- L'architecture
- L'origine des données et qualité (légalité de la collecte et des traitements)
- L'explication des processus, explication du périmètre de la partie non explicable
- La méthode d'apprentissage, d'inférence, etc.
- Les limites de l'utilisation de l'algorithme (FP, FN si classification)
- Les modalités de recours en cas d'erreurs
- L'implication des utilisateurs

Mise en œuvre des mécanismes de garantie humaine (PS, équipe de soins) pour assurer :

- Le contrôle par l'humain de l'IA
- L'autonomie décisionnelle
- Le maintien des compétences des utilisateurs
- L'intervalle de confiance de l'IA / garde-fou des erreurs de l'IA
- Les audits (désaccords IA / PS)

Définition de l'instance de régulation (audit, Label Ethique-IA)

Analyse d'impact organisationnel sur le parcours de soins

Analyse d'impact sociétal et environnemental

Cadrage

Définition de la finalité de la solution d'IA

Validation éthique de la finalité

Caractérisation des principes de gouvernance et de responsabilité

Collecte des données

Mise en œuvre des mesures pour assurer :

- Le consentement éclairé des patients à la réutilisation de leurs données au-delà de la finalité première du recueil (conformité RGPD, article 5 et 7)
- La proportionnalité des données collectées par rapport à la finalité du traitement servant l'élaboration de la solution d'IA (conformité RGPD, article 5, c)
- La non réidentification directe des données (agrégées, pseudonymisées, anonymisées) (conformité RGPD, article 32)
- La qualité des données (lutte contre les biais cognitifs)
- La représentativité de la population d'analyse / population cible / prévention des discriminations (lutte contre les biais de sélection)

Mise en œuvre des mesures de sécurité visant à assurer :

- Le transfert sécurisé des données (source unique, sources multiples, chaînage, intégrité)
- La qualité de l'hébergement des données, serveurs localisés en France / Europe (Cloud privé/public, centralisé/distribué) (conformité HDS)
- La cyber-sécurité à l'état de l'art

Mise en œuvre de mesures pour garantir la non-réutilisation non éthique des données (par ex. en cas de fusion de la société / modifications législatives (pouvant aller jusqu'à la destruction automatique des données))

Pré-traitement des données

Mise en œuvre des mesures de :

- Traitement des données manquantes (réduction des biais)
 - Rééquilibrage des populations minoritaires (réduction des biais)
- Séparation des données (deux jeux totalement distincts, un échantillon pour l'apprentissage et un échantillon pour l'évaluation) et représentativité des deux jeux de données par rapport à la population cible et la finalité du traitement)

Construction de l'algorithme

Choix de l'algorithme d'apprentissage en adéquation avec la finalité

Définition :

- De la politique qualité de l'algorithme
- Des mesures de transparence mises en œuvre
- De la politique de traçabilité de la démarche de construction de l'algorithme
- De la politique d'explicabilité des résultats explicables, du processus d'auditabilité des résultats non explicables

Implémentation des fonctionnalités et mécanismes visant à assurer :

- L'identification et élimination des biais
- La correction des erreurs
- La traçabilité des traitements (rendre les codes sources publics avec protection par dépôt (APP))
- L'adaptabilité
- L'intégration des évolutions réglementaires
- L'intégration des avancées médicales
- La maintenance et le versionning

Définition des indicateurs de dérive du système

Evaluation de l'algorithme et préparation de la mise en production

Mise en œuvre des principes d'évaluation externe :

- Technique (bugs), clinique (gold standard, score de précision)
- De l'utilisabilité (PS, patients, usagers)
- De la non-discrimination/équité
- De la robustesse/reproductibilité

Mise en œuvre des procédures en cas de cyberattaques (analyse d'impact sur la sécurité du système d'IA)

Mise en œuvre des mesures pour assurer l'information (juste et égalitaire) des utilisateurs (PS, patients) relative à :

- La finalité, gouvernance, responsabilité
- L'architecture
- L'origine des données et qualité (légalité de la collecte et des traitements)
- L'explication des processus, explication du périmètre de la partie non explicable
- La méthode d'apprentissage, d'inférence, etc.
- Les limites de l'utilisation de l'algorithme (FP, FN si classification)
- Les modalités de recours en cas d'erreurs
- L'implication des utilisateurs

Mise en œuvre des mécanismes de garantie humaine (PS, équipe de soins) pour assurer :

- Le contrôle par l'humain de l'IA
- L'autonomie décisionnelle
- Le maintien des compétences des utilisateurs
- L'intervalle de confiance de l'IA / garde-fou des erreurs de l'IA
- Les audits (désaccords IA / PS)

Définition de l'instance de régulation (audit, Label Ethique-IA)

Analyse d'impact organisationnel sur le parcours de soins

Analyse d'impact sociétal et environnemental

Cadrage

Définition de la finalité de la solution d'IA

Validation éthique de la finalité

Caractérisation des principes de gouvernance et de responsabilité

Collecte des données

Mise en œuvre des mesures pour assurer :

- Le consentement éclairé des patients à la réutilisation de leurs données au-delà de la finalité première du recueil (conformité RGPD, article 5 et 7)
- La proportionnalité des données collectées par rapport à la finalité du traitement servant l'élaboration de la solution d'IA (conformité RGPD, article 5, c)
- La non réidentification directe des données (agrégées, pseudonymisées, anonymisées) (conformité RGPD, article 32)
- La qualité des données (lutte contre les biais cognitifs)
- La représentativité de la population d'analyse / population cible / prévention des discriminations (lutte contre les biais de sélection)

Mise en œuvre des mesures de sécurité visant à assurer :

- Le transfert sécurisé des données (source unique, sources multiples, chaînage, intégrité)
- La qualité de l'hébergement des données, serveurs localisés en France / Europe (Cloud privé/public, centralisé/distribué) (conformité HDS)
- La cyber-sécurité à l'état de l'art

Mise en œuvre de mesures pour garantir la non-réutilisation non éthique des données (par ex. en cas de fusion de la société / modifications législatives (pouvant aller jusqu'à la destruction automatique des données))

Pré-traitement des données

Mise en œuvre des mesures de :

- Traitement des données manquantes (réduction des biais)
 - Rééquilibrage des populations minoritaires (réduction des biais)
- Séparation des données (deux jeux totalement distincts, un échantillon pour l'apprentissage et un échantillon pour l'évaluation) et représentativité des deux jeux de données par rapport à la population cible et la finalité du traitement)

Construction de l'algorithme

Choix de l'algorithme d'apprentissage en adéquation avec la finalité

Définition :

- De la politique qualité de l'algorithme
- Des mesures de transparence mises en œuvre
- De la politique de traçabilité de la démarche de construction de l'algorithme
- De la politique d'explicabilité des résultats explicables, du processus d'auditabilité des résultats non explicables

Implémentation des fonctionnalités et mécanismes visant à assurer :

- L'identification et élimination des biais
- La correction des erreurs
- La traçabilité des traitements (rendre les codes sources publics avec protection par dépôt (APP))
- L'adaptabilité
- L'intégration des évolutions réglementaires
- L'intégration des avancées médicales
- La maintenance et le versionning

Définition des indicateurs de dérive du système

Evaluation de l'algorithme et préparation de la mise en production

Mise en œuvre des principes d'évaluation externe :

- Technique (bugs), clinique (gold standard, score de précision)
- De l'utilisabilité (PS, patients, usagers)
- De la non-discrimination/équité
- De la robustesse/reproductibilité

Mise en œuvre des procédures en cas de cyberattaques (analyse d'impact sur la sécurité du système d'IA)

Mise en œuvre des mesures pour assurer l'information (juste et égalitaire) des utilisateurs (PS, patients) relative à :

- La finalité, gouvernance, responsabilité
- L'architecture
- L'origine des données et qualité (légalité de la collecte et des traitements)
- L'explication des processus, explication du périmètre de la partie non explicable
- La méthode d'apprentissage, d'inférence, etc.
- Les limites de l'utilisation de l'algorithme (FP, FN si classification)
- Les modalités de recours en cas d'erreurs
- L'implication des utilisateurs

Mise en œuvre des mécanismes de garantie humaine (PS, équipe de soins) pour assurer :

- Le contrôle par l'humain de l'IA
- L'autonomie décisionnelle
- Le maintien des compétences des utilisateurs
- L'intervalle de confiance de l'IA / garde-fou des erreurs de l'IA
- Les audits (désaccords IA / PS)

Définition de l'instance de régulation (audit, Label Ethique-IA)

Analyse d'impact organisationnel sur le parcours de soins

Analyse d'impact sociétal et environnemental

Cadrage

Définition de la finalité de la solution d'IA

Validation éthique de la finalité

Caractérisation des principes de gouvernance et de responsabilité



Collecte des données

Mise en œuvre des mesures pour assurer :

- Le consentement éclairé des patients à la réutilisation de leurs données au-delà de la finalité première du recueil (conformité RGPD, article 5 et 7)
- La proportionnalité des données collectées par rapport à la finalité du traitement servant l'élaboration de la solution d'IA (conformité RGPD, article 5, c)
- La non réidentification directe des données (agrégées, pseudonymisées, anonymisées) (conformité RGPD, article 32)
- La qualité des données (lutte contre les biais cognitifs)
- La représentativité de la population d'analyse / population cible / prévention des discriminations (lutte contre les biais de sélection)

Mise en œuvre des mesures de sécurité visant à assurer :

- Le transfert sécurisé des données (source unique, sources multiples, chaînage, intégrité)
- La qualité de l'hébergement des données, serveurs localisés en France / Europe (Cloud privé/public, centralisé/distribué) (conformité HDS)
- La cyber-sécurité à l'état de l'art

Mise en œuvre de mesures pour garantir la non-réutilisation non éthique des données (par ex. en cas de fusion de la société / modifications législatives (pouvant aller jusqu'à la destruction automatique des données))

// LeBrief du 15 novembre 2018



En intégrant Streams de DeepMind, Google est accusé de rompre une promesse faite aux patients



Dans un billet de blog, DeepMind (une société appartenant à Alphabet) annonce que l'application médicale Streams rejoint le vaisseau mère Google, avec l'idée de la développer au niveau mondial. Pour ce faire, « l'équipe travaillera au sein de Google ». Pour rappel, DeepMind et Google étaient jusqu'à présent deux entités distinctes au sein du groupe Alphabet.

Au Royaume-Uni, ce rapprochement est perçu comme la trahison d'une promesse faite il y a deux ans par Mustafa Suleyman (cofondateur de DeepMind) quand la société travaillait avec le National Health Service, **comme l'explique The Guardian**.

« DeepMind fonctionne de manière autonome vis-à-vis de Google. Dès le départ, nous avons clairement indiqué que les données des patients ne seraient jamais liées ni associées à des comptes, produits ou services Google », affirmait-il en 2016 **sur Medium**.

Selon un porte-parole de Google repris par nos confrères, il n'y aurait pas de changement : « toutes les données des patients restent sous le strict contrôle de nos partenaires, à qui incombent les décisions relatives à leur utilisation. Ces données restent soumises à des contrôles d'audit et d'accès stricts. Leur traitement dépend de nos contrats et de la législation sur la protection des données. Le passage à Google n'y change rien ».

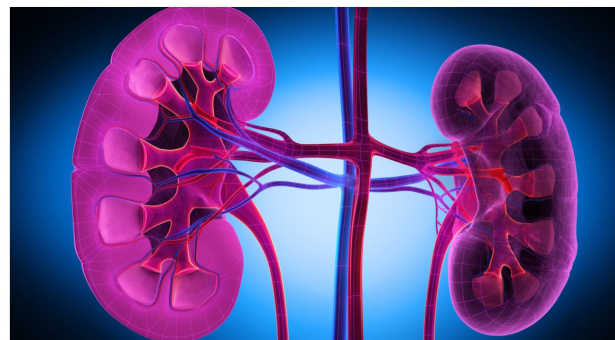
Julia Powles, spécialiste en protection de la vie privée, **n'est pas du tout du même avis**, jugeant ce rapprochement « totalement inacceptable » : « DeepMind a déclaré qu'il ne connecterait jamais Streams à Google. L'ensemble de l'application Streams est désormais un produit Google ! Il s'agit d'un abus de confiance ».

An AI model to predict kidney damage, trained on data from veterans, works less well in women



By Casey Ross | Dec. 5, 2022

Reprints



Cadrage

Définition de la finalité de la solution d'IA

Validation éthique de la finalité

Caractérisation des principes de gouvernance et de responsabilité



Collecte des données

Mise en œuvre des mesures pour assurer :

- Le consentement éclairé des patients à la réutilisation de leurs données au-delà de la finalité première du recueil (conformité RGPD, article 5 et 7)
- La proportionnalité des données collectées par rapport à la finalité du traitement servant l'élaboration de la solution d'IA (conformité RGPD, article 5, c)
- La non réidentification directe des données (agrégées, pseudonymisées, anonymisées) (conformité RGPD, article 32)
- La qualité des données (lutte contre les biais cognitifs)
- La représentativité de la population d'analyse / population cible / prévention des discriminations (lutte contre les biais de sélection)

Mise en œuvre des mesures de sécurité visant à assurer :

- Le transfert sécurisé des données (source unique, sources multiples, chaînage, intégrité)
- La qualité de l'hébergement des données, serveurs localisés en France / Europe (Cloud privé/public, centralisé/distribué) (conformité HDS)
- La cyber-sécurité à l'état de l'art

Mise en œuvre de mesures pour garantir la non-réutilisation non éthique des données (par ex. en cas de fusion de la société / modifications législatives (pouvant aller jusqu'à la destruction automatique des données))

But a **new study** shows how treacherous that journey can be. Researchers found that a replica of the AI system, trained on a predominantly male population of veterans, does not perform nearly as well on women. Their study, published recently in the journal Nature, reports that a model built to approximate Google's AI overestimated the risk for women in certain circumstances and was less accurate in predicting the condition for women overall.

"If we have this problem, then half the population won't benefit," said Jie Cao, a Ph.D. student at the University of Michigan and the lead author of the paper. She said the results reinforce the need to train models on diverse groups of patients and test them on local populations, where demographic differences among patients and variations in how health care is delivered may undermine an AI system's accuracy.

Cadrage

Définition de la finalité de la solution d'IA

Validation éthique de la finalité

Caractérisation des principes de gouvernance et de responsabilité

Collecte des données

Mise en œuvre des mesures pour assurer :

- Le consentement éclairé des patients à la réutilisation de leurs données au-delà de la finalité première du recueil (conformité RGPD, article 5 et 7)
- La proportionnalité des données collectées par rapport à la finalité du traitement servant l'élaboration de la solution d'IA (conformité RGPD, article 5, c)
- La non réidentification directe des données (agrégées, pseudonymisées, anonymisées) (conformité RGPD, article 32)
- La qualité des données (lutte contre les biais cognitifs)
- La représentativité de la population d'analyse / population cible / prévention des discriminations (lutte contre les biais de sélection)

Mise en œuvre des mesures de sécurité visant à assurer :

- Le transfert sécurisé des données (source unique, sources multiples, chaînage, intégrité)
- La qualité de l'hébergement des données, serveurs localisés en France / Europe (Cloud privé/public, centralisé/distribué) (conformité HDS)
- La cyber-sécurité à l'état de l'art

Mise en œuvre de mesures pour garantir la non-réutilisation non éthique des données (par ex. en cas de fusion de la société / modifications législatives (pouvant aller jusqu'à la destruction automatique des données))

Pré-traitement des données

Mise en œuvre des mesures de :

- Traitement des données manquantes (réduction des biais)
 - Rééquilibrage des populations minoritaires (réduction des biais)
- Séparation des données (deux jeux totalement distincts, un échantillon pour l'apprentissage et un échantillon pour l'évaluation) et représentativité des deux jeux de données par rapport à la population cible et la finalité du traitement)

Construction de l'algorithme

Choix de l'algorithme d'apprentissage en adéquation avec la finalité

Définition :

- De la politique qualité de l'algorithme
- Des mesures de transparence mises en œuvre
- De la politique de traçabilité de la démarche de construction de l'algorithme
- De la politique d'explicabilité des résultats explicables, du processus d'auditabilité des résultats non explicables

Implémentation des fonctionnalités et mécanismes visant à assurer :

- L'identification et élimination des biais
- La correction des erreurs
- La traçabilité des traitements (rendre les codes sources publics avec protection par dépôt (APP))
- L'adaptabilité
- L'intégration des évolutions réglementaires
- L'intégration des avancées médicales
- La maintenance et le versionning

Définition des indicateurs de dérive du système

Evaluation de l'algorithme et préparation de la mise en production

Mise en œuvre des principes d'évaluation externe :

- Technique (bugs), clinique (gold standard, score de précision)
- De l'utilisabilité (PS, patients, usagers)
- De la non-discrimination/équité
- De la robustesse/reproductibilité

Mise en œuvre des procédures en cas de cyber-attaques (analyse d'impact sur la sécurité du système d'IA)

Mise en œuvre des mesures pour assurer l'information (juste et égalitaire) des utilisateurs (PS, patients) relative à :

- La finalité, gouvernance, responsabilité
- L'architecture
- L'origine des données et qualité (légalité de la collecte et des traitements)
- L'explication des processus, explication du périmètre de la partie non explicable
- La méthode d'apprentissage, d'inférence, etc.
- Les limites de l'utilisation de l'algorithme (FP, FN si classification)
- Les modalités de recours en cas d'erreurs
- L'implication des utilisateurs

Mise en œuvre des mécanismes de garantie humaine (PS, équipe de soins) pour assurer :

- Le contrôle par l'humain de l'IA
- L'autonomie décisionnelle
- Le maintien des compétences des utilisateurs
- L'intervalle de confiance de l'IA / garde-fou des erreurs de l'IA
- Les audits (désaccords IA / PS)

Définition de l'instance de régulation (audit, Label Ethique-IA)

Analyse d'impact organisationnel sur le parcours de soins

Analyse d'impact sociétal et environnemental

s

Aller plus loin !



Home > OECD AI Principles > Transparency and explainability (Principle 1.3)



Transparency and explainability (Principle 1.3)



This principle is about transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes.



AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:

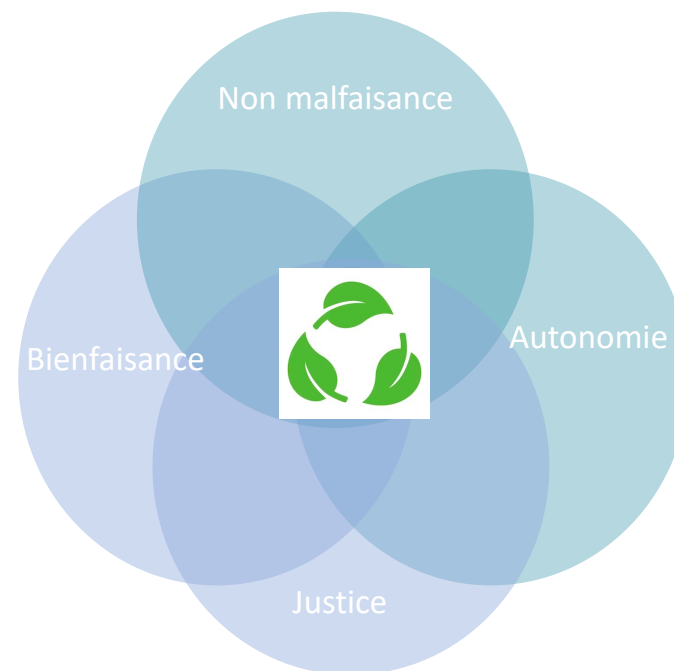
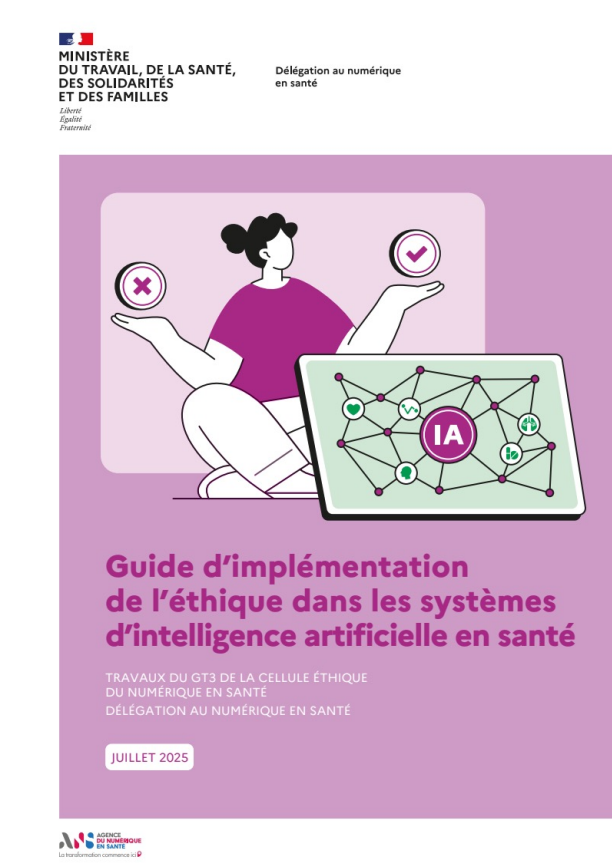
- › *to foster a general understanding of AI systems, including their capabilities and limitations,*
- › *to make stakeholders aware of their interactions with AI systems, including in the workplace,*
- › *where feasible and useful, to provide plain and easy-to-understand information on the sources of data/input, factors, processes and/or logic that led to the prediction, content, recommendation or decision, to enable those affected by an AI system to understand the output, and,*
- › *to provide information that enable those adversely affected by an AI system to challenge its output.*



Les acteurs de l'IA devraient s'engager à faire preuve de transparence et à divulguer de manière responsable les informations concernant les systèmes d'IA. À cette fin, ils devraient fournir des informations pertinentes, adaptées au contexte et cohérentes avec l'état de l'art :

- pour favoriser une **compréhension générale des systèmes d'IA, y compris leurs capacités et leurs limites** ;
- pour informer les parties prenantes de leurs **interactions avec les systèmes d'IA**, y compris sur le lieu de travail ;
- lorsque cela est faisable et utile, pour fournir des informations claires et faciles à comprendre **sur les sources de données/entrées, les facteurs, les processus et/ou la logique** ayant conduit à la prédiction, au contenu, à la recommandation ou à la décision, afin de permettre aux personnes concernées par un système d'IA de **comprendre** le résultat ;
- pour fournir des informations permettant aux personnes affectées négativement par un système d'IA de **contester** son résultat.

Guide d'implémentation



Guide d'implémentation de l'éthique dans les SIA



1. Cadrage

Mettre en place un CSE

Guide d'implémentation de l'éthique dans les SIA

1. Cadrage

Mettre en place un CSE

2. Collecte et préparation des données

Transparence sur l'origine des données
d'apprentissage du SIA

Minimisation des données vs singularité

Minimisation des données vs impact environnemental

Guide d'implémentation de l'éthique dans les SIA

1. Cadrage

Mettre en place un CSE

2. Collecte et préparation des données

Transparence sur l'origine des données
d'apprentissage du SIA
Minimisation des données vs singularité
Minimisation des données vs impact environnemental

3. Conception de l'algorithme du SIA

Réduire et éliminer les biais du SIA
Explicabilité du SIA
Empêcher l'IA générative de restituer des données à caractère personnel utilisées pour son entraînement (G)
Garantir l'absence de plagiat de l'IA générative (G)
Autonomie pour la désactivation d'un SIA modulaire
Ecoconception du SIA lutte contre l'obsolescence des équipements SIA
Bilan carbone de la fabrication du SIA

Guide d'implémentation de l'éthique dans les SIA

1. Cadrage

Mettre en place un CSE

2. Collecte et préparation des données

Transparence sur l'origine des données d'apprentissage du SIA
Minimisation des données vs singularité
Minimisation des données vs impact environnemental

3. Conception de l'algorithme du SIA

Réduire et éliminer les biais du SIA
Explicabilité du SIA
Empêcher l'IA générative de restituer des données à caractère personnel utilisées pour son entraînement (G)
Garantir l'absence de plagiat de l'IA générative (G)
Autonomie pour la désactivation d'un SIA modulaire
Ecoconception du SIA lutte contre l'obsolescence des équipements SIA
Bilan carbone de la fabrication du SIA

4. Conception des interfaces du SIA

Transparence sur l'interaction avec une IA
Vérifier la bonne compréhension de l'interaction avec une IA*
Marquer les résultats de l'IA
Traçabilité de la désactivation du SIA
Garantir l'autonomie décisionnelle de l'utilisateur
Transparence sur le fait de suivre les propositions du SIA
Transparence sur le fait de ne pas suivre les propositions du SIA et sur le motif de non suivi
Transparence sur le fait de ne pas avoir utilisé le SIA
Conformité RGAA des SIA avec une interface utilisateurs
Intelligibilité des interfaces des SIA comportant une interface utilisateurs

Guide d'implémentation de l'éthique dans les SIA

1. Cadrage

Mettre en place un CSE

2. Collecte et préparation des données

Transparence sur l'origine des données d'apprentissage du SIA
Minimisation des données vs singularité
Minimisation des données vs impact environnemental

3. Conception de l'algorithme du SIA

Réduire et éliminer les biais du SIA
Explicabilité du SIA
Empêcher l'IA générative de restituer des données à caractère personnel utilisées pour son entraînement (G)
Garantir l'absence de plagiat de l'IA générative (G)
Autonomie pour la désactivation d'un SIA modulaire
Ecoconception du SIA lutte contre l'obsolescence des équipements SIA
Bilan carbone de la fabrication du SIA

4. Conception des interfaces du SIA

Transparence sur l'interaction avec une IA
Vérifier la bonne compréhension de l'interaction avec une IA*
Marquer les résultats de l'IA
Traçabilité de la désactivation du SIA
Garantir l'autonomie décisionnelle de l'utilisateur
Transparence sur le fait de suivre les propositions du SIA
Transparence sur le fait de ne pas suivre les propositions du SIA et sur le motif de non suivi
Transparence sur le fait de ne pas avoir utilisé le SIA
Conformité RGAA des SIA avec une interface utilisateurs
Intelligibilité des interfaces des SIA comportant une interface utilisateurs

5. Evaluation du SIA

Cohérence des réponses dans la répétition d'un SIA générative(G)
Qualité des réponses d'un SIA générative (G)
Garantir que le SIA générative incite le patient utilisateur à consulter un professionnel de santé en cas d'alerte (G)
Performance du SIA identique pour tous les publics couverts par la population cible

Guide d'implémentation de l'éthique dans les SIA

1. Cadrage

Mettre en place un CSE

2. Collecte et préparation des données

Transparence sur l'origine des données d'apprentissage du SIA
Minimisation des données vs singularité
Minimisation des données vs impact environnemental

3. Conception de l'algorithme du SIA

Réduire et éliminer les biais du SIA
Explicabilité du SIA
Empêcher l'IA générative de restituer des données à caractère personnel utilisées pour son entraînement (G)
Garantir l'absence de plagiat de l'IA générative (G)
Autonomie pour la désactivation d'un SIA modulaire
Ecoconception du SIA lutte contre l'obsolescence des équipements SIA
Bilan carbone de la fabrication du SIA

4. Conception des interfaces du SIA

Transparence sur l'interaction avec une IA
Vérifier la bonne compréhension de l'interaction avec une IA*
Marquer les résultats de l'IA
Traçabilité de la désactivation du SIA
Garantir l'autonomie décisionnelle de l'utilisateur
Transparence sur le fait de suivre les propositions du SIA
Transparence sur le fait de ne pas suivre les propositions du SIA et sur le motif de non suivi
Transparence sur le fait de ne pas avoir utilisé le SIA
Conformité RGAA des SIA avec une interface utilisateurs
Intelligibilité des interfaces des SIA comportant une interface utilisateurs

5. Evaluation du SIA

Cohérence des réponses dans la répétition d'un SIA générative(G)
Qualité des réponses d'un SIA générative (G)
Garantir que le SIA générative incite le patient utilisateur à consulter un professionnel de santé en cas d'alerte (G)
Performance du SIA identique pour tous les publics couverts par la population cible

6. Déploiement, formation et utilisation du SIA

Accompagner les utilisateurs/la structure dans l'évaluation des impacts du déploiement d'un SIA
Co-construire avec les utilisateurs les ressources de formation aux enjeux de l'IA
Former les utilisateurs aux enjeux de l'IA (biais, limites de performance)
Vérifier la bonne compréhension par les utilisateurs des enjeux de l'IA en particulier des limites de performance*
Former à l'utilisation du SIA en amont de son introduction
Permettre un accès facile et intuitif à la documentation du SIA
Transparence sur le type d'IA employé
Transparence sur la performance du SIA
Détecter la dépendance au SIA
Alerter en cas de détection de dépendance au SIA
Garder un esprit critique et faire confiance à son expertise clinique
Rappeler la responsabilité de l'utilisateur sur sa décision
Sensibilisation à un usage écoresponsable
Souveraineté des données de l'organisation utilisatrice (G)

Guide d'implémentation de l'éthique dans les SIA

1. Cadrage

Mettre en place un CSE

2. Collecte et préparation des données

Transparence sur l'origine des données d'apprentissage du SIA
Minimisation des données vs singularité
Minimisation des données vs impact environnemental

3. Conception de l'algorithme du SIA

Réduire et éliminer les biais du SIA
Explicabilité du SIA
Empêcher l'IA générative de restituer des données à caractère personnel utilisées pour son entraînement (G)
Garantir l'absence de plagiat de l'IA générative (G)
Autonomie pour la désactivation d'un SIA modulaire
Ecoconception du SIA lutte contre l'obsolescence des équipements SIA
Bilan carbone de la fabrication du SIA

4. Conception des interfaces du SIA

Transparence sur l'interaction avec une IA
Vérifier la bonne compréhension de l'interaction avec une IA*
Marquer les résultats de l'IA
Traçabilité de la désactivation du SIA
Garantir l'autonomie décisionnelle de l'utilisateur
Transparence sur le fait de suivre les propositions du SIA
Transparence sur le fait de ne pas suivre les propositions du SIA et sur le motif de non suivi
Transparence sur le fait de ne pas avoir utilisé le SIA
Conformité RGAA des SIA avec une interface utilisateurs
Intelligibilité des interfaces des SIA comportant une interface utilisateurs

5. Evaluation du SIA

Cohérence des réponses dans la répétition d'un SIA générative(G)
Qualité des réponses d'un SIA générative (G)
Garantir que le SIA générative incite le patient utilisateur à consulter un professionnel de santé en cas d'alerte (G)
Performance du SIA identique pour tous les publics couverts par la population cible

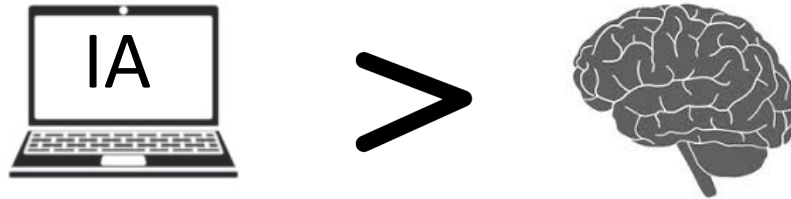
6. Déploiement, formation et utilisation du SIA

Accompagner les utilisateurs/la structure dans l'évaluation des impacts du déploiement d'un SIA
Co-construire avec les utilisateurs les ressources de formation aux enjeux de l'IA
Former les utilisateurs aux enjeux de l'IA (biais, limites de performance)
Vérifier la bonne compréhension par les utilisateurs des enjeux de l'IA en particulier des limites de performance*
Former à l'utilisation du SIA en amont de son introduction
Permettre un accès facile et intuitif à la documentation du SIA
Transparence sur le type d'IA employé
Transparence sur la performance du SIA
Détecter la dépendance au SIA
Alerter en cas de détection de dépendance au SIA
Garder un esprit critique et faire confiance à son expertise clinique
Rappeler la responsabilité de l'utilisateur sur sa décision
Sensibilisation à un usage écoresponsable
Souveraineté des données de l'organisation utilisatrice (G)

7. Suivi, mise à jour et amélioration continue

Evaluer le SIA en continu
Mettre en place une démarche d'amélioration continue de l'IA
Transparence sur la mise à jour d'un SIA et ses conséquences
Vérifier la bonne compréhension des conséquences d'une mise à jour d'un SIA*

Conclusion (1/3)



L'impact sur la capacité à avoir un esprit critique



The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers

Hao-Ping (Hank) Lee
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
haopingl@cs.cmu.edu

Ian Drosos
Microsoft Research
Cambridge, United Kingdom
t-iandrosos@microsoft.com

Advait Sarkar
Microsoft Research
Cambridge, United Kingdom
advait@microsoft.com

Sean Rintel
Microsoft Research
Cambridge, United Kingdom
serintel@microsoft.com

Nicholas Wilson
Microsoft Research
Cambridge, United Kingdom
niwilson@microsoft.com

Lev Tankelevitch
Microsoft Research
Cambridge, United Kingdom
levt@microsoft.com

Richard Banks
Microsoft Research Cambridge
Cambridge, United Kingdom
rbanks@microsoft.com

L'impact sur la capacité à avoir un esprit critique

- Effets à court terme : une pensée moins critique
 - 72 à 79% des utilisateurs déclarent avoir réduit leur effort cognitif dans **des activités telles que l'analyse ou la recherche**. Si l'IA simplifie certaines tâches, **elle diminue aussi l'engagement dans une réflexion profonde** avec une corrélation significative : plus la confiance en l'IA est élevée, moins la réflexion est poussée
 - Conclusion : une dépendance aux réponses toutes faites

The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers

Hao-Ping (Hank) Lee
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
haopingl@cs.cmu.edu

Ian Drosos
Microsoft Research
Cambridge, United Kingdom
t-iandrosos@microsoft.com

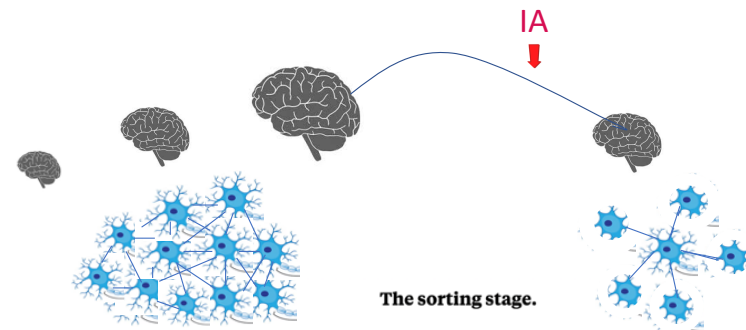
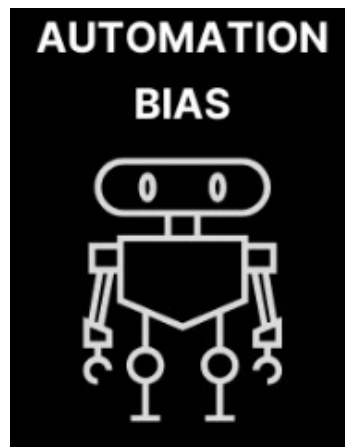
Advait Sarkar
Microsoft Research
Cambridge, United Kingdom
advait@microsoft.com

Sean Rintel
Microsoft Research
Cambridge, United Kingdom
serintel@microsoft.com

Nicholas Wilson
Microsoft Research
Cambridge, United Kingdom
niwilson@microsoft.com

Lev Tankelevitch
Microsoft Research
Cambridge, United Kingdom
levt@microsoft.com

Richard Banks
Microsoft Research Cambridge
Cambridge, United Kingdom
rbanks@microsoft.com



The least used connections, which encode the least often repeated experiences, will gradually become weaker and be eliminated. In contrast, the most often used connections, which encode the experiences which the child has most often, will be reinforced. This is called *synaptic pruning*. But the key point is that when the brain is doing its synaptic pruning, it pays no attention to the nature of what it is erasing: it sorts according to frequency.

<https://www.celinealvarez.org/en/plasticity%E2%80%93for-better-or-worse>

L'impact sur la capacité à avoir un esprit critique



- Effets à court terme : une pensée moins critique

- 72 à 79% des utilisateurs déclarent avoir réduit leur effort cognitif dans **des activités telles que l'analyse ou la recherche**. Si l'IA simplifie certaines tâches, **elle diminue aussi l'engagement dans une réflexion profonde** avec une corrélation significative : plus la confiance en l'IA est élevée, moins la réflexion est poussée
- Conclusion : une dépendance aux réponses toutes faites

- Effets à moyen terme : une dépendance cognitive

- A force de déléguer la résolution de problèmes et la création à l'IA, 62% des participants ont signalé une baisse de leur pensée critique. Le concept « d'atrophie cognitive induite par les chatbots IA » désigne une dépendance croissante qui affaiblit le raisonnement autonome.
- Conclusion : sans pratique régulière, nos compétences analytiques vont s'atrophier

- Effets à long terme : un impact sur le QI

- Une sur-utilisation inévitable va entraîner un déclin cognitif profond, affectant la mémoire (hippocampe), l'attention et la créativité, ce qui risque d'entraîner une baisse globale du QI. Les circuits neuronaux liés à la mémoire ou à la créativité pourraient être sous-stimulés, tandis que ceux dédiés à la vérification des outputs de l'IA (cortex pré-frontal) s'activeraient davantage

The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers

Hao-Ping (Hank) Lee
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
haopingl@cs.cmu.edu

Ian Drosos
Microsoft Research
Cambridge, United Kingdom
t-iandrosos@microsoft.com

Advait Sarkar
Microsoft Research
Cambridge, United Kingdom
advait@microsoft.com

Sean Rintel
Microsoft Research
Cambridge, United Kingdom
serintel@microsoft.com

Nicholas Wilson
Microsoft Research
Cambridge, United Kingdom
niwilson@microsoft.com

Lev Tankelevitch
Microsoft Research
Cambridge, United Kingdom
levt@microsoft.com

Richard Banks
Microsoft Research Cambridge
Cambridge, United Kingdom
rbanks@microsoft.com

Plus récemment ...

L'étude multicentrique a été conduite dans **quatre centres** polonais entre septembre 2021 et mars 2022. Tous ont progressivement adopté le système ENDO-AID (Olympus) de **détection assistée par IA**. La période d'exposition à l'IA a été d'environ **trois mois**. Après cette période, les coloscopies ont à nouveau été réalisées sans l'aide de l'IA. Les chercheurs ont comparé **1443 procédures standards non assistées par IA : 795 avant et 648 après** la période d'utilisation de l'IA en routine. Dix-neuf endoscopistes expérimentés (16 gastro-entérologues et 3 chirurgiens, tous avec plus de 2000 coloscopies à leur actif) ont participé.

Le résultat principal est préoccupant : l'ADR en coloscopie standard **a chuté de manière significative**, passant de 28,4 % avant exposition à l'IA, à 22,4 % après (-6 points absolus ; IC 95 % -10,5 à -1,6 ; $p = 0,009$). Autrement dit, **une fois habitués au « filet de sécurité » algorithmique, les endoscopistes détectaient moins bien les lésions** lorsqu'ils devaient à nouveau compter uniquement sur leurs propres yeux.

■ Assistés par IA, les endoscopistes perdraient en compétence

PAR ELSA BELLANGER - PUBLIÉ LE 13/08/2025

27 RÉACTIONS COMMENTER    

Article réservé aux abonnés

Une étude polonaise publiée dans le « *Lancet Gastroenterology and Hepatology* » est la première en vie réelle à mettre en évidence que l'usage de l'IA en médecine entraîne un recul rapide du savoir-faire des praticiens.

<https://www.lequotidiendumedecin.fr/actu-medicale/recherche-science/assistes-par-ia-les-endoscopistes-perdraient-en-compete#:::text=Une%20C3%A9tude%20polonaise%20publi%C3%A9e%20dans,du%20savoir%20faire%20des%20praticiens.>

THE LANCET
Gastroenterology & Hepatology

 **Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study**

www.thelancet.com/gastrohep Vol 10 October 2025

<https://www.thelancet.com/action/showPdf?pii=S2468-1253%2825%2900133-5>



Et aussi

Digital medicine

Preserving clinical skills in the age of AI assistance



How are clinicians to preserve core clinical skills in an era of algorithmic assistance? As artificial intelligence (AI) assumes a growing role in clinical practice, concern is mounting that off-loading clinical tasks and reasoning will lead to loss of skills (deskilling), adopting errors or bias from AI (mis-skilling), or failure to achieve competence (never-skilling; figure). Evidence for such skill attrition has been seen with automated interpretation of electrocardiograms or radiological images. An observational study published earlier this year, however, sharpens this concern, suggesting that experienced colonoscopists lost some proficiency in colon polyp detection when routine AI support was switched off.

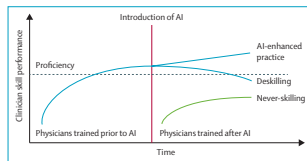


Figure: Potential effects of AI on clinical skills over time

At four Polish endoscopy centres, 19 gastroenterologists and surgeons who used an AI polyp detection system for 3 months showed a decline in adenoma detection rates when subsequently performing colonoscopies without AI. As a group, their post-AI polyp detection performance fell below their own baseline pre-AI performance. Some endoscopists experienced dramatic drops, whereas others maintained their performance. This finding suggests deskilling is not inevitable, and that we must look for approaches to mitigate physician deskilling as AI becomes woven into routine care.

We can learn from other industries that have long grappled with balancing automation and skill retention. In aviation, autopilot systems have brought safety gains, yet also raised concerns about erosion of manual flying skills. Regulators responded by requiring routine hands-on flight time and simulator sessions that mimic system failure. Nuclear plant operators rehearse critical scenarios at regular intervals to ensure that if automation fails, core skills remain intact.

The potential for AI-related deskilling extends well beyond gastrointestinal endoscopy. Surgical and interventional specialties require constant vigilance and readiness to respond to unexpected events. Dermatology, pathology, and radiology demand meticulous visual analysis and pattern recognition. Beyond procedures or image interpretation, clinical reasoning for diagnosis, decision making, and patient management requires critical thinking that could be eroded by AI interactions. Patient encounters are often hurried, leaving little space for deeper analysis and reasoning. Small wonder, then, that physicians would welcome the promise of AI to share the load. But when AI becomes entrenched in clinical practice, new risks emerge. Vigilance may diminish, skills could decay, and the more physicians rely on AI, the less confident we may become in our own, unaided judgement.

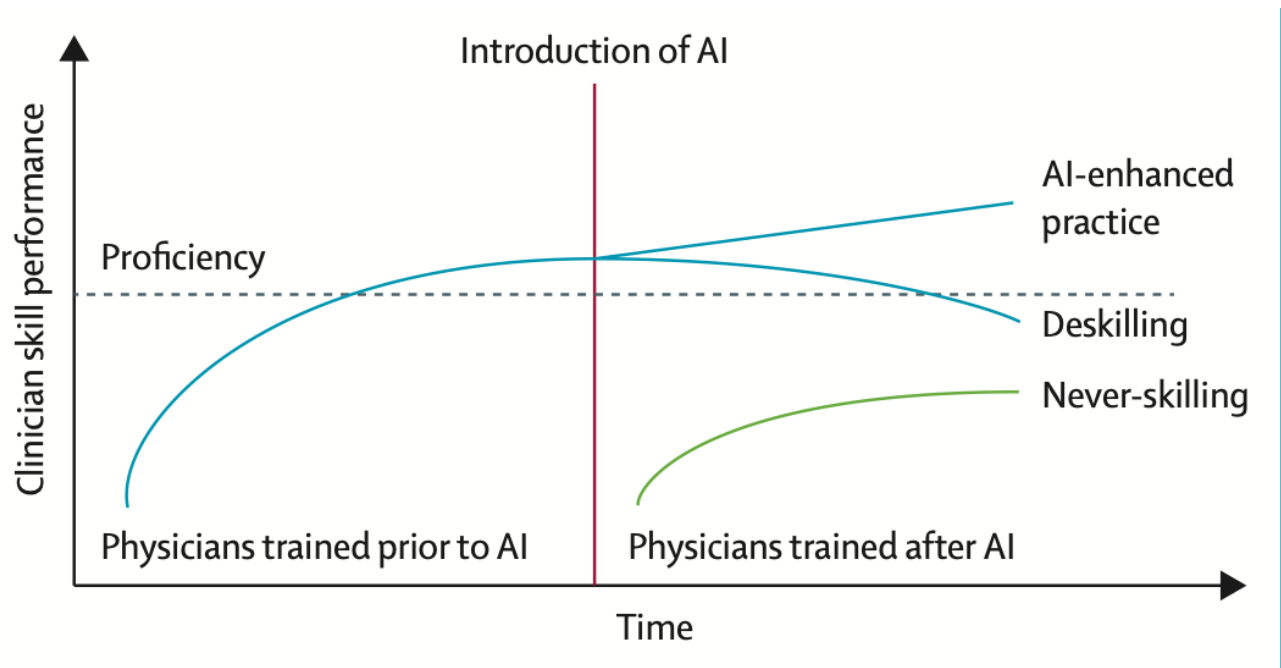
These concerns are especially relevant during medical education, where doctors-in-training may depend on AI before fully developing core competencies (never-skilling), leaving little foundation to draw on if technology fails, and curtailing the formative experiences that allow clinical

intuition and judgement to fully develop. Protecting physician skills and critical thinking in an AI-enabled environment will not have a single, simple solution. The trajectory of deskilling, over-reliance, or loss of vigilance will differ across algorithms, specialties, and contexts, as will potential solutions. One safeguard might be to build short, deliberate "AI-off" or "AI-delay" intervals into clinical workflows, to recalibrate vigilance and capture objective measures of unaided performance. During training or skill-maintenance sessions, AI interpretations could be set to appear only after the reader logs their own findings. Another approach is to draw firmer boundaries: AI takes on rule-based, high-volume, low-ambiguity work, freeing physicians to focus more on contextual, ambiguous, and high-stakes decisions. For instance, in radiology, an AI algorithm might safely exclude normal chest x-rays so that radiologists can concentrate on subtle or complex findings. Such boundaries would help reduce automation bias (over-trusting incorrect AI outputs) and automation neglect (ignoring correct ones), which can arise when tasks are supposedly shared. Whatever the strategy, the foundation must be clinicians' ongoing ability to evaluate the trustworthiness of any AI tool.

The colonoscopy AI deskilling study is not an indictment of AI. Evidence from over 40 randomised trials for AI polyp detection shows it helps physicians detect more precancerous colon polyps. We should welcome such progress, but remain vigilant to how physicians' behaviour and clinical care will change as AI becomes embedded in practice. AI adoption is accelerating, yet workflows and habits are open to shaping. The choices we make now about how we design, integrate, and train around AI will determine whether these systems elevate our profession or quietly erode the skills that define it.

*Tyler M Berzin, Eric J Topol
Center for Advanced Endoscopy, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215, USA (TMB); Scripps Research Translational Institute, La Jolla, CA, USA (EJT)
tberzin@bidmc.harvard.edu

Further reading
Abdulnour E, Gin B, Boscardin C, et al. Educational strategies for clinical supervision of artificial intelligence use. *N Engl J Med* 2025; 393: 766-67.
Budyk K, Romaszko M, Kitala D, et al. Endoscopic deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol* 2025; 10: 896-903.
Gade S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit Med* 2021; 4: 31.
Kostick-Quemet KM, Gerke S. AI in the hands of imperfect users. *NPJ Digit Med* 2022; 5: 197.
TMB reports consulting fees from Boston Scientific, Medtronic, Sanofi, and Beckman Coulter. EJT is supported by the NIH/National Center for Advancing Translational Sciences grant UL1TR001114 and is an advisor to Tempus Labs, Pheno AI, and Abridge.

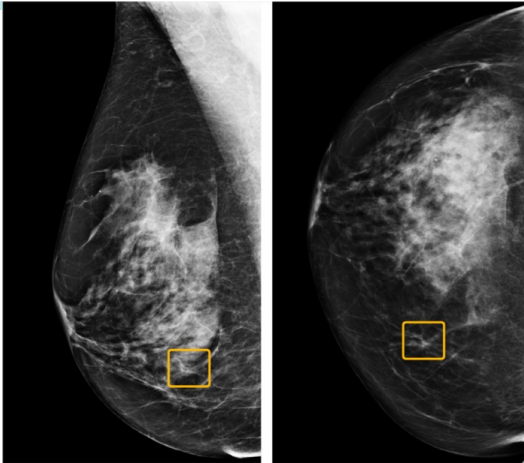


La baisse de compétences : « deskilling »

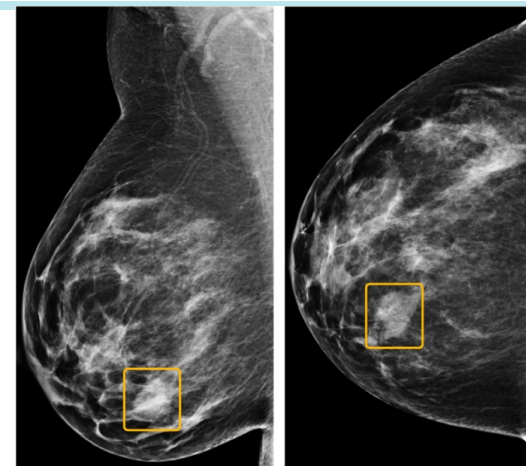
- L'IA, c'est bien pour ce que l'on sait faire, ça fait gagner du temps !
 - Encore faut-il savoir faire !
 - Renforcer l'acquisition de connaissances
 - Permettre l'acquisition de compétences
 - Ex : les services de radiologie qui débranchent l'IA pour les internes au début de leur stage (premiers 3 mois)
- Se donner les moyens d'être TOUJOURS en état de challenger les propositions de l'IA
 - Pour une performance de l'IA de 95%
 - En cas de désaccord entre le jugement clinique et la proposition de l'IA
 - Est-on dans les 95% des cas où l'IA a raison ?
 - Est-on dans les 5% où l'IA a tort ?



L'IA de Google, meilleure que les radiologues

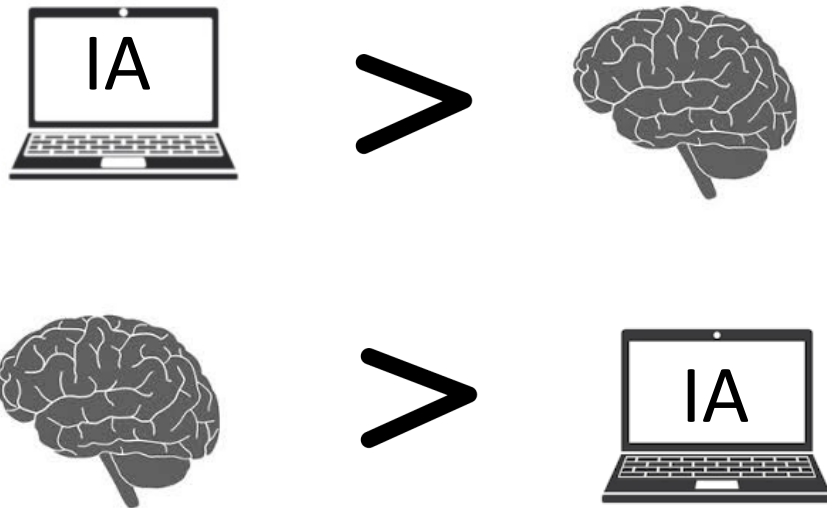


Exemple d'un cancer du sein détecté par aucun des experts humains mais identifié par l'IA. La lésion encadrée en jaune est une petite masse irrégulière associée à des microcalcifications, localisée dans le quadrant inféro-interne du sein droit.

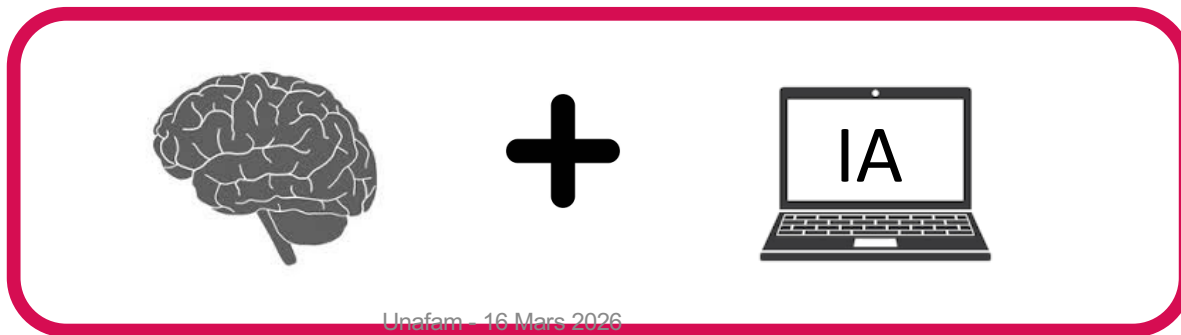
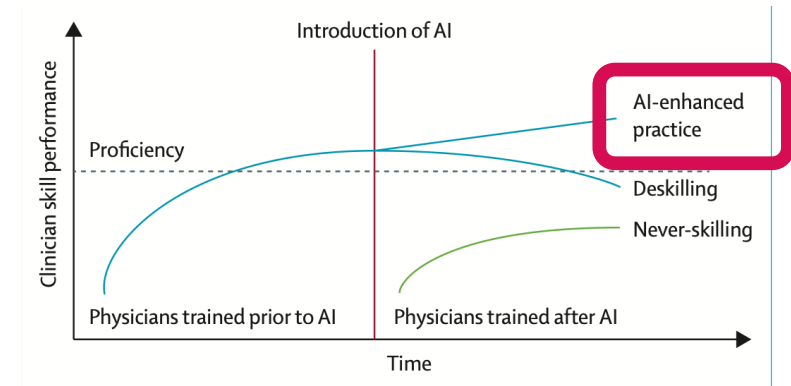
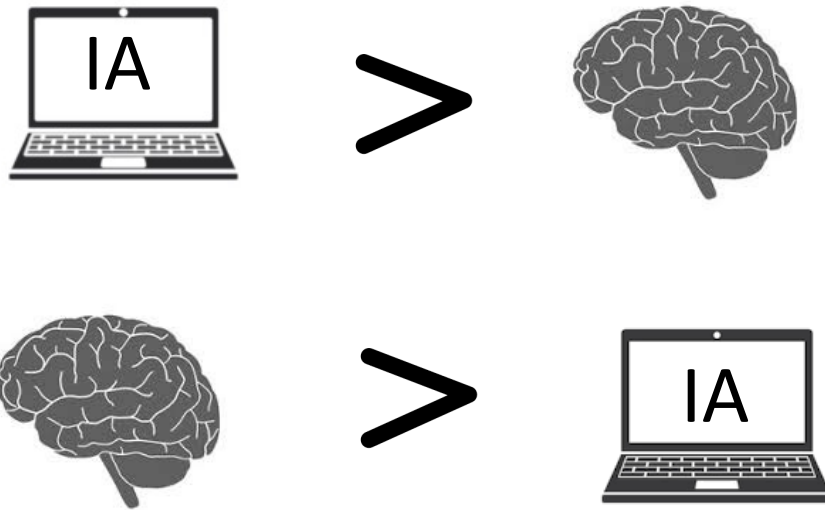


Exemple d'un cancer du sein détecté par les experts humains, mais raté par l'IA. Dans ce cas, la tumeur est une masse dense localisée dans le quadrant inféro-interne du sein droit.

Conclusion (2/3)



Conclusion (3/3)





Merci

brigitte.seroussi@aphp.fr



Demain j'arrête!!!



Demain j'arrête!!!

