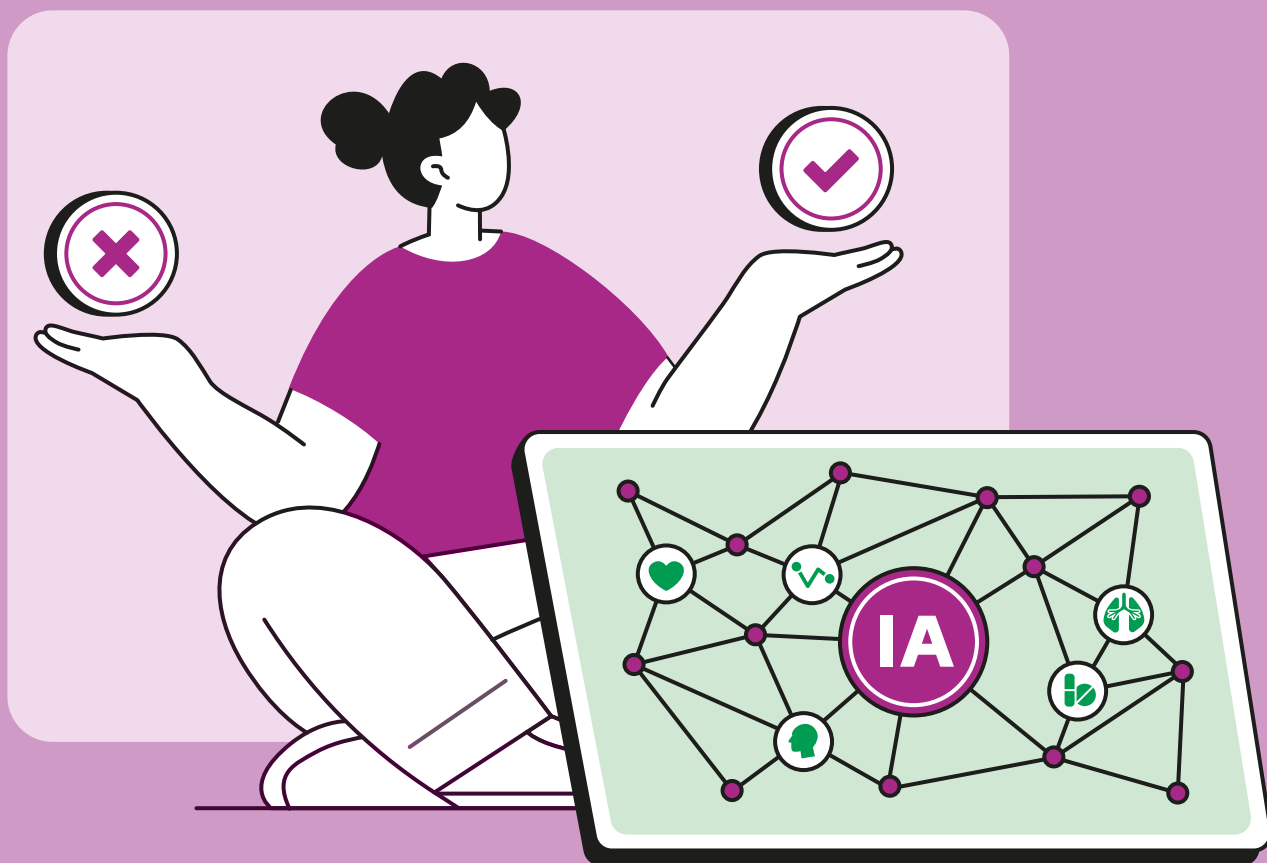




MINISTÈRE
DU TRAVAIL, DE LA SANTÉ,
DES SOLIDARITÉS
ET DES FAMILLES

*Liberté
Égalité
Fraternité*

Délégation au numérique
en santé



Guide d'implémentation de l'éthique dans les systèmes d'intelligence artificielle en santé

TRAVAUX DU GT3 DE LA CELLULE ÉTHIQUE
DU NUMÉRIQUE EN SANTÉ
DÉLÉGATION AU NUMÉRIQUE EN SANTÉ

JUILLET 2025

Sommaire

Éditorial de l'expert	4
Éditorial des pilotes du GT3	6
Membres du groupe de travail GT3	8
1. Introduction	12
1.1 Objet du document	14
1.2 Contexte général	14
1.2.1 La nécessité d'un cadre éthique de l'IA	14
1.2.2 Contexte national	16
1.2.3 Contexte international	18
1.3 Le rôle de la cellule éthique de la Délégation au numérique en santé	20
2. Méthode de travail adoptée par le GT3	22
2.1 Présentation des travaux	22
2.2 Analyse et définition du périmètre	22
2.3 Construction d'une première version V0.1	25
2.4 Comparaison de la V0.1 avec les labels de l'éthique de l'IA	26
2.5 Comparaison de la V0.1 avec le Règlement européen sur l'IA	28
2.6 Finalisation itérative du guide d'implémentation : les versions V0.2, V0.3 et V0.4	30
3. Version finale du guide d'implémentation de l'éthique dans les SIA en santé	32
3.1 Phase de cadrage	32
3.2 Phase de Développement	32
3.3 Phase de Déploiement et Usage	35
4. Conclusion	38

5. Annexe 1 : Comparaison de la V0.1 du guide avec les labels de l'éthique de l'IA	40
5.1 Rappel des différents labels sur l'éthique de l'IA	40
5.2 Méthodologie de l'analyse comparative entre les critères de la V0.1 du guide DNS et les critères des labels	41
5.3 Comparaison des critères de la V0.1 du guide DNS avec les recommandations « Éthique by design »	41
5.4 Comparaison des critères de la V0.1 du guide DNS et ceux du label Labelia	42
5.5 Comparaison des critères de la V0.1 du guide DNS et ceux de Ethical AI	43
5.6 Analyse comparative des labels non disponibles dans l'espace public : Ethik-IA & Ekitia	44
5.6.1 Label Ethik-IA	44
5.6.2 Label Ekitia	46
5.7 Limites de l'analyse	47
6. Annexe 2 : Règlements et normes associés aux critères du guide d'implémentation de l'éthique dans les SIA en santé et propositions d'éléments d'auto-évaluation	48

Éditorial

de l'expert

Le numérique, dont les systèmes dits d'intelligence artificielle (SIA), change profondément nos sociétés et tous les domaines de la santé humaine. Cette transformation remonte au tout début de l'informatisation dès les années 1950, elle s'amplifie au début des années 2000 et s'accélère considérablement depuis 15 ans. Ce domaine encore émergent de la santé numérique nous mets face à des tensions nouvelles entre avancées remarquables et risques à appréhender, promesses et réalisations effectives, temps de la compréhension et urgence de solution.

Quand il s'agit de réfléchir précisément aux enjeux éthiques de la santé numérique pour mettre au point un dispositif médical qui s'appuie sur ces SIA, la situation est difficile ; que ce soit pour un professionnel de santé, la direction d'un hôpital, une entreprise du domaine ou pour ses salariés. En effet comment gérer ces tensions, comment concevoir et réaliser, comment mettre en œuvre et utiliser, comment financer et décider, comment expliquer et qualifier, comment établir la confiance dans ces innovations ? Bien sûr en s'appuyant sur les réflexions éthiques développées de multiple manières au niveau d'entreprises, d'organisations nationales ou internationales. Mais ces réflexions, telles que celles développées par le Comité

national pilote d'éthique du numérique, le Comité consultatif national d'éthique ou l'OMS, parmi bien d'autres, nécessitent au-delà de leur indispensable réflexion générale sur le sujet, de comprendre de manière opérationnelle, comment en pratique mettre en œuvre ces recommandations, ces schémas de réflexion en partageant le plus objectivement possible les approches recommandées et leurs difficultés de mise en application. On passe alors d'une réflexion générale à des recommandations pratiques partagées par tous les acteurs de la santé numérique. C'est précisément l'objet de ce guide qui vient éclairer la scène française de manière remarquable.

Si cette réflexion opérationnelle est indispensable, elle est aussi redoutablement difficile. L'implémentation, résultat de l'action d'implémenter, anglicisme largement utilisé par les professionnels de l'informatique, est décrit dans le JO du 20 avril 2007 comme « Effectuer l'ensemble des opérations qui permettent de définir un projet et de le réaliser, de l'analyse du besoin à l'installation et la mise en service du système ou du produit. ». C'est difficile ici car le sujet est au croisement de trois domaines majeurs, la santé, le numérique incluant les systèmes d'intelligence artificielle et l'éthique. Tous trois ont des cultures, des pratiques et des

implémentations jusqu'à présent disjointes, mais dont la convergence est maintenant essentielle.

Ce guide, avec une méthodologie clairement explicitée et l'association de toutes les parties prenantes, aide à s'entourer des réflexions nécessaires à la mise en œuvre d'une démarche éthique opérationnelle clairement énoncée. Plutôt que de proposer un nouveau règlement ou une nouvelle charte, il s'inspire et se compare aux nombreuses chartes et avis publiés. Il permet ainsi aux différentes parties prenantes pour l'élaboration d'un SIA en santé de comprendre et d'élaborer sa raison d'être, d'aider à comprendre et à prendre en compte les autres parties prenantes, de rendre intelligents tous les acteurs humains impliqués dans son élaboration. Il nous permet de réfléchir aux conduites humaines et aux valeurs qui les fondent en vue d'élaborer des dispositifs médicaux de qualité fondant la confiance dans leurs utilisations. Ce guide sert alors de référence, à s'approprier et à partager dans une dynamique qui devra prendre en compte les évolutions qui enrichissent et défient tous les jours les domaines de la santé.



Claude Kirchner
Président du Comité
Consultatif National
d'Éthique du Numérique





Editorial

des pilotes du GT3

À chaque avancée technique, la médecine s'est réinventée. Le stéthoscope, l'imagerie, la génétique ont ouvert des mondes nouveaux. L'intelligence artificielle ne fait pas exception. Mais son irruption marque un tournant : pour la première fois, des machines prétendent raisonner, diagnostiquer, décider. Le questionnement change alors de nature, il ne s'agit plus seulement de se demander ce qu'on peut faire de cette nouvelle technologie, mais ce qu'on doit en faire.

Car en santé, l'enjeu n'est pas seulement de gagner en rapidité. Il est de préserver ce qui fait que soigner n'est pas seulement traiter. C'est précisément pourquoi la vigilance éthique ne peut être facultative. Car les risques sont réels : biais cachés dans les algorithmes, dilution des responsabilités, perte d'autonomie professionnelle, standardisation des décisions. Et surtout, cette tentation du soulagement cognitif : laisser la machine décider, pour aller plus vite, pour avoir moins peur de se tromper. Mais que reste-t-il du soin, lorsque penser devient optionnel ?

Le progrès technologique est une promesse. Il devient menace quand il n'est plus interrogé. L'éthique, ici, n'est pas un frein, mais une respiration. Elle nous oblige à nous souvenir que la santé n'est pas un champ d'expérimentation technique parmi d'autres. C'est un lieu de confiance, de fragilité, de responsabilité partagée. Dans ce lieu, toute innovation doit d'abord se plier à une exigence : celle de servir l'humain, non de s'y substituer.

C'est pourquoi il est urgent de structurer une intelligence artificielle à la hauteur de nos valeurs du soin. Une IA qui soutient sans supplanter, qui alerte sans dicter, qui aide sans désengager.

Brigitte SEROUSSI

Directrice de projets
Responsable de la Cellule
Éthique
Délégation au numérique
en santé



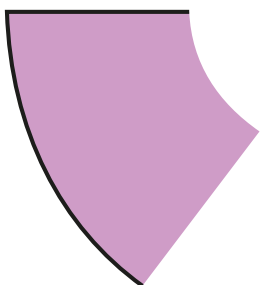
David Gruson

Directeur de Programme
Santé à Domicile
La Poste Santé
& Autonomie
Fondateur ETHIK-IA

Pour cela, l'éthique n'est pas une option : elle ne freine pas l'innovation mais lui donne un sens, un cap, une légitimité durable.

Ce guide d'implémentation de l'éthique d'un système d'IA en santé ne donne pas des leçons de morale. Il propose des repères, des outils, une méthode pour intégrer l'éthique au cœur même de la conception, du déploiement et de l'évaluation des systèmes d'IA en santé. Il permet d'interroger les finalités avant les fonctionnalités. Il trace une voie pour que la puissance des algorithmes ne marginalise pas l'intelligence humaine, mais l'amplifie. Il assume que dans un monde numérique, la décision doit rester profondément humaine. Le travail collectif ayant conduit à ce guide s'est d'ailleurs prolongé dans le Règlement européen sur l'IA – l'AI Act qui est entré en vigueur dès le 1er août 2024 – qui a introduit le principe d'une supervision humaine des systèmes d'IA en santé.

Nous avons aujourd'hui l'opportunité de faire de l'IA en santé un levier de justesse, d'équité et de progrès. À condition de se poser la question essentielle : que veut-on faire de cette puissance nouvelle, et au nom de quelle idée du soin voulons-nous en guider l'usage ?



Membres

du groupe de travail GT3

Liste des contributeurs aux travaux du GT3 ayant permis l'élaboration du guide d'implémentation de l'éthique dans les systèmes d'Intelligence artificielle en santé.

NOM	PRÉNOM	TITRE	ORGANISATION
ABAD	Sébastien	Praticien Hospitalier CHU ROUEN	NAULIMUS
ALLOGGIA	Marie	Juriste	MALAKOFF HUMANIS
AMRANI	Sarah	Juriste IA	ETHIK IA
AUGRIS	Arnaud	Responsable affaires réglementaires	SNITEM
AVRIN	Guillaume	Coordonnateur national pour l'intelligence artificielle	DGE
BAROT	Charlotte	Analyste IA	CNIL
BEN OTHMAN	Atef	Délégué aux usages numériques en IA	NUMEUM
BERANGER	Jérôme	PDG – Expert en éthique du numérique	GOODALGO
BERGERON	Christine	Présidente	CERBAPATH
BOURNONNEUX	Agathe	Juriste	DNS
BRAC DE LA PERRIERE	Marguerite	Avocate	FIELDFISHER
CASTELS	Bernard	PDG	HORIZON 2 INNOVATION & TRANSFORMATION
CATALANO	Lydie	PDG & Co-Fondatrice	IA MEDICAL
CIMINO	Mariane	Déléguée numérique en santé	NUMEUM
CLAEYS	Sébastien	Professeur associé	Conseil national consultatif des personnes handicapées

NOM	PRÉNOM	TITRE	ORGANISATION
CLOUT	Emmanuel	Directeur de projets	DNS
COLLIGNON	Corinne	Cheffe de service Mission numérique en santé	HAS
COLLOC	Joël	Professeur Émérite d'Informatique – Université du Havre	PU – UNIVERSITÉ DU HAVRE
DAMEZ-FONTAINE	Marc	Directeur Data Analytics	PWC
DAUPHIN	Arthur	Chargé de Mission	FRANCE ASSOS SANTE
DEGOS	Sandrine	Présidente	CARE INSIGHT
DENIS	Marielle	Juriste	MALAKOFF HUMANIS
DIDIER	Emmanuel	Sociologue	CNPEN
ETIENNE	Loïc	Président	MEDVIR
FOEX	Olivier	Responsable Éthique & Conformité	DOCTOLIB
FRAYSSE	Jean-Louis	Co-fondateur	BOTDESIGN
FUCHS	Alexandra	Responsable de programmes Technologies de Santé	CEA
GOGLIN	Jean-François	DGA	CONNECTIVE SANTE
GRUSON	David	Co-pilote du GT3 – Fondateur ETHIK-IA – Directeur de Programme Santé à Domicile La Poste – Santé Autonomie	ETHIK-IA
GUILLOT	Caroline	Directrice citoyenne	HEALTH DATA HUB
KIRCHNER	Claude	Président	CNPEN – CCNE Numérique
LAINE	Katya	Présidente de la Commission IA	NUMEUM
LE GUILLOU	Frédéric	Pneumologue	ASSOCIATION SANTE RESPIRATOIRE FRANCE
LEFEVRE	Xavier	PDG et Fondateur	FAIR&SMART

NOM	PRÉNOM	TITRE	ORGANISATION
LETHEUX	Corinne	Médecin	PRESANSE
LOLLCHAND	Kunal	Responsable Développement et Innovation Santé et Action Sociale	AFNOR
LOURO	Nelsa	Responsable de mission	ANS
MARIAUX	Charles	Délégué E-Santé - Secteur Public	NUMEUM
MARY	Vincent	Consultant en SI	SIB / ASINHPA
MENAGER	Kenann	Chef de projet numérique en santé	HEALTH DATA HUB
MORILHAT	Louis	Chargé de Mission	AFNOR
MOTTE DEJEAN	Clélia	Sage-femme	NAULIMUS
PARROT	Daniella	DPO	MINISTERE DE LA SANTE
RENNER	Simon	Chef de projet	HAS
ROUANET	Flavie	Chargée de Mission	ETHIK IA
SEROUSSI	Brigitte	Co-pilote du GT3 - Directrice de Projets Délégation au numérique en santé Responsable de la Cellule Éthique	DNS
SINEGRE	Geoffroy	Juriste – Enseignant en droit numérique	HUB France IA
STREIFF	Philippe	Délégué général	POSITIVE AI
THEARD - JALLU	Cécile	Avocate	DE GAULLE FLEURANCE
THEBAUT	Jean-François	Vice-Président	FEDERATION FRANÇAISE DES DIABETIQUES
THERON	Patrick	Médecin	CNOM
THOMASSIN - NAGGARA	Isabelle	Professeur de radiologie	APHP - TENON

NOM	PRÉNOM	TITRE	ORGANISATION
TOPALLIAN	Stéphane	Responsable de projet	DGD SOLIDARITES DU DEPARTEMENT DES YVELINES
TORTEL	Alexandre	Directeur de mission	MUTUALITE FRANÇAISE
VAROQUAUX	Gaël	Chercheur	INRIA
VAUGELADE	Cécile	Directeur Affaires Technico-Réglementaires	SNITEM
WILLIATTE	Lina	Professeur de droit de la santé numérique – Experte éthique à la délégation au numérique en santé	SPE WT AVOCAT – UNIVERSIE CATHOLIQUE DE LILLE
WILTZ	Pierre	Référent Sciences Humaines et Sociales	CEA
ZAMBROWSKI	Jean-Jacques	Consultant	NEXTEP HEALTH

Synthèse

L'intelligence artificielle (IA) connaît une expansion rapide dans le secteur de la santé, apportant des solutions innovantes, notamment en matière de diagnostic et de personnalisation des soins. Toutefois, cette évolution technologique soulève des préoccupations éthiques majeures : risques de biais discriminatoires, atteinte à la vie privée, opacité algorithmique, dépendance des utilisateurs à la technologie, ou encore impact environnemental. Face à ces enjeux, la Délégation au numérique en santé (DNS), à travers sa cellule éthique et le groupe de travail GT3, a élaboré un guide d'implémentation de l'éthique d'un système d'intelligence artificielle (SIA) en santé. Ce guide vise à accompagner les fournisseurs de SIA dans le respect des valeurs fondamentales de l'éthique du soin, tout en favorisant une innovation responsable.



Le processus de construction du guide s'est appuyé sur une méthodologie collaborative et itérative entre 2023 et 2025, incluant des ateliers d'experts, des analyses comparatives avec plusieurs labels de l'éthique de l'IA (Labelia, Ethical AI, Ethik-IA, Ekitia), une confrontation avec les exigences du Règlement européen sur l'IA, ainsi qu'une phase de concertation auprès des acteurs de l'écosystème, et plus largement par la mise en œuvre d'une concertation publique (fin mai – début juin 2025). Le résultat est un guide composé de 43 critères classés selon les étapes chronologiques de développement d'un SIA : cadrage, collecte des données, construction de l'algorithme du SIA, développement des interfaces, évaluation du SIA, déploiement et formation, suivi et amélioration continue. À chaque critère est associé un principe éthique fondamental du Cadre de l'éthique du numérique en santé (CENS) : bienfaisance, non-malfaisance, autonomie, justice/équité et écoresponsabilité. Les références juridiques sont associées à chaque critère pour lequel elles existent. Des éléments d'autoévaluation permettent à chaque fournisseur de mesurer sa conformité au guide et de s'inscrire dans une démarche d'amélioration continue.

Parmi les principales exigences du guide, on retrouve l'obligation de constituer un comité scientifique et éthique représentatif dès la phase de cadrage, afin d'assurer une validation du projet en continu. La collecte des données doit être rigoureuse, limitée aux stricts besoins du SIA, en tenant compte à la fois de la singularité des patients et de la réduction de l'empreinte environnementale. La conception des algorithmes doit viser la réduction des biais, l'explicabilité des décisions, la traçabilité des usages et la possibilité, pour l'utilisateur, d'être autonome pour désactiver une IA modulaire. Une attention particulière est portée à l'IA générative, avec des mesures contre le plagiat, la restitution de données à caractère personnel et la production d'hallucinations.

Le guide insiste également sur l'importance de l'élaboration de ressources pédagogiques en co-construction avec les utilisateurs finaux visant à assurer une bonne compréhension des utilisateurs des enjeux, des limites de l'IA et de ses implications. La dépendance aux SIA doit être anticipée et surveillée, notamment via des mécanismes d'alerte et des rappels de la responsabilité humaine dans toute prise de décision. L'évaluation régulière des performances, l'intégration d'indicateurs d'impact environnemental, la transparence des mises à jour et la souveraineté des données constituent les piliers du suivi des SIA.

En proposant des préconisations à l'attention des fournisseurs de SIA pour les engager vers la mise en œuvre d'une IA en santé éthique, ce guide constitue un outil de référence ayant vocation à être décliné dans des référentiels sectoriels, comme cela a déjà été le cas pour la téléconsultation, par exemple. Il se distingue par son approche pragmatique, sa lisibilité et sa prise en compte des réalités du terrain. Il cherche à équilibrer les impératifs éthiques avec les contraintes techniques et économiques des acteurs du numérique en santé. En inscrivant les pratiques d'IA dans un cadre de valeurs partagées, il ambitionne de renforcer la confiance des professionnels de santé, des patients et de la société dans son ensemble à l'égard de ces technologies, et de garantir que le soin, dans sa dimension humaine, reste au cœur du numérique en santé.

Introduction

1.1 Objet du document

L'objectif de ce document est de présenter la démarche développée par le GT3 de la cellule éthique de la Délégation au numérique en santé (DNS) dans le but de construire un guide d'implémentation de l'éthique d'un système d'intelligence artificielle (SIA) en santé. Il s'agit également de mettre en lumière la cohérence de cette réflexion avec les travaux antérieurs de la DNS sur l'élaboration d'un guide de recommandations de bonnes pratiques concrétisant une approche éthique by design des solutions d'IA en santé, la réglementation internationale (Règlement européen sur l'Intelligence Artificielle ou RIA, et le Règlement européen relatif aux dispositifs médicaux ou MDR), ainsi que les réflexions autour de l'éthique de l'IA portées par les labels existants au niveau national.

1.2 Contexte général

La nécessité d'un cadre éthique de l'IA

Promouvoir l'innovation en IA est essentiel pour relever les grands défis de la santé. Toutefois, cette dynamique ne peut s'affranchir d'un cadre éthique. En effet, l'IA en santé peut présenter des dérives potentielles que l'adoption d'un cadre éthique permet de prévenir : la reproduction ou l'amplification des biais présents dans les données d'entraînement peuvent conduire à des propositions de décision discriminatoires, aggravant les inégalités sociales et territoriales de santé ; la mauvaise utilisation des données à caractère personnel et sensibles des patients peut être à l'origine d'atteintes à la vie privée ; l'opacité des processus et le fonctionnement en « boîte noire » de certains modèles d'IA, ne permet pas l'explicabilité et la compréhension des propositions de l'IA mettant à mal la confiance des professionnels et des patients ; des modèles insuffisamment validés ou mal adaptés au contexte clinique peuvent produire des diagnostics erronés ou des traitements inappropriés ; une automatisation excessive, au détriment des interactions humaines, peut menacer l'écoute empathique avec un risque de déshumanisation de la relation de soins.



Face à ces enjeux, la mise en œuvre d'un cadre éthique est une exigence indispensable :

Pour garantir les fondements de l'éthique du soin par le respect des principes qui la composent :

le principe de bienfaisance, de non-malfaisance, de justice/équité, et le respect de l'autonomie.

Pour garantir la protection des données

à caractère personnel en garantissant que le traitement est strictement limité aux finalités déclarées, et que le consentement éclairé du patient a été obtenu.

Pour garantir que les biais ont été identifiés et corrigés

et que la représentativité de la population cible dans les jeux d'apprentissage est assurée afin de promouvoir l'équité.

Pour garantir la fiabilité et ainsi la confiance des utilisateurs :

exiger des normes de sécurité élevées, une validation clinique rigoureuse, une traçabilité des décisions et une mise à jour en continu des algorithmes.

Pour garantir l'utilisabilité des SIA en santé :

développement de modèles explicables pour les utilisateurs, accompagnement à l'usage et articulation avec les processus de surveillance humaine (human oversight).

Ces dernières années, des travaux ont été réalisés aux niveaux national, européen et international afin de définir l'éthique de l'IA. La communauté scientifique, les pouvoirs publics, les industriels, les entrepreneurs et les organisations issues de la société civile se sont mobilisés pour construire une IA digne de confiance au travers de l'élaboration et de la mise en œuvre de réglementations ([Règlement général pour la protection des données ou RGPD](#)), ainsi que le Règlement européen pour l'IA ou [RIA](#)), de référentiels éthiques ([lignes directrices du groupe d'experts de haut niveau de la Commission européenne](#), [les recommandations de l'UNESCO](#)), de labels et de certifications.

Contexte national

En Avril 2022, la Délégation au numérique en santé a émis des recommandations de bonnes pratiques visant à intégrer l'éthique dès les premières étapes du développement des systèmes d'IA en santé au travers d'un [guide de mise en œuvre de « l'éthique by design »](#). Ce guide expose la démarche méthodologique sur lequel il a été créé et les questionnements éthiques posés à chaque étape de construction d'un système d'IA en santé : (i) le cadrage, (ii) la collecte des données, (iii) le pré-traitement des données, (iv) la construction de l'algorithme, (v) l'évaluation de l'algorithme en amont de la mise en production de la solution numérique. Les recommandations ont ainsi pour objectif d'orienter les organisations dans la définition d'une démarche de cadrage permettant de définir la finalité du système d'IA, la définition des principes de gouvernance et les rôles et les responsabilités des différentes parties prenantes.

Par ailleurs, il existe des initiatives issues des secteurs public et privé qui proposent une labellisation éthique de l'IA dont l'objectif est de garantir des pratiques responsables et transparentes. Ces labels sont décernés aux organisations et aux produits qui répondent à des critères éthiques spécifiques.

- [Le label « Labelia »](#) lancé par Labelia Labs en 2019, gratuit et en libre accès, évalue la maturité des organisations en matière de pratiques responsables et éthiques en IA. Le label repose sur six sections : (i) protéger les données personnelles ou confidentielles, (ii) prévenir les biais et élaborer des modèles non discriminatoires, (iii) évaluer la performance de manière rigoureuse, (iv) assurer la reproductibilité des modèles et en établir la chaîne de responsabilité, (v) utiliser des modèles en confiance et de manière responsable, (vi) anticiper, suivre et minimiser les externalités négatives de l'activité de data science.
- [Le label GEEIS IA](#) (Gender Equality European & International Standard for AI), lancé en septembre 2020 par le fond de dotation Arborus, vise à promouvoir une IA inclusive et responsable par la valorisation des pratiques des entreprises en matière d'égalité professionnelle. Il repose sur un audit de critères lié aux ressources humaines, la diversité et la déontologie, et sensibilise l'ensemble de la chaîne de valeur à un développement responsable en IA.

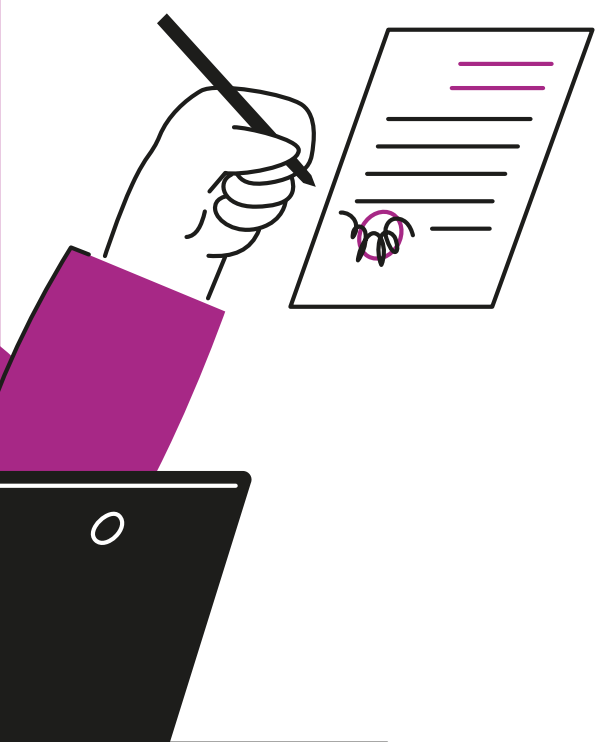
- [Le guide Ethical AI](#), lancé par NUMEUM en 2020, propose 117 recommandations pour opérationnaliser les principes éthiques sur l'ensemble du cycle de vie des systèmes d'IA pour une IA respectueuse, sûre, équitable, maîtrisée, transparente.
- [Le label Positive IA](#), lancé en novembre 2022 par l'association Positive AI (fondée par le BCG GAMMA, l'Oréal, Malakoff Humanis et Orange France), repose sur six principes de l'IA responsables définis par la Commission européenne, recouvrant des dimensions telles que la justice, la transparence, la supervision humaine, l'impact social et environnemental, la responsabilité, la protection de la vie privée et la robustesse technique. Le processus de labellisation, ouvert aux membres de l'association, repose sur un audit externe et dépasse les seuls aspects algorithmiques pour inclure l'ensemble des pratiques de gouvernance liées à l'IA.
- [Le label EKITIA](#), lancé en octobre 2022 par Ekitia (ex-Occitanie Data), valorise les projets ou solutions respectueux qui s'engagent à respecter la Charte Éthique des Usages des Données, actualisée en 2023 avec la contribution d'experts sectoriels et de citoyens. Le processus de labellisation comprend une évaluation d'éligibilité, suivie d'un questionnaire déclaratif (environ 100 questions) dont les réponses sont vérifiées par un évaluateur indépendant à l'aide de justificatifs.



- Le label d’Ethik IA, lancé en février 2023 en collaboration avec le Digital Medical Hub (DMH) vise à garantir la conformité en Garantie Humaine de l’IA. Il permet d’évaluer l’éthique et la conformité juridique des solutions d’IA selon des critères pré-définis. Cette labellisation s’inscrit dans une démarche de normalisation menée avec l’AFNOR.
- Le label ADEL-AI act, lancé en décembre 2023 par GoodAlgo, une entreprise spécialisée dans l’IA responsable, la science des données et l’éthique numérique, vise à certifier les solutions d’IA pour leurs qualités éthiques et leur conformité aux exigences du RIA. Une première auto-évaluation complétée par un audit externe permet de clôturer le processus de labellisation.

Par ailleurs, en 2023, le Comité consultatif national d’éthique (CCNE) et le Comité National Pilote d’Ethique du Numérique (CNPEN) ont publié un avis commun « Diagnostic médical et intelligence artificielle : Enjeux éthiques », en réponse à une saisine par le Premier ministre sur les enjeux d’éthique de l’utilisation de l’intelligence artificielle dans le champ du diagnostic médical, exprimée dans sa lettre de mission du 15 juillet 2019. Le document met en avant l’importance de remettre la garantie humaine au cœur de l’usage de l’IA afin de s’assurer du maintien d’un regard critique de la part des professionnels. Il met également en garde contre les dérives possibles de l’utilisation de l’IA, les risques liés au mésusage des données et au niveau de maturité hétérogène des patients face au numérique. De plus il souligne le manque d’étude permettant d’évaluer le rapport bénéfice-risque de l’IA, des outils et services qui ne sont pas candidats au

remboursement par l’assurance maladie, mais qui représentent un nombre important d’outils.



Contexte international

Au niveau international, plusieurs référentiels et réglementations ont été élaborés pour encadrer l'IA par l'éthique. Ces initiatives montrent un effort concerté pour créer un cadre éthique global portant sur des valeurs communes.

Les principes de l'OCDE sur l'IA, adoptés en mai 2019, constitue la première norme intergouvernementale sur l'IA. L'objectif est de promouvoir l'utilisation d'une IA innovante et digne de confiance, qui respecte les droits de l'Homme et les valeurs démocratiques. Ces principes sont généraux et ne concernent pas spécifiquement l'utilisation de l'IA en santé. Ainsi sont ciblés les impératifs de croissance inclusive, le développement durable et le bien-être ; le respect de l'état de droit, des droits humains et des valeurs démocratiques, y compris l'équité et la vie privée ; la transparence et l'explicabilité ; la robustesse, la sûreté et la sécurité ; la responsabilité.

Après une période de transition, le Medical Device Regulation (MDR), règlement de l'UE entré en application en 2021, encadre la mise sur le marché, la mise en service et l'utilisation de dispositifs médicaux (DM), y compris ceux qui mobilisent l'IA. Il met l'accent sur la sécurité des patients, la précision des diagnostics et l'efficacité des traitements. Afin de garantir un niveau élevé de sécurité et de performance des systèmes d'IA, le MDR impose plusieurs exigences parmi lesquelles, (i) l'évaluation des risques associés à l'utilisation de l'IA, (ii) la mise à disposition d'une documentation technique détaillée, (iii) l'application de mesures appropriées pour assurer une surveillance continue des DM après leur mise sur marché afin de détecter et corriger d'éventuels problèmes de sécurité ou de performance, (iv) le développement de systèmes d'IA transparents et explicables, permettant aux utilisateurs de comprendre comment les propositions du système d'IA ont été élaborées, (v) des systèmes d'IA robustes, sécurisés, précis, fiables et reproductibles afin de minimiser les risques, et (vi) le marquage CE des DM indiquant la conformité des DM aux exigences du MDR et de l'Union Européenne.

Les recommandations sur l'éthique de l'IA de l'UNESCO, adoptées en novembre 2021 par 193 pays, visent à définir des valeurs et principes communs pour assurer le développement d'une IA transparente, responsable, qui protège les données à caractère personnel. Ces recommandations ont pour objectif de protéger et promouvoir les droits de l'homme, la dignité humaine et les libertés fondamentales en insistant sur l'importance de l'équité, de la non-discrimination, et de l'inclusion, afin de prévenir les biais et les inégalités que l'IA pourrait exacerber. Elles proposent la mise en place d'une gouvernance éthique pour maximiser les avantages de l'IA tout en minimisant les risques et soulignent la nécessité de la transparence et de l'explicabilité des algorithmes et des données utilisées en encourageant l'utilisation de l'IA de manière durable et dans le respect de l'environnement. Enfin ces recommandations mettent l'accent sur l'interdiction de l'utilisation des systèmes dotés d'une IA pour la surveillance de masse et le scoring social (système qui attribue une note ou un score à des individus en fonction de leur comportement, de leurs interactions et de leurs activités, souvent sur les réseaux sociaux).

Les lignes directrices sur l'éthique et la gouvernance de l'IA pour les grands modèles de langues (Large Language Models ou LLMs) de l'OMS, publiées en janvier 2024, proposent plus de 40 recommandations appelant à un déploiement contrôlé, éthique et équitable des LLMs dans le domaine de la santé. L'objectif de l'OMS est de mettre en garde contre les risques des LLMs : production d'informations biaisées ou inexacts, qualité insuffisante des données, atteinte à la vie privée, ou encore automatisation excessive des décisions médicales. Les dérives potentielles incluent également une dépendance injustifiée aux algorithmes, un accès inégal à ces technologies, et des failles de cybersécurité. L'OMS souligne la nécessité d'une gouvernance inclusive et transparente, impliquant toutes les parties prenantes (gouvernements, développeurs, soignants, patients, ...). Elle encourage les États à : investir dans des infrastructures ouvertes, garantir le respect des droits humains via la législation, créer des instances de régulation, imposer des audits indépendants sur l'impact des LLMs. Les développeurs doivent intégrer les usagers finaux dès les phases de conception, anticiper les effets indirects des LLMs, et s'assurer que leurs produits soient conçus pour répondre à des finalités éthiquement justifiables.

Le Règlement européen sur l'IA (RIA ou AI Act), entré en vigueur le 1er août 2024, vise à établir un cadre juridique harmonisé pour le développement, le déploiement et l'usage de l'IA au sein de l'UE. Il promeut une IA centrée sur l'humain, respectueuse des droits fondamentaux, de la démocratie et de l'état de droit. Il complète le RGPD, en ajoutant des exigences spécifiques à l'IA sur la transparence, la responsabilité et la protection des données à caractère personnel. Il est aligné sur les recommandations de l'UNESCO (droits humains, équité, transparence, durabilité), les principes de l'OCDE pour une IA responsable, et les réglementations sectorielles dont le MDR qu'il complète en imposant des exigences supplémentaires (sécurité, robustesse, traçabilité, éthique). L'AI Act catégorise les SIA selon quatre niveaux de risque (inacceptable, élevé, limité et minimal) avec des règles spécifiques pour les systèmes d'IA à haut risque, comme ceux utilisés dans le domaine de la santé. Il s'applique à plusieurs acteurs impliqués dans le cycle de vie des SIA, dont les fournisseurs de solutions d'IA, les fabricants intégrant des solutions d'IA dans leurs produits, les importateurs et les distributeurs, les représentants des fournisseurs ou des fabricants, les déployeurs (toute personne physique ou morale, autorité publique, agence ou autre organisme utilisant un système d'IA sous sa propre autorité à l'exception des usages personnels non professionnels). A titre d'exemple, dans le secteur de la santé, une structure (e.g., un hôpital) ayant recours à un SIA est qualifiée de déployeur.



1.3. Le rôle de la cellule éthique de la Délégation au numérique en santé

Donner la main au citoyen sur ses données de santé est un pilier éthique fort des actions de la Délégation au numérique en santé (DNS) du Ministère de la Santé et de l'accès aux soins. Comme c'est écrit dans la deuxième feuille de route du numérique en santé (2023-2027), « Le numérique en santé doit être massivement inclusif et solidaire afin que chacun puisse 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 ».

Pour inscrire cette vision dans la réalité, la DNS s'est dotée dès 2019 d'une cellule Éthique. Sa mission est de faire de l'éthique un élément central du virage numérique en santé, notamment grâce à l'élaboration d'outils pratiques de sensibilisation et d'évaluation éthique à destination des industriels, des professionnels de santé, des usagers du système de santé et des pouvoirs publics.

La cellule a défini l'éthique du numérique en santé comme issue de deux sources : d'une part, l'éthique clinique, telle que décrite dans le serment d'Hippocrate et reprise par les principes de la bioéthique, et, d'autre part, l'éthique du numérique, qui conçoit la technologie comme un outil au service des utilisateurs. L'éthique du numérique en santé repose aujourd'hui sur cinq principes bien établis : la bienfaisance, la non-malfaisance, le respect de l'autonomie des personnes, la justice ou l'équité, l'écoresponsabilité et le développement durable.

Afin de structurer et de diffuser ces principes, le Cadre de l'éthique du numérique en santé (CENS) a été co-construit avec les acteurs de l'écosystème. Il répond à la recommandation du [rapport Pon-Coury](#) d'inscrire l'accélération du virage numérique dans un cadre de valeurs éthiques. Ce cadre vise à rendre opérable, de manière pragmatique et humaniste, les principes éthiques, dans l'objectif de garantir la confiance des citoyens et des soignants.

La cellule Éthique est soutenue dans ses travaux par le Comité de l'éthique du numérique en santé (COMENS). Ce comité pilote les grandes orientations, la production et la mise à jour des documents du CENS. Il veille à leur alignement avec les évolutions technologiques, sociales, réglementaires, ainsi qu'avec les retours de la [plateforme d'éthico-vigilance](#).

Les travaux de la cellule Éthique sont portés par 15 groupes de travail, parmi lesquels un groupe de travail, le GT3, est chargé des questions relatives à l'éthique de l'IA en santé. En mai 2022, le GT3 a publié un guide de « [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](#) ». Face à la montée de l'IA générative et la prolifération d'une offre de solutions d'IA en santé non évaluées, la cellule Éthique de la DNS a souhaité aller plus loin en initiant des travaux sur un référentiel de l'éthique de l'IA en santé. Cependant, la publication en août 2024 du règlement européen sur l'IA a conduit à réorienter ces travaux vers la publication d'un guide d'implémentation à destination des fournisseurs de systèmes d'intelligence artificielle (SIA) en santé, évitant ainsi le risque de sur-spécification du cadre légal.

Ainsi, le guide d'implémentation de l'éthique d'un SIA en santé n'a pas vocation à être opposable en tant que tel. Il vise à accompagner les acteurs de l'IA en santé sur le chemin de l'éthique de l'IA à travers des critères pratiques et pragmatiques, explicites et auto-évaluables. Néanmoins, certains de ces critères pourront, à terme, être repris dans des référentiels sectoriels opposables pour l'obtention de certains bénéfices, à l'image de ce qui a déjà été mis en œuvre pour l'éthique de la téléconsultation, ou de ce qui est prévu pour le référentiel de sécurité, d'interopérabilité et d'éthique des dispositifs médicaux numériques.



Méthode de travail adoptée par le GT3

2.1 Présentation des travaux

Les travaux menés par la cellule Éthique de la DNS se sont déroulés entre 2023 et 2025 selon les modalités suivantes :

- **Animation d'ateliers de réflexion et de sessions de travail** en 4 phases menant à l'élaboration d'une première version du guide fin 2024.
- **Analyse des recommandations « Éthique by design »** et identification des critères potentiels à inclure dans le guide DNS d'implémentation de l'éthique d'un SIA en santé.
- **Identification des sujets clés** à explorer et à inclure dans le guide DNS d'implémentation de l'éthique d'un SIA en santé.
- **État des lieux des travaux en cours sur l'IA** (OMS, OCDE, MDR, RIA, CCNE, CNPEN, CNIL...) pour enrichir les réflexions.
- **Élaboration d'une première version V0.1 du guide d'implémentation de l'éthique** d'un SIA en santé fin 2024.

1.

Analyse comparative des critères convergents et divergents de la V0.1 du guide DNS et des critères des labels de l'éthique de l'IA : Labelia, Ethical IA, Ethik-IA et Ekitia.

2.

Analyse comparative des critères convergents et divergents de la V0.1 du guide DNS et du RIA, et élaboration d'une V0.2 du guide DNS.

3.

Présentation de la V0.2 du guide DNS au GT3, mais également à différentes institutions et acteurs de l'écosystème, analyse des retours et rédaction d'une V0.3 soumis à la concertation publique en mai 2025.

4.

Intégration des retours de la concertation publique et publication de la version finale V0.1 du guide en juillet 2025.

IMPLICATION DU CONSEIL SCIENTIFIQUE, TECHNIQUE ET ÉTHIQUE ET

GESTION DES RISQUES

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 p

- Le consentement éclairé d à la réutilisation de leurs d de la finalité première du (conformité RGPD, art. 5 e
- La proportionnalité des d par rapport à la finalité du l'élaboration de la solution RGPD, art.5, c)
- La non réidentification dir (agrégées, pseudonymisée (conformité RGPD, art. 32)
- La qualité des données (lu cognitifs)
- La représentativité de la p /population cible / préven discriminations (lutte cont sélection)

Mise en œuvre des mesures c à assurer :

- Le transfert sécurisé des d unique, sources multiples,
- La qualité de l'hébergement serveurs localisés en Franc privé/public, centralisé/dis HDS)
- La cybersécurité à l'état d

Mise en œuvre de mesures p la non-réutilisation non éthic (par ex. en cas de fusion de l modifications législatives (p la destruction automatique c

Figure 1 – Questionnements éthiques attachés à chaque étape des solutions d'Intelligence Artificielle en Santé : mise en œuvre

2.2 Analyse et définition du périmètre

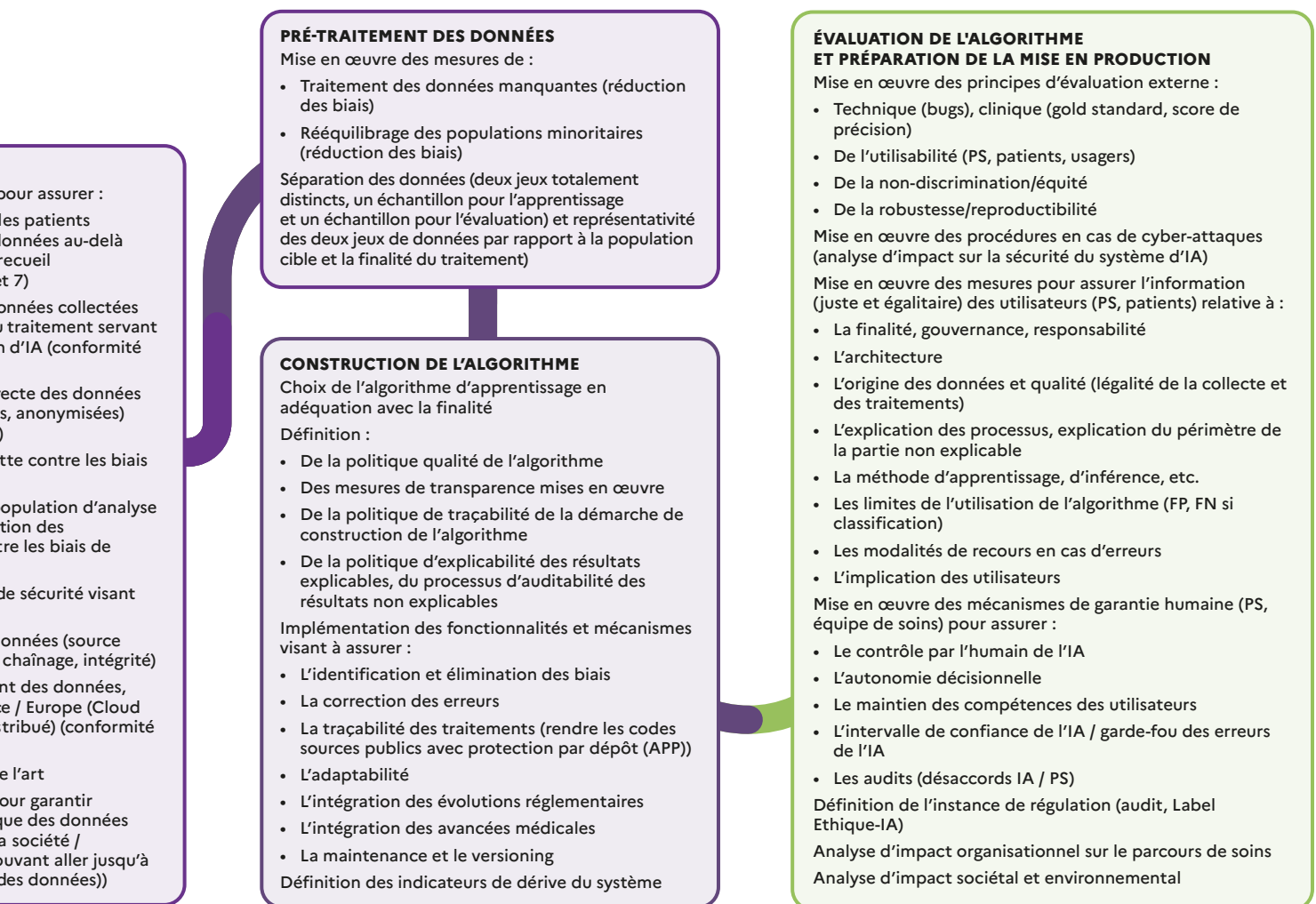
La nécessité d'un cadre éthique de l'IA

Les travaux du GT3 ont débuté par l'identification des besoins spécifiques et des enjeux éthiques liés à l'implémentation d'un SIA en santé. Pour cela, le groupe de travail s'est appuyé sur les recommandations « Éthique by design » publiées en 2022 (cf. Figure 1), afin d'identifier les sujets clés à approfondir à chaque étape de la démarche. Ces thématiques ont ensuite été croisées avec les réflexions menées par les institutions nationales et internationales sur l'éthique de l'IA (cf. Figure 2).

L'analyse a pris en compte :

- **les démarches internationales**, telles que le code de conduite du G7 sur l'IA générative ;
- **les réglementations européennes**, notamment le Règlement de l'IA (RIA) en cours d'élaboration au moment des travaux, le Règlement sur les dispositifs médicaux (MDR) et le Règlement général sur la protection des données (RGPD) ;
- **ainsi que des références nationales**, parmi lesquelles l'avis 141 du Comité Consultatif National d'Éthique (CCNE) et l'avis 4 du Comité National Pilote de l'Éthique (CNPEN).

DES UTILISATEURS DANS LE DÉVELOPPEMENT ET LE DESIGN DE LA SOLUTION D'IA



de la construction d'un SIA en santé (extrait des *Recommandations de bonnes pratiques pour intégrer l'éthique dès le développement* de « l'éthique by design »).

5 PHASES PROJETS ONT ÉTÉ IDENTIFIÉES AVEC DES QUESTIONS À SE POSER À CHACUNE DES ÉTAPES

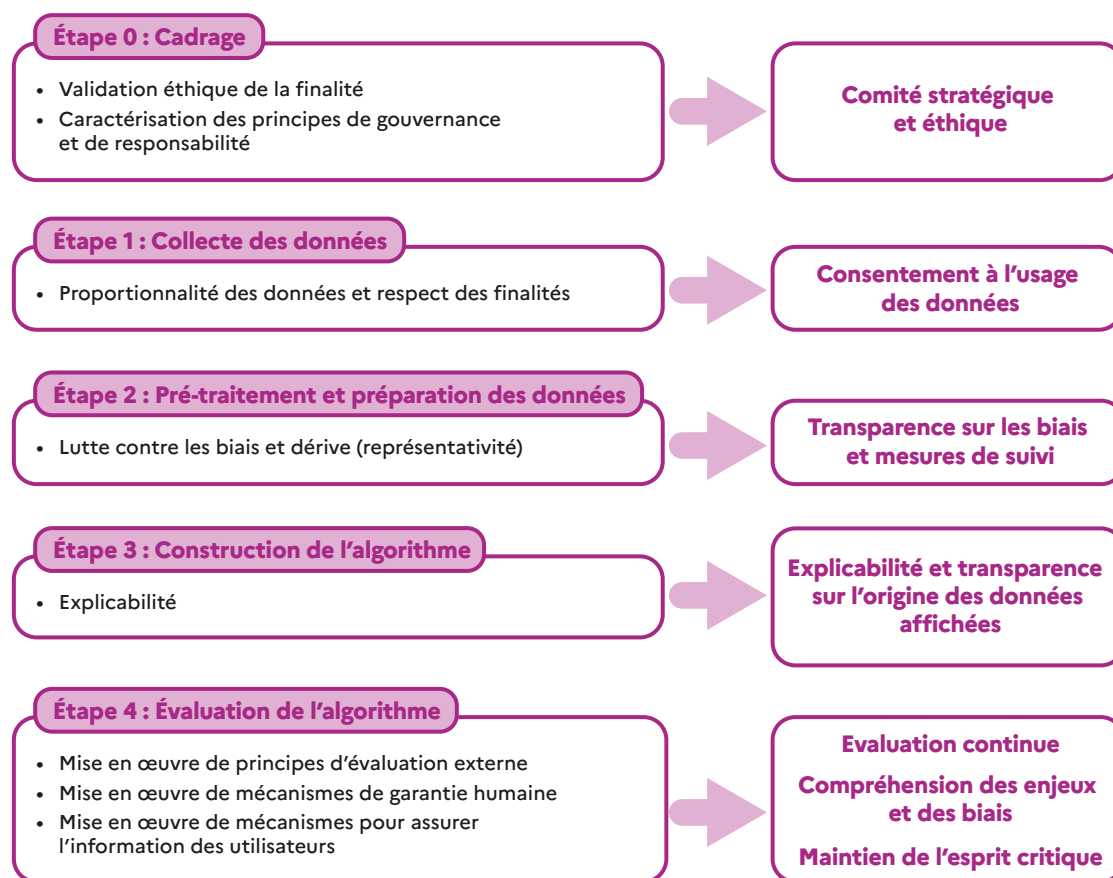


Figure 2 – Identification des critères issus des recommandations Éthique by design à conserver.

2.3 Construction d'une première version V0.1

Sur la base des recommandations pour une « Éthique by design » des SIA en santé, le GT3 a défini des principes directeurs et a formalisé en Mai 2024 une première liste de critères pour guider l'implémentation de l'éthique d'un SIA en santé. Chaque critère était associé à un principe du CENS (cf. Figure 3). Les critères étaient ainsi classés selon la bienfaisance, la non malfaisance, la justice et l'équité, l'autonomie, et l'écoresponsabilité. Les travaux d'élaboration du Règlement européen sur l'Intelligence Artificielle (RIA) étant alors en cours, le GT3 a élaboré une première version V0.1 du guide DNS indépendamment du RIA.

Au vu de l'essor de l'IA générative, des LLMs et de leurs implications, le groupe de travail a décidé en septembre 2024 de définir des critères spécifiques aux systèmes intégrant de l'IA générative venant enrichir et compléter la première liste, de façon à aboutir à une première version V0.1 de 38 critères, dont 6 spécifiquement dédiés à l'IA générative.

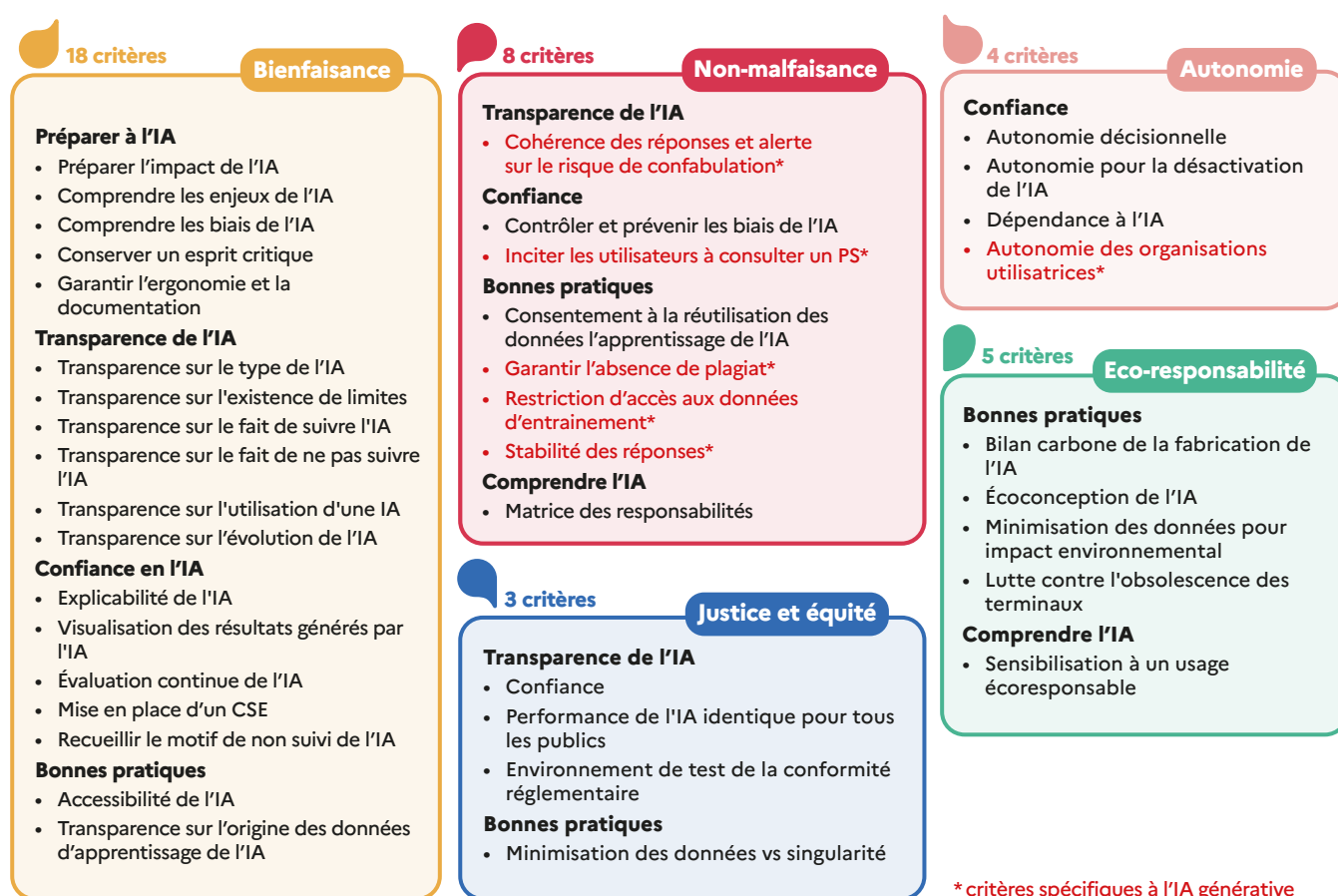


Figure 3 – Libellés courts des critères de la V0.1 du guide (38 critères), les critères avec * sont spécifiques à l'IA générative.

2.4 Comparaison de la V0.1 avec les labels de l'éthique de l'IA

Les labels de l'éthique de l'IA jouent un rôle important dans la promotion des pratiques responsables au cours du développement et de l'utilisation d'un SIA. Dans le cadre de ses travaux, le GT3 a identifié 7 labels, qui ne s'appliquent pas tous au secteur de la santé, mais dont l'analyse et l'approche éthique sont pertinentes dans le cadre de cette étude : Labelia, Ethical IA, Ethik-IA, Ekitia, GEEIS IA, Adel AI act et Positive AI.

Une analyse comparative entre les 38 critères de la V0.1 du guide DNS et les critères ou les recommandations des différents labels a pu être menée auprès de 4 des 7 labels identifiés. Deux d'entre eux étant disponibles dans l'espace public, le label Labelia et le guide Ethical IA, l'analyse comparative a pu être effectuée par les équipes de la DNS. Pour deux autres labels, Ethik-IA et Ekitia, les critères d'évaluation n'étant pas publics, les porteurs des labels ont été sollicités pour l'analyse.

En identifiant les critères de la V0.1 du guide DNS qui étaient couverts, non couverts ou partiellement couverts par les labels, cette analyse avait pour objectif d'aboutir à une estimation du taux de couverture globale de la V0.1 du guide DNS par chaque label. Dans chaque cas, un score de recouvrement a été associé à chaque critère du référentiel : 2 pour un critère totalement couvert, 1 pour un critère partiellement couvert, 0 pour un critère non couvert. La somme de ces scores a ensuite permis de déterminer un pourcentage de recouvrement de la V0.1 du guide DNS par les labels.

Cette analyse de recouvrement a également été réalisée pour comparer les critères de la V0.1 du guide DNS avec les recommandations « Éthique by design ». Les résultats sont synthétisés dans le Tableau 1. Les détails de cette évaluation sont disponibles en Annexe 1.

Les pourcentages de recouvrement obtenus pour chaque label ne doivent pas être interprétés comme une « note », mais plutôt comme une indication de leur degré de similarité avec le projet de guide d'implémentation de l'éthique d'un SIA en santé proposé par la DNS.

V0.1 du guide DNS	« Éthique by design »	Labelia	Ethical AI	Ethik-IA	Ekitia
Critères spécifiques DNS (non couverts)	18	29	24	3	13
Critères totalement couverts	14	3	9	27	7
Critères partiellement couverts	7	6	5	8	18
Score de recouvrement	46%	16%	30%	82%	42%

Tableau 1 : Synthèse des recouvrements de la V0.1 du guide DNS par les critères des labels.

Cette similarité dépend principalement des objectifs poursuivis et du périmètre couvert par chaque label. Ainsi, le label Labelia poursuit des objectifs distincts de ceux du guide de la DNS (recouvrement de 16%). Le label Ethical AI ne comporte pas de critères relatifs au développement durable ou à l'écoresponsabilité (recouvrement de 30%). Le label Ethik-IA, avec un taux de recouvrement de 82 %, est celui dont le périmètre est le plus proche de celui du guide DNS, en particulier parce qu'il est également centré sur l'IA en santé. Le label Ekitia, quant à lui, n'intègre pas de critères liés à la compréhension par l'utilisateur, à l'IA générative, et n'est pas spécifique au secteur de la santé (recouvrement de 42%).

Cette analyse a également pu mettre en évidence que certains critères de la V0.1 du guide DNS étaient non couverts par une majorité des labels (i.e. par au moins 3 des 4 labels évalués) :

- **Transparence sur le fait de suivre ou ne pas suivre l'IA**
- **Transparence sur l'évolution de l'IA**
- **Recueil du motif de non suivi de l'IA**

- **Dépendance à l'IA et autonomie de désactivation de l'IA**
- **Les critères liés au développement durable**
- **Les critères liés à l'IA générative**

À l'inverse, certaines thématiques communes aux différents labels n'étaient pas couvertes par la V0.1 du guide DNS, en particulier :

- **Les notions de sécurité de l'information et du système d'IA**
- **Les critères liés à la qualité de données** (évaluation, amélioration continue, transparence)
- **La notion de traçabilité des données**
- **Les questions liées à la finalité de l'IA** (intérêt général, balance bénéfice/risque)

Enfin, on observe que les catégories du CENS de bienfaisance, justice et équité, et écoresponsabilité apparaissent comme les plus couvertes par 3 des 4 labels étudiés (cf. Figure 4).

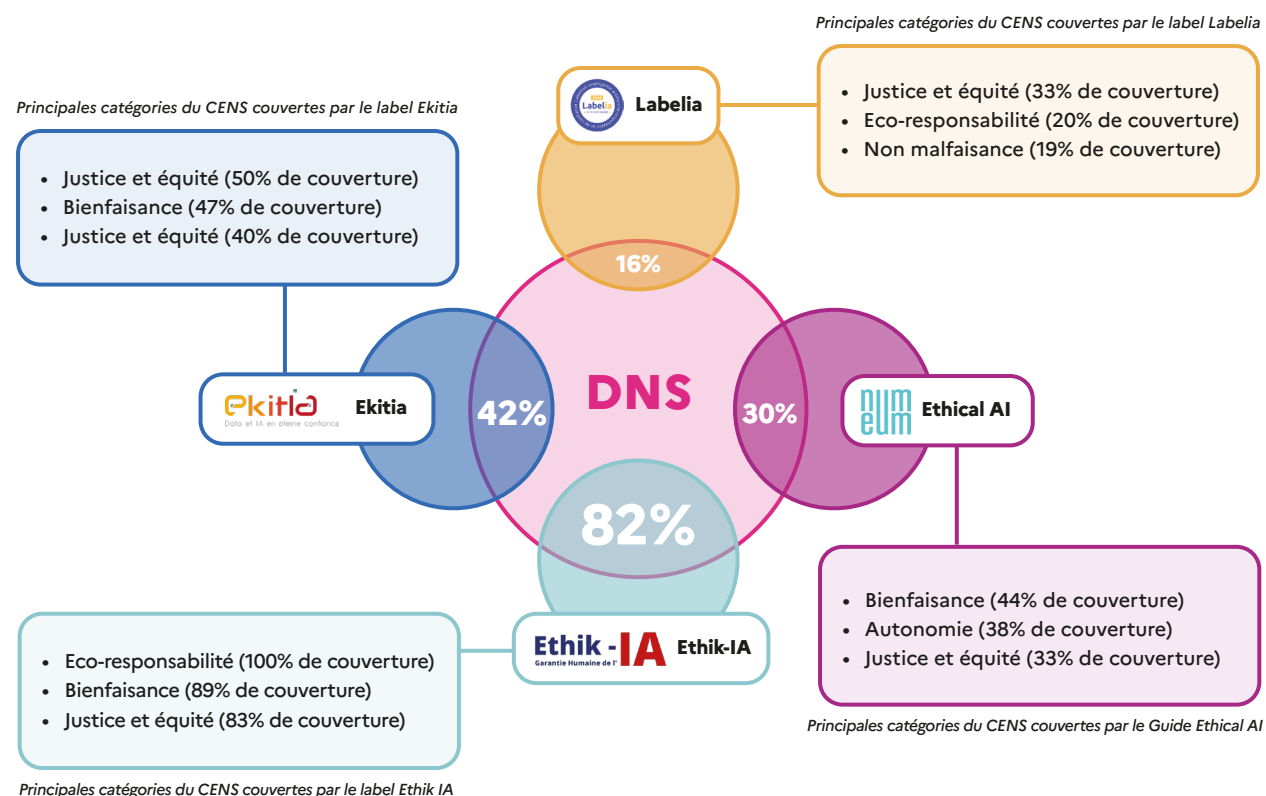


Figure 4 – Synthèse du recouvrement de la V0.1 du guide DNS par les différents labels.

2.5 Comparaison de la V0.1 avec le Règlement européen sur l'IA

Les labels de l'éthique de l'IA jouent un rôle. Dès la publication du Règlement européen sur l'IA (RIA) en juillet 2024, un travail de comparaison entre la V0.1 du guide DNS et les articles du RIA a été engagé en collaboration avec le SNITEM. Cette analyse a été menée en parallèle et de manière indépendante par le SNITEM, la DNS et l'ANS.

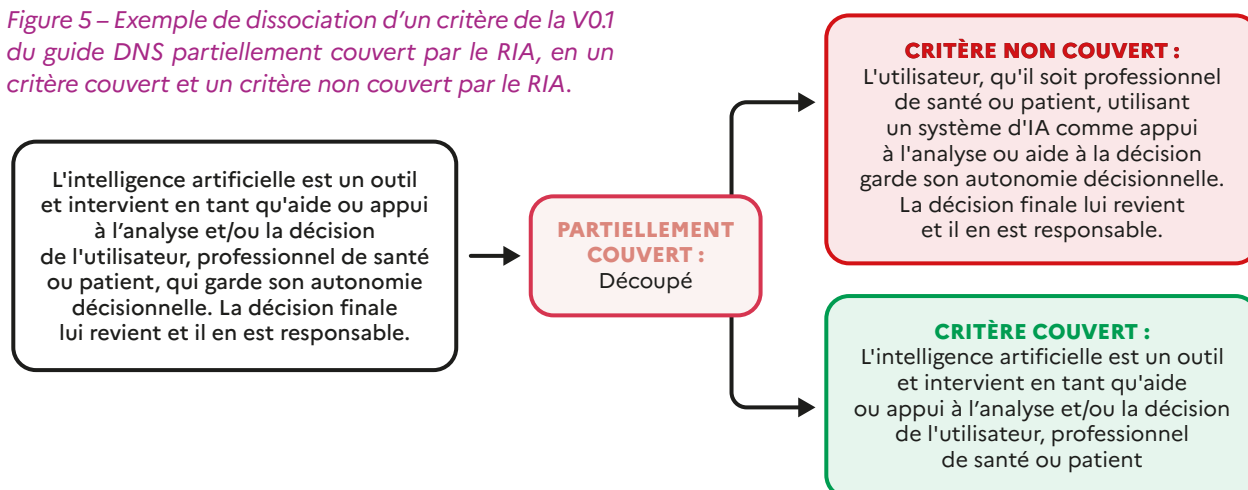
L'objectif de ce travail était de déterminer, pour chaque critère de la V0.1 du guide DNS, s'il était totalement couvert, partiellement couvert ou non couvert par le RIA. Les résultats initiaux ont mis en évidence des divergences notables : le SNITEM considérait que 31 critères du guide étaient couverts par le RIA, dont 2 partiellement et 5 non

couverts, tandis que la DNS et l'ANS s'accordaient pour dire que seuls 3 critères étaient totalement couverts, 12 partiellement couverts et 23 non couverts.

Afin d'harmoniser ces positions, des séances de travail communes ont été organisées. Elles ont permis de constater que de nombreux critères initialement considérés comme couverts par le SNITEM ne l'étaient en réalité que partiellement. Pour clarifier l'analyse et rendre le guide plus lisible, il a été décidé de scinder les critères partiellement couverts en deux sous-critères distincts, chacun étant soit totalement couvert, soit non couvert par le RIA. La Figure 5 illustre sur un exemple ce principe de découpage.

Ce processus a conduit à la publication d'une V0.2 du guide DNS, qui est passé de 38 à 54 critères (cf. Figure 6). Parmi ceux-ci, 17 critères étaient désormais considérés comme couverts par le RIA, tandis que 37 ne l'étaient pas.

Figure 5 – Exemple de dissociation d'un critère de la V0.1 du guide DNS partiellement couvert par le RIA, en un critère couvert et un critère non couvert par le RIA.



28 critères

Bienfaisance

Comprendre l'IA

- Préparer l'impact de l'IA
- Co-construire les ressources de formation aux enjeux de l'IA
- Vérifier la bonne compréhension des enjeux de l'IA
- Réduire et éliminer les biais de l'IA
- Expliquer les biais de l'IA
- Vérifier la bonne compréhension des limites de l'IA
- Former les utilisateurs en amont de l'introduction d'un SIA
- Garder un esprit critique
- Accès facile et intuitif à la documentation d'un SIA
- Garantir l'ergonomie et l'intelligibilité des interfaces du SIA

Transparence de l'IA

- Transparence sur l'existence de limites
- Transparence sur le type de l'IA
- Transparence sur le fait de suivre l'IA
- Transparence sur le fait de ne pas suivre l'IA
- Transparence sur le fait de ne pas avoir utilisé l'IA
- Transparence sur l'évolution de l'IA
- Vérifier la bonne compréhension de l'impact d'une évolution de l'IA
- Transparence sur l'interaction avec une IA
- Vérifier la bonne compréhension de l'interaction avec une IA

Confiance en l'IA

- Mise en place d'un CSE
- Recueillir le motif de non suivi de l'IA
- Explicabilité de l'IA
- Visualisation des résultats générés par l'IA
- Évaluation continue de l'IA
- Mettre en place une démarche d'amélioration continue de l'IA

Bonnes pratiques

- Conformité RGAA des SIA avec IHM
- Évaluation de l'utilisabilité de l'IA
- Évaluer le SIA en continu
- Transparence sur l'origine des données d'apprentissage de l'IA

5 critères

Eco-responsabilité

Bonnes pratiques

- Bilan carbone de la fabrication de l'IA
- Écoconception de l'IA
- Minimisation des données pour impact environnemental
- Lutte contre l'obsolescence des terminaux

Comprendre l'IA

- Sensibilisation à un usage écoresponsable

10 critères

Non-malfaisance

Transparence de l'IA

- Cohérence des réponses*
- Qualité des réponses et avertissement sur les risques de confabulation*

Confiance

- Transparence sur les mécanismes mis en place sur le contrôle et la prévention des biais

- Inciter les utilisateurs à consulter un PS*

Bonnes pratiques

- Anonymisation des données de l'apprentissage
- Consentement à la réutilisation des données pour l'apprentissage de l'IA (pseudonymisation des données)
- Stabilité des réponses*
- Garantir l'absence de plagiat*
- Restriction d'accès aux données d'entraînement*

Comprendre l'IA

- Matrice des responsabilités

4 critères

Justice et équité

Transparence de l'IA

- Transparence sur la performance de l'IA
- Performance de l'IA identique pour tous les publics
- Environnement de test de la conformité réglementaire

Bonnes pratiques

- Minimisation des données vs singularité

7 critères

Autonomie

Confiance

- Autonomie décisionnelle
- Rappeler la responsabilité de l'utilisateur sur sa décision
- Modularité de l'IA
- Autonomie pour la désactivation de l'IA
- Traçabilité de la désactivation de l'IA
- Dépendance à l'IA
- Alerter en cas de détection d'une dépendance à l'IA
- Autonomie des organisations utilisatrices*

* critères spécifiques à l'IA générative

Figure 6 – Libellés courts des 54 critères de la V0.2 du guide DNS.

2.6 Finalisation itérative du guide d'implémentation : les versions V0.2, V0.3 et V0.4

À l'issue de l'étape précédente, la V0.2 du guide DNS a été présentée aux membres du GT3, ce qui a donné lieu à plusieurs propositions d'amélioration, toutes intégrées dans une nouvelle version V0.3 du guide DNS :

- **L'organisation de la liste des critères a été repensée afin d'améliorer la lisibilité du guide.** Alors qu'ils étaient initialement organisés selon les grandes familles de l'éthique, issues des fondements du Cadre de l'éthique du numérique en santé (CENS), les critères ont été présentés selon l'ordre chronologique des phases de développement d'un système d'intelligence artificielle (SIA). Cette nouvelle présentation, plus opérationnelle, est mieux adaptée aux principaux destinataires du guide – les fournisseurs de SIA – et facilite ainsi son appropriation. La référence aux familles éthiques reste toutefois précisée pour chaque critère.
- **Les critères ont été simplifiés :** certains critères ont été fusionnés, d'autres supprimés en raison de redondances, et plusieurs ont été reformulés pour renforcer leur lisibilité et leur accessibilité.
- **Lorsqu'un critère était associé à une réglementation** (RGPD, RIA, Règlement de l'Espace Européen des Données de Santé), à une norme ou à un standard, les références précises – articles réglementaires, normes ou standards – ont été explicitement indiqués.
- **Afin de faciliter la mise en œuvre effective des critères,** chacun a été complété par des pistes d'autoévaluation, précisant les attendus, et illustrés par des exemples concrets de mise en œuvre. Ces éléments visent à permettre aux fournisseurs de SIA d'évaluer concrètement leur conformité aux recommandations éthiques. Ils constituent une première base pour démontrer la conformité des solutions et pourront être précisés et adaptés lors de l'intégration des critères dans des référentiels sectoriels.
- **20 janvier 2025 :** Le Comité Consultatif National d'Éthique du numérique (CCNE Numérique), instance nationale indépendante, ayant pour mission de contribuer à la réflexion sur les enjeux d'éthique soulevés par les avancées des sciences, des technologies, des usages et des innovations dans le domaine du numérique, comme par exemple en Intelligence Artificielle (Claude Kirchner, Ingrid Callies).
- **28 janvier 2025 :** La Commission Nationale de l'Informatique et des Libertés (CNIL), autorité administrative indépendante, chargée de veiller à la protection des données personnelles, contenues dans les fichiers et traitements informatiques ou papiers (Hélène Guimiot, Aurore Gaignon, Léo Ifergan).
- **3 mars 2025 :** Le Service de l'économie numérique de la Direction Générale des Entreprises (DGE) : service du ministère de l'Économie qui propose, met en œuvre et évalue les politiques de l'État dans les technologies numériques en veillant au développement de leurs usages ainsi qu'à l'intérêt des utilisateurs (Marie-Léa Rols, Armand Lacombe, Jeanne Blain).
- **6 mars 2025 :** NUMEUM, organisation professionnelle représentant les entreprises du numérique en France (Mariane Cimino, Anne-Sophie Bouy Plagnard, Charles Mariaux, Atef Ben Othman).
- **14 mars 2025 :** France Assos Santé, l'union nationale des associations d'utilisateurs du système de santé (Arthur Dauphin, Celia Verland, Alain Olympie, Alain Rochon, Patrick Brillet, Christiane Gachet, Daniel Cadet, Fauzi el Mazzouz, Francine Laigle, Gérard le Goff, Gérard Perrier, Jean-Claude Bourquin, Jean-Claude Malaize, Jean-Philippe Rigaud, Jeannette Guedmi, Alain Laforêt, Marie-Catherine Time, Marie-Françoise Calmus, Mathilde Dissoubray, Mathilde Pitaval, Maureen Pollart, Michel Darde, Micheline Claes, Hervé Nabarette, Nathalie Torres, Olivier Dauptain, Régine Ernst, Serge Rieupeyrou, Sylvie Crelier).
- **20 mars 2025 :** Le Comité Déontologique du LEEM (CODEEM), organe de régulation éthique de l'industrie pharmaceutique représentée par le LEEM (Sofia Afonso, Valérie Coignard, Gonzague Pelletier, Dora Talvard).

La V0.3 du guide a été ensuite présentée à une communauté élargie, rassemblant de nombreuses instances et acteurs de l'écosystème du numérique en santé (cf. Figure 7) :

- **24 mars 2025** : Le Syndicat National de l'Industrie des Technologies Médicales (SNITEM), organisation représentant les fabricants de dispositifs médicaux (Arnaud Augris, Cécile Vaugelade).
- **28 mars 2025** : L'Alliance des Syndicats Indépendants des Industries du Numérique en Santé, de l'Hébergement et de la Protection des Données de Santé (ASINHPA), représentant des acteurs du numérique en santé autour des questions d'hébergement et de sécurité des données (Elise Guittard, Sarah Peigne, Elodie Pospieszynski, Vincent Mary, Mohamed Karout, David Gruson).
- **3 avril 2025** : La Haute Autorité de Santé (HAS), autorité publique indépendante visant à développer la qualité dans le champ sanitaire, social et médico-social, au bénéfice des personnes, évalue les produits de santé, élabore des recommandations pour améliorer la qualité des soins (Corinne Collignon, Simon Renner).
- **30 avril 2025** : NABLA, éditeur d'un assistant médical permettant de générer les comptes rendus des consultations médicales et de les intégrer dans les logiciels métier des professionnels de santé (Léa Collier)
- **7 mai 2025** : L'Union Nationale des Professions de Santé (UNPS), organisation nationale représentant l'ensemble des professions de santé libérales (Mathilde Guest, Mélanie Ordines).
- **13 mai 2025** : Le Comité éthique de la Fédération Hospitalière de France (FHF), organisation représentant l'ensemble des établissements publics de santé et médico-sociaux, qu'elle accompagne dans la défense de leurs intérêts et le développement du service public hospitalier (Pr Bertrand Guidet)
- **15 mai 2025** : Délégation au numérique en santé, Vincent Vercamer, expert accès au marché des innovations, en charge des Dispositifs Médicaux incluant ceux avec Intelligence artificielle, mais aussi Line Farah en charge de piloter les travaux sur la stratégie de l'IA en santé.
- **28 mai 2025** : Le Département Santé et Transformation Numérique de la Direction Générale de l'Offre de Soins (DGOS) : direction du ministère de la Santé en charge de l'organisation et de la performance du système de soins (Dr Yann-Maël Le Douarin, Judicaël Thevenard).
- **4 juin 2025** : Le Domaine de la Conformité et de la Régulation, Direction Expertise, Innovation et International de l'Agence du Numérique en Santé (Christophe Pichot-Meuger, Gilles Braud, Yves Lorillard, Frédéric Law-Dune)

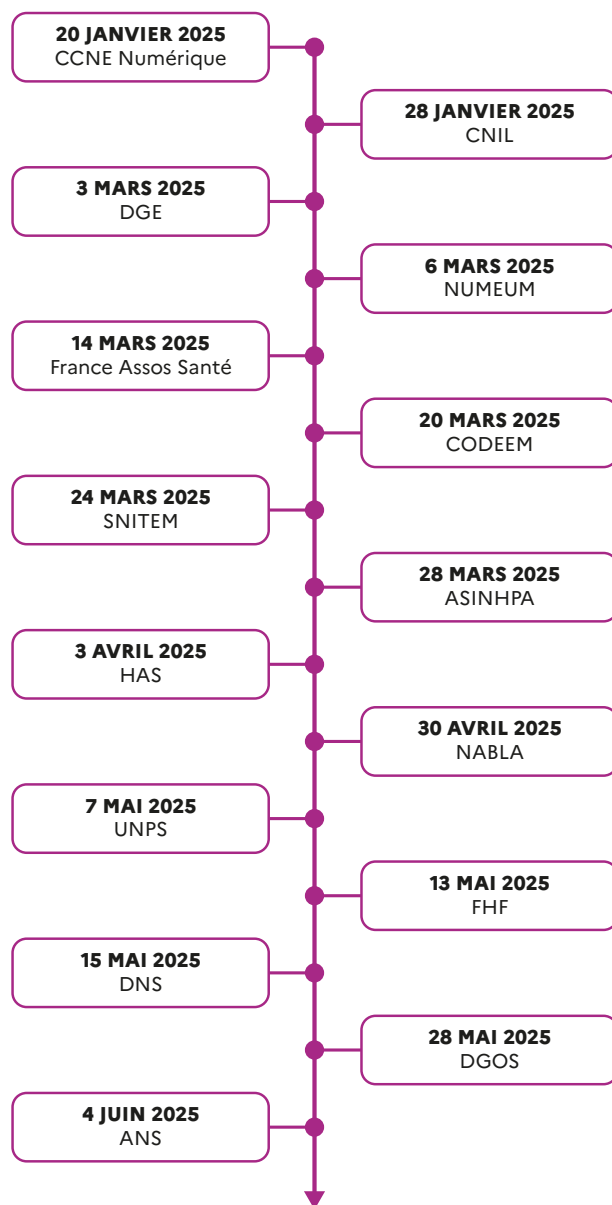


Figure 7 – Consultation de la V0.3 du guide DNS auprès des acteurs de l'écosystème.

A l'issue de ces consultations, une V0.4 du guide d'implémentation de l'éthique d'un SIA en santé, intégrant l'ensemble des retours reçus a été élaborée. Elle est constituée de 44 critères. De la même manière, chaque critère est associé à une famille éthique du CENS, à savoir, la bienfaisance, la non-malfaisance, l'autonomie, la justice ou l'équité, et l'éco-responsabilité. Les critères sont présentés suivant l'ordre chronologique des phases de développement d'un SIA, et comprennent, le cas échéant, la référence aux règlements ou normes existant attachés au critère (par exemple l'article correspondant du RGPD ou du RIA). Des explications et des exemples d'éléments d'autoévaluation sont également proposés pour chaque critère. La V0.4 a été mise en concertation publique, du 12 mai au 06 juin 2025, sur le site de l'ANS.

Version finale du guide d'implémentation de l'éthique dans les SIA en santé

Suite à l'intégration des retours reçus à l'issue de la concertation publique, la version finale du guide d'implémentation de l'éthique d'un SIA en santé est constituée de 43 critères. De même que précédemment, chaque critère est caractérisé par un numéro, une phase de développement du système, une étape d'élaboration du SIA, le principe éthique du CENS dont il relève, un libellé court et un libellé long. Les articles de lois, règlements ou normes associés à chaque critère ainsi que les éléments d'auto-évaluation proposés aux fournisseurs de SIA figurent en Annexe 2.

3.1 Phase de cadrage

Critère 1 : Mettre en place un Comité Scientifique et Éthique

Bienfaisance

- **Étape : Cadrage**
- Le fournisseur d'un SIA en santé devrait mettre en place, en amont de la conception de sa solution, un comité scientifique et éthique, constitué d'un panel représentatif des utilisateurs potentiels du SIA mais également des personnes pour lesquelles le SIA sera utilisé (usagers, patients). Ce comité s'inscrit dans une démarche projet et valide à toutes les étapes du cycle de vie du SIA le caractère éthique de la solution.

3.2 Phase de Développement

Critère 2 : Transparence sur l'origine des données d'apprentissage du SIA

Bienfaisance

- **Étape : Collecte et préparation des données**
- Lorsque le fournisseur d'un SIA en santé utilise des données de santé issues du soin pour entraîner l'algorithme du SIA, il devrait s'assurer que les patients ont été informés et ont pu exercer leurs droits concernant l'utilisation secondaire de leurs données de santé.

Critère 3 : Minimisation des données vs singularité des patients

Justice

- **Étape : Collecte et préparation des données**
- Le fournisseur d'un SIA en santé devrait conduire une analyse lui permettant de minimiser les données utilisées pour l'entraînement de l'algorithme, la production et l'exploitation du SIA, aux seules données nécessaires et suffisantes pour tenir compte de la singularité des patients qui constituent la cible d'usage du SIA.

Critère 4 : Minimisation des données vs impact environnemental

Éco-responsabilité

- **Étape : Collecte et préparation des données**
- Le fournisseur d'un SIA en santé devrait conduire une analyse lui permettant de minimiser les données utilisées pour l'entraînement de l'algorithme, la production et l'exploitation du SIA, aux seules données nécessaires et suffisantes afin de réduire l'impact environnemental des traitements.

Critère 5 : Réduire et éliminer les biais du SIA

Bienfaisance

- **Étape : Conception de l'algorithme du SIA**
- Le fournisseur d'un SIA en santé devrait mettre en œuvre des mécanismes afin de réduire les biais du SIA et idéalement de les éliminer.

Bienfaisance

Critère 6 : Explicabilité du SIA

- **Étape : Conception de l'algorithme du SIA**
- Le fournisseur d'un SIA en santé qui propose une aide à la décision avec une interface utilisateur, qu'ils soient professionnels de santé ou patients, devrait mettre en place des mécanismes permettant au SIA d'expliquer les propositions faites aux utilisateurs qui en feraient la demande. Lorsque le SIA n'est pas en capacité d'expliquer ses propositions, le fournisseur du SIA devrait décrire la méthode d'apprentissage qui a été mise en œuvre dans le SIA en indiquant pourquoi les propositions du SIA ne peuvent pas être expliquées.

Critère 7 : Empêcher l'IA générative de restituer des données à caractère personnel qui auraient été utilisées pour son entraînement (G)

Non-malfaisance

- **Étape : Conception de l'algorithme du SIA**
- Lorsqu'un SIA générative a été entraîné à partir de données à caractère personnel, le fournisseur du SIA devrait mettre en œuvre des mécanismes permettant de garantir que les données à caractère personnel utilisées pour l'entraînement ne peuvent pas être utilisées par le SIA dans le texte de ses réponses.

Critère 8 : Garantir l'absence de plagiat avec les IA génératives (G)

Non-malfaisance

- **Étape : Conception de l'algorithme du SIA**
- Le fournisseur d'un SIA générative devrait mettre en œuvre des mécanismes permettant de garantir que les réponses apportées ne sont pas la recopie exacte de textes d'auteurs utilisés pour son entraînement, à moins d'explicitement qu'il s'agit de citations et d'indiquer les sources garantissant ainsi l'absence de plagiat dans les réponses produites.

Critère 9 : Autonomie pour la désactivation d'un SIA modulaire

Autonomie

- **Étape : Conception de l'algorithme du SIA**
- Le fournisseur d'un SIA modulaire en santé (i.e. intégré dans un système constitué de différentes applications pouvant être désactivées séparément les unes des autres) devrait mettre en œuvre des mécanismes permettant à l'utilisateur, qu'il soit professionnel de santé ou patient, d'être autonome pour désactiver le SIA, sans devoir arrêter le fonctionnement des autres applications du système.

Critère 10 : Écoconception du SIA et lutte contre l'obsolescence des équipements

Éco-responsabilité

- **Étape : Conception de l'algorithme du SIA**
- Le fournisseur d'un SIA en santé devrait vérifier le référentiel général d'écoconception des services numériques (RGESN) de la DINUM dans la conception du SIA et participer à la lutte contre l'obsolescence des équipements, en favorisant une démarche d'économie circulaire, et en permettant que le SIA puisse être utilisé sur des équipements ne nécessitant pas d'être de dernière génération, ceci dans la limite des contraintes propres aux systèmes de santé.

Critère 11 : Bilan carbone de la fabrication du SIA

Éco-responsabilité

- **Étape : Conception de l'algorithme du SIA**
- Le fournisseur d'un SIA en santé devrait réaliser un bilan carbone de la fabrication du SIA (entraînement de l'algorithme).

Critère 12 : Transparence sur l'interaction avec une IA

Bienfaisance

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé avec une interface utilisateur devrait informer les utilisateurs, qu'ils soient professionnels de santé ou patients, qu'ils interagissent avec une IA, au moment de leur utilisation du SIA.

Critère 13 : Vérifier la bonne compréhension de l'interaction avec une IA

Bienfaisance

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé avec une interface utilisateur devrait mettre en place des mécanismes afin de vérifier que les utilisateurs, qu'ils soient professionnels de santé ou patients, ont compris qu'au moment de leur utilisation du SIA, ils interagissent avec une IA.

Critère 14 : Marquer les résultats de l'IA

Bienfaisance

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé devrait mettre en place des mécanismes techniques afin de permettre le marquage des résultats qui ont été générés par l'IA.

Critère 15 : Traçabilité de la désactivation du SIA

Autonomie

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA modulaire en santé (i.e. intégré dans un système constitué de différentes applications pouvant être désactivées séparément les unes des autres) devrait mettre en œuvre des mécanismes permettant de garantir que toute désactivation du SIA est tracée (nom de la personne, date et durée de la désactivation).

Critère 16 : Garantir l'autonomie décisionnelle de l'utilisateur

Autonomie

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé, avec une interface utilisateur, qui propose une aide à la décision à destination des utilisateurs, qu'ils soient professionnels de santé ou patient, devrait garantir qu'ils peuvent ne pas suivre les propositions du SIA et conserver leur autonomie décisionnelle.

Critère 17 : Transparence sur le fait de suivre les propositions du SIA

Bienfaisance

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé, avec une interface utilisateur, qui propose une aide à la décision à destination des professionnels de santé devrait mettre en place des mécanismes techniques permettant de tracer, pour chaque patient, que le SIA a été utilisé, que les propositions du SIA ont été suivies, et que le patient en a été informé.

Critère 18 : Transparence sur le fait de ne pas suivre les propositions du SIA et sur le motif de non suivi

Bienfaisance

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé, avec une interface utilisateur, qui propose une aide à la décision à destination des professionnels de santé devrait mettre en place des mécanismes techniques permettant de tracer, pour chaque patient, que le SIA a été utilisé, que les propositions du SIA n'ont pas été suivies et la raison pour laquelle les propositions du SIA n'ont pas été suivies, et que le patient en a été informé.

Critère 19 : Transparence sur le fait de ne pas avoir utilisé le SIA

Bienfaisance

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé, avec une interface utilisateur, qui propose une aide à la décision à destination des professionnels de santé devrait mettre en place des mécanismes techniques permettant, lorsque le professionnel n'a pas utilisé le SIA, de tracer, pour chaque patient, que le professionnel de santé n'a pas utilisé le SIA, et que le patient en a été informé.

Critère 20 : Conformité RGAA du SIA avec une interface utilisateurs

Bienfaisance

- **Étape : Conception des interfaces**
- Tout SIA en santé disposant d'une interface utilisateur devrait être conforme au Référentiel Général d'Amélioration de l'Accessibilité (RGAA).

Critère 21 : Intelligibilité des interfaces des SIA comportant une interface utilisateurs

Bienfaisance

- **Étape : Conception des interfaces**
- Le fournisseur d'un SIA en santé doté d'une interface utilisateur met en œuvre des mécanismes permettant de garantir l'utilisabilité des interfaces et l'utilisation facile et intuitive du SIA par tout type d'utilisateur.

Critère 22 : Cohérence des réponses dans la répétition d'un SIA générative (IAgen)

Non-malfaisance

• **Étape : Évaluation du SIA**

Le fournisseur d'un SIA générative devrait mettre en œuvre des mécanismes permettant de garantir que des questions identiques posées au même moment produisent des réponses au moins comparables sur le fond.

Critère 23 : Qualité des réponses d'un SIA générative (IAgen)

Non-malfaisance

• **Étape : Évaluation du SIA**

- Le fournisseur d'un SIA générative devrait mettre en œuvre des mécanismes permettant de garantir la qualité des réponses proposées et d'avertir sur les risques d'hallucinations du SIA.

Critère 24 : Garantir que le SIA générative incite le patient utilisateur à consulter un professionnel de santé en cas d'alerte (IAgen)

Non-malfaisance

• **Étape : Évaluation du SIA**

- Le fournisseur d'un SIA générative en santé, doté d'une interface utilisateur, qui propose une aide à la décision à destination des patients devrait mettre en place des mécanismes techniques permettant de garantir que le SIA n'outrepasse pas ses compétences et incite effectivement les patients à consulter un professionnel de santé au moindre signe d'urgence sanitaire pour le patient.

Critère 25 : Performance du SIA identique pour tous les publics couverts par la population cible

Justice

• **Étape : Évaluation du SIA**

- Le fournisseur d'un SIA en santé qui propose une aide à la décision, diagnostique ou thérapeutique, des patients, devrait garantir que l'aide à la décision produite est juste et équitable pour chacun des patients de la population cible pour laquelle la solution peut être utilisée.

3.3 Phase de Déploiement et Usage

Critère 26 : Accompagner les utilisateurs/la structure à l'impact du déploiement d'un SIA

Bienfaisance

• **Étape : Déploiement, formation et utilisation du SIA**

- Lorsqu'une structure de soin souhaite s'équiper d'un SIA en santé, le fournisseur du SIA devrait accompagner les utilisateurs/la structure dans l'évaluation des impacts du déploiement du SIA sur les relations entre les professionnels et avec les patients (analyse des risques), préalablement au déploiement du SIA.

Critère 27 : Co-construire avec les utilisateurs les ressources de formation aux enjeux de l'IA

Bienfaisance

• **Étape : Déploiement, formation et utilisation du SIA**

- Les ressources pédagogiques utilisées par le fournisseur d'un SIA en santé afin de former les utilisateurs du SIA, qu'ils soient professionnels de santé ou patients, aux enjeux associés à l'usage d'un SIA sont co-construites avec un panel représentatif des utilisateurs du SIA.

Critère 28 : Former les utilisateurs aux enjeux de l'IA (biais, limites de performance)

Bienfaisance

• **Étape : Déploiement, formation et utilisation du SIA**

- Le fournisseur d'un SIA en santé devrait former le personnel concerné (déployeur ou utilisateurs selon le cas, qu'ils soient professionnels de santé ou patients) aux enjeux de l'IA, afin d'expliquer en particulier les biais et les limites de performance du SIA, ainsi que les mécanismes qui ont été mis en œuvre afin de les réduire et idéalement les éliminer.

Critère 29 : Vérifier la bonne compréhension par les utilisateurs des enjeux de l'IA en particulier les limites de performance

Bienfaisance

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait mettre en place des mécanismes afin de vérifier que les utilisateurs du SIA, qu'ils soient professionnels de santé ou patients, ont compris les enjeux associés à l'IA, notamment les limites de performance du SIA, et qu'ils sont en capacité de choisir d'utiliser le SIA ou pas, et lorsque le SIA est doté d'une interface utilisateur et produit une aide à la décision, d'adhérer, ou pas, de façon éclairée aux propositions du SIA.

Critère 30 : Former à l'utilisation du SIA en amont de son introduction

Bienfaisance

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait être responsable de la formation du personnel concerné (déployeur, responsable de garantie humaine, et selon le cas les professionnels de santé utilisateurs) à l'utilisation du SIA en amont de son déploiement.

Critère 31 : Permettre un accès facile et intuitif à la documentation du SIA

Bienfaisance

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait mettre en œuvre des mécanismes lui permettant de garantir l'accessibilité de la documentation relative au SIA afin d'en faciliter l'utilisation.

Critère 32 : Transparence sur les types d'IA employés

Bienfaisance

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait indiquer dans la documentation ou lors des sessions de formation à destination des utilisateurs, qu'ils soient professionnels de santé ou patients, le ou les types d'intelligence artificielle utilisés dans le SIA.

Critère 33 : Transparence sur la performance du SIA

Justice

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait indiquer les caractéristiques de la population cible pour laquelle le SIA peut être utilisé et le taux de performance du SIA attendu pour cette population.

Critère 34 : Détecter la dépendance au SIA

Autonomie

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait implémenter des mécanismes permettant de détecter la dépendance d'un utilisateur, qu'il soit professionnel de santé ou patient, aux propositions du SIA.

Critère 35 : Alerter en cas de détection d'une dépendance au SIA

Autonomie

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait implémenter des mécanismes permettant d'alerter en cas de détection d'une dépendance d'un utilisateur aux propositions du SIA.

Critère 36 : Garder un esprit critique et faire confiance à son expertise clinique

Bienfaisance

- **Étape : Déploiement, formation et utilisation du SIA**
- La formation des professionnels de santé à l'utilisation d'un SIA devrait insister sur l'importance de conserver un esprit critique vis à vis des propositions du SIA et d'avoir confiance en sa propre compétence clinique.

Critère 37 : Rappeler la responsabilité de l'utilisateur sur sa décision

Autonomie

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé, doté d'une interface utilisateur, qui propose une aide à la décision, devrait mettre en place des mécanismes permettant d'accompagner les propositions du SIA d'un rappel à l'attention de l'utilisateur du SIA, qu'il soit professionnel de santé ou patient, lui indiquant qu'il conserve l'entière responsabilité de ses décisions.

Critère 38 : Sensibilisation à un usage écoresponsable

Éco-responsabilité

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait inclure dans ses formations une information claire et facile à comprendre sur les bonnes pratiques d'utilisation écoresponsable du SIA. Il devrait également mettre à disposition des utilisateurs, qu'ils soient professionnels de santé ou patients, une documentation pédagogique et didactique sur l'impact environnemental de l'utilisation de l'IA.

Critère 39 : Souveraineté des données de l'organisation utilisatrice

Autonomie

- **Étape : Déploiement, formation et utilisation du SIA**
- Le fournisseur d'un SIA en santé devrait proposer des solutions techniques permettant que les données ne soient pas soumises à des réglementations extra-européennes.

Critère 40 : Évaluer le SIA en continu

Bienfaisance

- **Étape : Suivi, mise à jour et amélioration continue**
- Le fournisseur d'un SIA en santé devrait procéder à des évaluations régulières des résultats du SIA.

Critère 41 : Mettre en place une démarche d'amélioration continue du SIA

Bienfaisance

- **Étape : Suivi, mise à jour et amélioration continue**
- Le fournisseur d'un SIA en santé devrait s'inscrire dans une démarche d'amélioration continue en évaluant régulièrement les résultats du SIA et en mettant en place un plan d'actions suite aux résultats de l'évaluation du SIA (intégrant notamment la détection des dérives et la mise en œuvre de processus de contrôle humain).

Critère 42 : Transparence sur la mise à jour d'un SIA et ses conséquences

Bienfaisance

- **Étape : Suivi, mise à jour et amélioration continue**
- Le fournisseur d'un SIA en santé devrait informer les utilisateurs, qu'ils soient professionnels de santé ou patients, des mises à jour du SIA pouvant induire une modification de son utilisation ou de l'exercice du jugement de ses utilisateurs.

Critère 43 : Vérifier la bonne compréhension des conséquences d'une mise à jour d'un SIA

Bienfaisance

- **Étape : Suivi, mise à jour et amélioration continue**
- Le fournisseur d'un SIA en santé devrait mettre en place des mécanismes afin de vérifier que les utilisateurs du SIA en santé, qu'ils soient professionnels de santé ou patients, ont compris les impacts liés aux mises à jour du SIA notamment celles qui induisent une modification de l'utilisation du SIA ou de l'exercice du jugement de ses utilisateurs.

Conclusion

L'intelligence artificielle (IA) représente une avancée majeure dans le domaine de la santé, avec un potentiel significatif pour améliorer la qualité des soins, optimiser les diagnostics, personnaliser les traitements et renforcer l'organisation des parcours de soin. Loin d'être un outil unique, l'IA regroupe un ensemble de technologies comme le machine learning, le deep learning ou le traitement du langage naturel, qui interviennent déjà dans les domaines cliniques, médico-sociaux, organisationnels et de recherche. L'IA permet par exemple de détecter automatiquement des anomalies sur des images médicales, de prédire des complications, d'automatiser certaines tâches administratives ou encore d'améliorer l'analyse des bases de données en recherche. Elle promet de renforcer l'accessibilité aux soins, la sécurité des patients, et de recentrer les professionnels sur l'essentiel : la relation humaine.

Cependant, l'IA ne transforme pas uniquement les techniques : elle modifie en profondeur les pratiques et les rapports au savoir, ce qui exige une vigilance éthique forte. L'IA en santé ne peut être encadrée uniquement par la conformité réglementaire ou par la performance technique. Les SIA sont souvent opaques, auto-apprenants et peu explicables. Cela soulève des questions cruciales liées à la transparence, à l'équité des traitements, à l'autonomie des patients, à la responsabilité en cas d'erreur, et à la gouvernance démocratique de ces outils. L'encadrement éthique devient ainsi une condition indispensable pour que l'innovation reste au service de l'humain.

Le Règlement européen sur l'IA (RIA), entré en vigueur en 2024, constitue une avancée importante en fixant des exigences en matière de sécurité, de transparence et de traçabilité selon le niveau de risque des systèmes. Toutefois, il reste généraliste et ne répond pas pleinement aux spécificités du secteur de la santé, où les enjeux humains sont particulièrement sensibles. C'est pourquoi l'éthique joue un rôle complémentaire essentiel : elle permet d'interroger les finalités collectives des usages technologiques, les effets sur la relation de soin, sur les inégalités d'accès, et sur la capacité de chacun à demeurer acteur de sa propre santé. Loin d'entraver le progrès, l'éthique permet de le rendre soutenable, responsable et aligné avec les valeurs démocratiques fondamentales telles que la dignité, la justice, la solidarité et le respect de la personne.

Parmi les dérives les plus préoccupantes, le biais d'automatisation se traduit par une confiance excessive accordée par les professionnels de santé aux recommandations d'un SIA, au détriment de leur propre jugement clinique. Cela peut entraîner des erreurs graves et un relâchement de la vigilance. Ce biais appelle des réponses systémiques, comme la formation à la vigilance cognitive, la mise en place de systèmes explicables, et l'encouragement à la discussion critique. De plus, un usage répété et non questionné des SIA peut induire un déclin cognitif progressif chez les professionnels, en réduisant leur esprit critique, leur mémoire de travail et leur autonomie décisionnelle. Ce phénomène est particulièrement préoccupant pour les jeunes professionnels en formation, chez qui l'apprentissage se construit sur l'expérience, le raisonnement et la gestion de l'incertitude.

Un autre risque structurel concerne l'auto-alimentation des systèmes d'IA, lorsque ces derniers apprennent à partir de leurs propres productions, sans critique ou modification humaine. Cela peut conduire à une boucle

fermée du savoir, où les erreurs sont recyclées, les modèles se standardisent et les décisions s'éloignent de la singularité du réel et du patient. Face à ces défis, l'IA doit être conçue comme un partenaire de l'intelligence humaine, non comme un substitut. Elle peut enrichir la pratique médicale, détecter des signaux faibles, accélérer certaines décisions, mais c'est à l'humain d'en juger la pertinence et de garder le pouvoir de décision.

Dans ce cadre, l'action publique a un rôle central à jouer. Il ne s'agit pas de freiner le progrès, mais de créer un environnement de confiance pour les patients, les professionnels et les innovateurs. Cela implique le développement de référentiels éthiques concrets, le soutien à une recherche responsable, et l'association des citoyens aux décisions technologiques. C'est dans cette optique que la DNS a coordonné l'élaboration d'un guide d'implémentation de l'éthique de l'IA en santé. Ce guide propose des critères concrets, accompagnés d'outils d'autoévaluation, permettant aux développeurs et aux institutions de progresser dans une logique d'amélioration continue. Bien que non contraignants, ces critères offrent un cadre clair, adaptable et structurant pour intégrer les enjeux éthiques dans tous les projets de SIA en santé. Ce guide pourra à terme s'inscrire dans les référentiels sectoriels du numérique en santé, renforçant ainsi une approche éthique et partagée de l'innovation technologique.



Annexe 1 : Comparaison de la V0.1 du guide avec les labels de l'éthique de l'IA

5.1 Rappel des différents labels sur l'éthique de l'IA

Il existe des initiatives de labellisation éthiques de l'IA dont l'objectif est de garantir des pratiques responsables et transparentes. Ces labels sont décernés aux organisations et aux

produits qui répondent à des critères éthiques spécifiques. Les porteurs de ces labels sont issus des secteurs public et privé : le label «Labelia», le label d'Ethik IA, le guide Ethical AI, le label GEEIS IA (Gender Equality European & International Standard for AI), le label Positive IA, le label ADEL-AI act, le label EKITIA (cf. Figure 8).

Labels	Processus de labellisation	Niveaux de certification	Durée du label
	- Audit éthique et juridique - Mise en place de « Collèges de Garantie Humaine » associant concepteurs, professionnels et représentants des usagers	N/A	2 ans
	- Signature de la Charte Internationale pour une IA Inclusive. - Audit externe (7 critères) - Mise en place d'outils de pilotage par l'organisation - Évaluation continue	N/A	4 ans avec une évaluation intermédiaire après 2 ans
	- Auto-évaluation (référentiel d'évaluation composé de 36 questions couvrant 6 sections) - Audit de l'organisation par LABELIA	Labelia intermédiaire : maturité avancée dans les pratiques d'IA responsable Labelia avancé : maturité très avancée et excellence dans les pratiques d'IA éthique et responsable	2 ans
	- Auto-évaluation (50 questions couvrant 6 principes) - Audit externe de la gouvernance d'entreprise, des aspects techniques des systèmes d'IA, de la conformité aux principes éthiques et bonnes pratiques - Évaluation continue et audits réguliers	Niveau 1 : engagement initial de l'organisation et mesures de bases garantissant la transparence et l'équité Niveau 2 : maturité accrue en termes de gouvernance, politiques et procédures bien établies Niveau 3 : haut degré de maturité et excellentes pratiques d'IA responsable, leader dans leur domaine et engagée dans une démarche d'amélioration continue	2 ans
	- Auto-évaluation (référentiel technique composé de 50 questions couvrant 6 critères) - Audit externe de la gouvernance d'entreprise, des aspects techniques des systèmes d'IA, de la conformité aux principes éthiques et bonnes pratiques - Évaluation continue et audits réguliers	N/A	2 ans
	- Entretien de lancement pour clarifier les attentes - Questionnaire préliminaire (20 questions couvrant 4 critères) - Évaluation externe de la conformité du projet aux critères éthiques définis par la charte Ekitia	N/A	3 ans
	- Auto-évaluation (couvrant 6 critères) - Audit externe	N/A	3 ans

Figure 8 – Labels de l'éthique de l'IA existant en 2025.

5.2 Méthodologie de l'analyse comparative entre les critères de la V0.1 du guide DNS et les critères des labels

Une analyse comparative entre les critères et recommandations des labels et les critères de la V0.1 du guide DNS a pu être menée auprès de 4 des 7 labels identifiés. Deux d'entre eux étaient disponibles dans l'espace public : le label Labelia, le guide Ethical IA. Pour deux autres labels, Ethik-IA et Ekitia, les critères d'évaluation n'étaient pas publics, mais les équipes des labels ont été sollicitées pour l'analyse. Pour les 3 autres labels, les porteurs n'ont pas souhaité participer à la comparaison avec le guide DNS.

Cette analyse avait pour objectifs :

- D'identifier les critères du guide DNS couverts, non couverts ou partiellement couverts par le label
- D'identifier les critères et recommandations des labels qui n'étaient pas couverts par le guide DNS

Dans chaque cas, un score de recouvrement a été associé à chaque critère du guide DNS :

- 2 pour un critère totalement couvert
- 1 pour un critère partiellement couvert
- 0 pour un critère non couvert

La somme de ces scores a ensuite permis de déterminer un pourcentage de recouvrement du guide DNS par chacun des labels (100% correspondant à une couverture totale de l'ensemble des critères DNS, soit uniquement des scores de 2).

5.3 Comparaison des critères de la V0.1 du guide DNS avec les recommandations « Éthique by design »

Sur les 38 critères de la V0.1 du guide DNS, 16 n'ont pas été conservés dans le guide (considérés comme n'étant pas strictement propres à l'éthique, ou remplacés par d'autres critères plus adaptés), 14 ont été estimés totalement couverts par les recommandations, 7 partiellement couverts, et 18 nouveaux critères ont été ajoutés, soit un score de recouvrement de 46% (cf. Figure 9).

		Critères Communs 14 totalement couverts, 7 partiellement couverts		Recommandations Référentiel Ethique By Design 16 recommandations non couvertes
Bienséance 18 critères	Critères Spécifiques DNS 18 critères non couverts - Comprendre les enjeux de l'IA - Comprendre les biais de l'IA - Conserver un esprit critique - Garantir l'ergonomie et la documentation - Recueillir le motif de non suivi de l'IA	- Préparer à l'impact de l'IA - Transparence sur le type de l'IA - Transparence sur l'existence de limites - Transparence sur le fait de suivre l'IA - Transparence sur le fait de ne pas suivre l'IA - Mise en place d'un CSE - Visualisation des résultats générés par l'IA - Évaluation continue de l'IA - Transparence sur l'origine des données d'apprentissage de l'IA	- Transparence sur l'utilisation d'une IA - Transparence sur l'évolution de l'IA - Explicabilité de l'IA - Accessibilité de l'IA	- Non-ré-identification directe des données - Qualité des données - Transfert sécurisé des données - Hébergement HDS des données - Cybersécurité - Traitement des données manquantes - Qualité de l'algorithme - Traçabilité de la construction de l'algorithme - Correction des erreurs - Définition stratégie de maintenance et versionning - Évaluation de la non-discrimination/équité - Évaluation de la robustesse / reproductibilité - Information sur la finalité, la gouvernance - Mécanismes d'audit - Analyse d'impact sociétal - Analyse des risques
	Non-malfaisance 8 critères - Garantir l'absence de plagiat* - Inciter les utilisateurs à consulter un PS* - Restriction d'accès aux données d'entraînement* - Cohérence des réponses et alerte sur le risque de confabulation* - Stabilité des réponses*	- Contrôler et prévenir les biais de l'IA - Consentement à la réutilisation des données pour l'apprentissage de l'IA - Matrice des responsabilités		
	Justice et équité 3 critères Minimisation des données vs singularité	- Performance de l'IA identique pour tous les publics - Environnement de test de la conformité réglementaire		
	Autonomie 4 critères - Dépendance à l'IA - Autonomie pour la désactivation de l'IA - Autonomie des organisations utilisatrices*	Autonomie décisionnelle		
	Eco-responsabilité 5 critères - Sensibilisation à un usage écoresponsable - Écoconception de l'IA - Minimisation des données pour impact environnemental - Lutte contre l'obsolescence des terminaux	Bilan carbone de la fabrication de l'IA		

* critères spécifiques à l'IA générative

Critères DNS partiellement couverts par le label

Critères DNS totalement couverts par le label

Figure 9 – Comparaison des critères spécifiques DNS et des recommandations « Éthique by design ».

5.4 Comparaison des critères de la V0.1 du guide DNS et ceux du label Labelia

D'après l'analyse réalisée par la DNS (cf. Figure 10), il apparaît que 3 critères du guide DNS sont totalement couverts par le label Labelia, 6 critères ne sont que partiellement couverts, et 29 critères ne sont pas couverts. En suivant la méthodologie de scoring décrite précédemment, ceci correspond à un score de recouvrement de seulement 16% du guide DNS par le label. Cette différence peut notamment s'expliquer par une différence d'objectifs entre le label et le guide DNS.

Critères Communs		
3 totalement couverts, 6 partiellement couverts		
Critères Spécifiques DNS		
29 critères non couverts		
Bienfaisance 18 critères	• Préparer à l'impact de l'IA • Comprendre les enjeux de l'IA • Comprendre les biais de l'IA • Conserver un esprit critique • Garantie de l'ergonomie et de la documentation • Transparence sur le type de l'IA • Transparence sur l'existence de limites • Transparence sur le fait de suivre l'IA • Transparence sur le fait de ne pas suivre l'IA • Transparence sur l'évolution de l'IA • Recueillir le motif de non suivi de l'IA • Visualisation des résultats générés par l'IA • Accessibilité de l'IA	• Transparence sur l'utilisation d'une IA • Mettre en place d'un CSE • Explicabilité de l'IA • Évaluation continue de l'IA • Transparence sur l'origine des données d'apprentissage de l'IA
	• Consentement à la ré-utilisation des données pour l'apprentissage de l'IA • Stabilité des réponses* • Garantir de l'absence de plagiat* • Inciter les utilisateurs à consulter un PS* • Restriction d'accès aux données d'entraînement* • Cohérence des réponses et alerte sur le risque de confabulation*	• Contrôler et prévenir les biais de l'IA • Matrice des responsabilités
	• Performance de l'IA identique pour tous les publics • Environnement de test de la conformité réglementaire	• Minimisation des données vs singularité
	• Autonomie décisionnelle • Dépendance à l'IA • Autonomie pour désactivation de l'IA • Autonomie de organisation utilisatrices*	
	• Sensibilisation à un usage écoresponsable • Écoconception de l'IA • Minimisation des données pour impact environnemental • Lutte contre l'obsolescence des terminaux	• Bilan carbone de la fabrication de l'IA
* critères spécifiques à l'IA générative		
Critères DNS partiellement couverts par le label		
Critères DNS totalement couverts par le label		

Figure 10 – Comparaison des critères spécifiques DNS et ceux du label Labelia.

5.5 Comparaison des critères de la V0.1 du guide DNS et ceux de Ethical AI

L'analyse comparative montre que 9 critères du guide DNS sont totalement couverts par le guide Ethical AI, 5 critères sont partiellement couverts et 24 critères ne sont pas couverts, ce qui correspond à score de recouvrement de 30% du guide DNS par le label. La différence provient en partie du fait que le label ne comporte pas de critères en rapport avec le développement durable ou l'écoresponsabilité (cf. Figure 11).

Critères Communs 9 totalement couverts, 5 partiellement couverts		
Bienfaisance 18 critères	Critères Spécifiques DNS 24 critères non couverts	Critères Spécifiques Guide Ethical AI
	- Préparer à l'impact de l'IA - Comprendre les enjeux de l'IA - Comprendre les biais de l'IA - Conserver un esprit critique - Garantie de l'ergonomie et de la documentation - Transparence sur le fait de suivre l'IA - Transparence sur le fait de ne pas suivre l'IA - Transparence sur l'évolution de l'IA - Recueillir le motif de non suivi de l'IA	- Transparence sur le type de l'IA - Transparence sur l'utilisation d'une IA - Mettre en place d'un CSE - Explicabilité de l'IA - Visualisation des résultats générés par l'IA - Accessibilité de l'IA - Transparence sur l'origine des données d'apprentissage de l'IA
	- Matrice des responsabilités - Garantir l'absence de plagiat* - Inciter les utilisateurs à consulter un PS* - Restriction à l'accès aux données d'entraînement* - Stabilité des réponses* - Cohérence des réponses et alerte sur le risque de confabulation*	- Transparence sur l'existence de limites - Évaluation continue de l'IA
	- Performance de l'IA identique pour tous les publics - Environnement de test de la conformité réglementaire	- Contrôler et prévenir les biais de l'IA - Consentement à la ré-utilisation des données pour l'apprentissage de l'IA
	- Autonomie décisionnelle - Autonomie pour désactivation de l'IA - Autonomie de organisation utilisatrices*	Minimisation des données vs singularité
Non-malfaisance 8 critères	- Autonomie décisionnelle - Autonomie pour désactivation de l'IA - Autonomie de organisation utilisatrices*	- Autonomie décisionnelle - Dépendance à l'IA
	- Sensibilisation à un usage écoresponsable - Bilan carbone de la fabrication de l'IA - Écoconception de l'IA - Minimisation des données pour impact environnemental - Lutte contre l'obsolescence des terminaux	
Justice et équité 3 critères		
Autonomie 4 critères		
Eco-responsabilité 5 critères		

* critères spécifiques à l'IA générative

Critères DNS partiellement couverts par le label

Critères DNS totalement couverts par le label

Figure 11 – Comparaison des critères spécifiques DNS et ceux du guide Ethical AI.

5.6 Analyse comparative des labels non disponibles dans l'espace public : Ethik-IA & Ekitia

Les critères des labels Ethik-IA et Ekitia n'étant pas publics, les porteurs des labels ont été sollicités pour participer à l'exercice de comparaison.

Label Ethik-IA

Le label Ethik-IA de garantie humaine dans le secteur hospitalier se concentre sur 6 thématiques :

- **Thème #1** : Équité et universalité
- **Thème #2** : Transparence et traçabilité
- **Thème #3** : Gouvernance des données

- **Thème #4** : Facilité d'utilisation
- **Thème #5** : Robustesse
- **Thème #6** : Développement durable

L'équipe Ethik-IA a elle-même proposé les associations entre les critères du guide DNS et les principales thématiques abordés dans le label, et a calculé le pourcentage de recouvrement en suivant la même méthode de scoring.

D'après l'analyse effectuée par Ethik-IA, le label présente une forte proportion de similarité avec les critères du guide DNS : 27 critères du guide DNS sont totalement couverts par les thématiques du label Ethik-IA, 8 critères sont partiellement couverts, et 3 critères ne sont pas couverts, ce qui correspond à un score de recouvrement de 82% du guide DNS par le label (Cf. Figure 12). Ce très bon score peut s'expliquer par le fait que ce label est plus spécifique à la santé que les précédents.

		Critères Communs 14 totalement couverts, 7 partiellement couverts		Thématiques Spécifiques Ethik-IA  Garantie Humaine de l'IA - Partie Technique de l'algorithme (structure/squelette de l'algorithme) - Traçabilité - Qualité de la donnée	
Bienfaisance 18 critères	Critères Spécifiques DNS 3 critères	- Préparer à l'impact de l'IA - Transparence sur le type de l'IA - Transparence sur l'existence de limites - Transparence sur le fait de suivre l'IA - Transparence sur le fait de ne pas suivre l'IA - Mise en place d'un CSE - Visualisation des résultats générés par l'IA - Évaluation continue de l'IA - Transparence sur l'origine des données d'apprentissage de l'IA	- Explicabilité de l'IA - Visualisation des résultats générés par l'IA - Accessibilité de l'IA - Évaluation continue de l'IA - Garantie de l'ergonomie et de la documentation - Mise en place d'un CSE - Recueillir le motif de non suivi de l'IA - Transparence sur l'origine des données d'apprentissage de l'IA		
	Non-malfaisance 8 critères	Restriction à l'accès aux données d'entraînement*	- Matrice des responsabilités - Consentement à la réutilisation des données l'apprentissage de l'IA - Stabilité du modèle d'IA*		
	Justice et équité 3 critères		- Performance de l'IA identique pour tous les publics - Environnement de test de la conformité réglementaire - Minimisation des données vs singularité		
	Autonomie 4 critères	- Dépendance à l'IA - Autonomie des organisations utilisatrices*	- Autonomie décisionnelle - Autonomie pour désactivation de l'IA		
	Eco-responsabilité 5 critères		- Sensibilisation à un usage écoresponsable - Bilan carbone de la fabrication de l'IA - Écoconception de l'IA		- Minimisation des données pour impact environnemental - Lutte contre l'obsolescence des terminaux

* critères spécifiques à l'IA générative

Critères DNS partiellement couverts par le label

Critères DNS totalement couverts par le label

Figure 12 – Comparaison des critères spécifiques DNS et ceux du label Ethik-IA.

Le travail effectué par l'équipe Ethik-IA a également permis d'associer à chaque critère du guide DNS une ou plusieurs des thématiques du label Ethik-IA, et de calculer un pourcentage de couverture par catégorie DNS comme décrit ci-dessous en Figure 13. Les principes de bienfaisance, justice et équité, et éco-responsabilité sont particulièrement bien couverts par le label, avec respectivement 89%, 83% et 100% de couverture des critères DNS associés.

Enfin, il apparaît que certains critères DNS sont couverts par plusieurs thématiques du label Ethik-IA. Quatre critères sont notamment couverts par 3 thématiques différentes (cf. Figure 14) :

- Transparence sur le fait de suivre l'IA
- Transparence sur le fait de ne pas suivre l'IA
- Évaluation continue de l'IA
- Explicabilité de l'IA

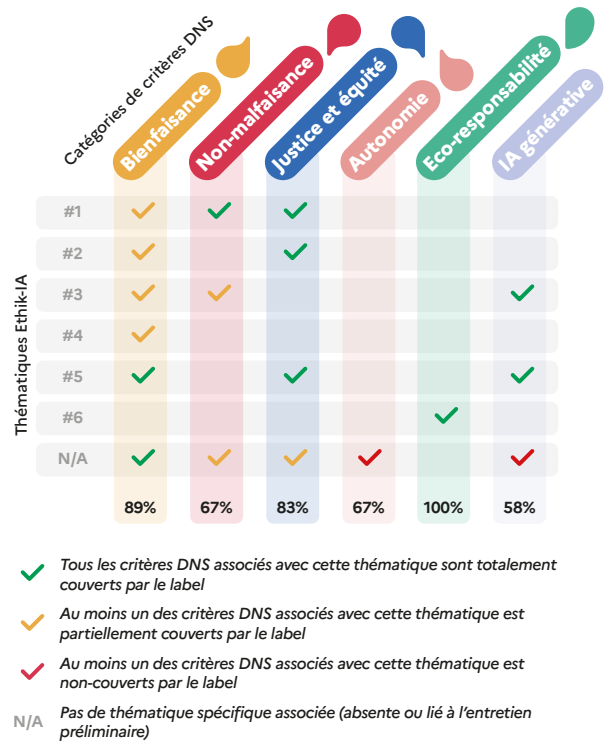


Figure 13 – Association entre les critères du guide DNS par catégories du CENS et les thématiques Ethik-IA.

Thématiques Ethik-IA

?: Critères Ethik-IA non publics

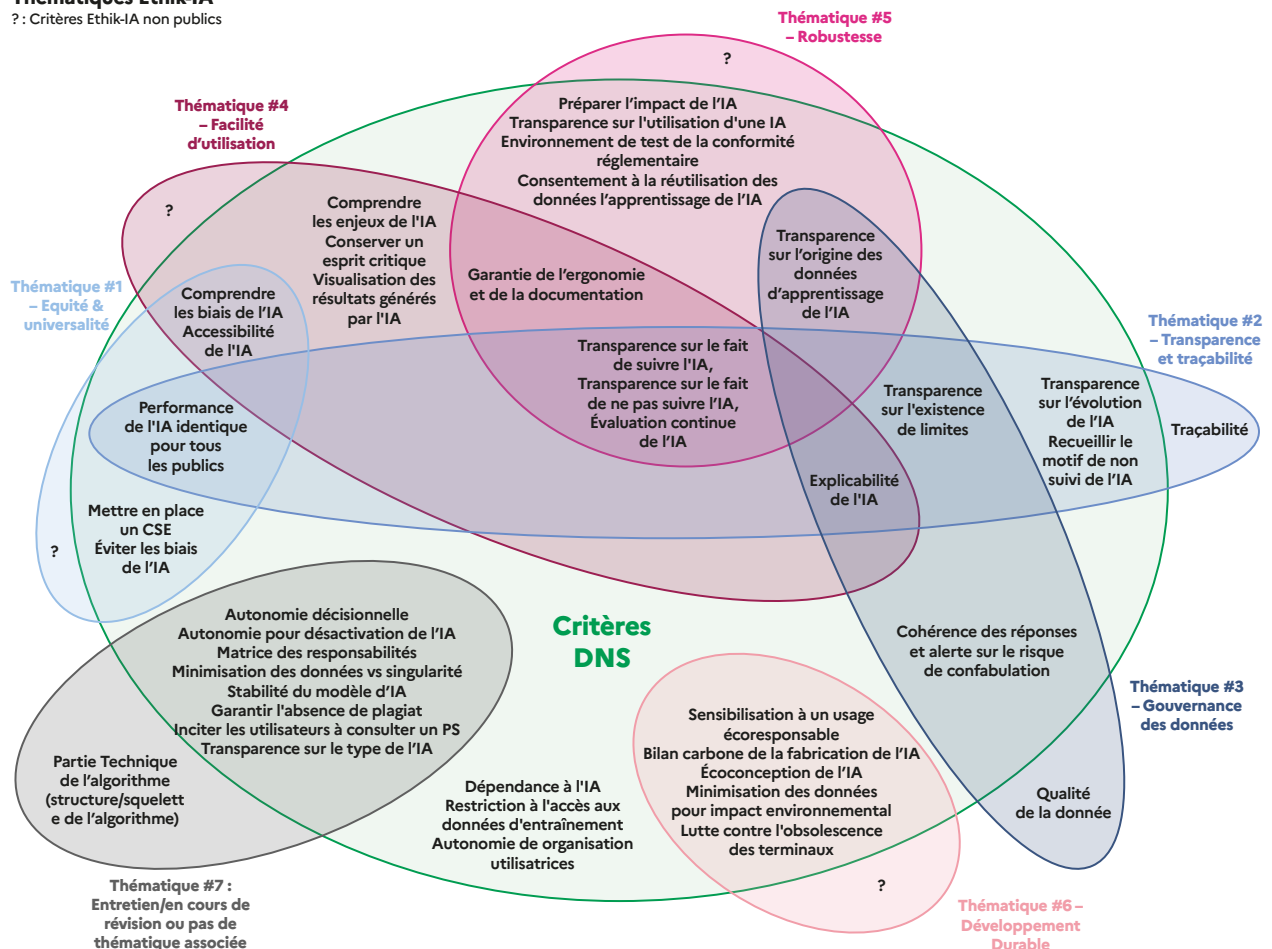


Figure 14 – Projection des critères DNS dans les thématiques du label Ethik-IA.

Label Ekitia

Le label Ekitia est basé sur une grille d'évaluation comportant 102 questions, déclinées selon 6 catégories de principes éthiques pertinents en matière d'usage des données :

- **Principe #1** : Science des données et société
- **Principe #2** : Science des données et individu
- **Principe #3** : Qualité des données et sécurité du système d'information
- **Principe #4** : Transparence
- **Principe #5** : Gouvernance des données dans un cadre de confiance
- **Principe #6** : Réciprocité

Pour ce label, l'analyse comparative a été effectuée à l'aveugle à la fois par l'équipe de la DNS et par l'équipe du label, toujours en suivant la même méthode que celle que décrite précédemment. Les résultats consolidés ont permis de conclure

que 7 critères du guide DNS étaient totalement couverts par le label, 18 étaient partiellement couverts et 13 étaient non couverts par le label, ce qui correspond à un score de recouvrement de 42% du guide DNS (Cf. Figure 15).

Plusieurs observations sont à noter :

- Bien que la grille d'évaluation du label Ekitia couvre un périmètre plus large (usages des données) que celui du guide DNS (IA en santé), avec plus de questions posées (102 questions contre 38 critères pour le référentiel), on observe un taux de recouvrement similaire du guide DNS par le label et du label par le guide DNS (42% et 40% respectivement).

Les critères du guide DNS non couverts par le label portent notamment sur les garanties de **compréhension de l'utilisateur, l'IA générative** (pas de questions spécifiques à l'IA générative dans la grille Ekitia dont le périmètre est plus large : usages des données), et le **secteur de la santé** (le label Ekitia n'étant pas spécifique à ce secteur). La vue complète du recouvrement des critères du guide DNS par le label Ekitia se trouve en Figure 15.

Critères Communs (Focus DNS) 7 totalement couverts, 18 partiellement couverts			
Bienfaisance (18 critères, 47% de couverture par le label) Non-malfaisance (8 critères, 38% de couverture par le label) Justice et équité (3 critères, 50% de couverture par le label) Autonomie (4 critères, 25% de couverture par le label) Eco-responsabilité (5 critères, 40% de couverture par le label)	Critères Spécifiques DNS (13 non couverts) - Transparence sur le fait de suivre l'IA - Transparence sur le fait de ne pas suivre l'IA - Transparence sur l'évolution du module d'IA - Recueillir le motif de non suivi de l'IA - Visualisation des résultats générés par l'IA	- Préparer l'impact de l'IA - Garantir l'ergonomie et de la documentation - Transparence sur l'origine des données d'apprentissage de l'IA - Évaluation continue de l'IA	Critères Spécifiques Ekitia  - En particulier les sous-thématiques suivantes 1.1 Bienfaisance (Intérêt général, finalité du projet). 1.3 Solidarité, diversité, et non-discrimination. 2.1 respect et renforcement de l'autonomie humaine. 3.2 Sécurité et hébergement des données. 4.1. Information claire et accessible sur le projet 5.3 Apprentissage collectif. 5.4 Intégrité Scientifique. 6. Réciprocité.
	- Protection des données d'entraînement* - Garantie de l'absence de plagiat* - Cohérence des réponses et avertissement du risque d'hallucination*	Matrice des responsabilités	
		- Compréhension des enjeux de l'IA - Compréhension des biais de l'IA - Conserver un esprit critique - Transparence sur le type de l'IA - Transparence sur l'existence de limites - Transparence sur l'utilisation d'une IA - Mettre en place un CSE - Explicabilité de l'IA - Accessibilité de l'IA	
		- Performance de l'IA identique pour tous les publics - Minimisation des données vs singularité - Environnement de test de la conformité réglementaire	
	- Dépendance à l'IA - Autonomie pour désactivation de l'IA - Autonomie de l'organisation utilisatrice*	Autonomie décisionnelle	
	- Sensibilisation à un usage écoresponsable - Lutte contre l'obsolescence des terminaux	Minimisation des données pour impact environnemental	- Bilan carbone de la fabrication de l'IA - Ecoconception de l'IA

* critères spécifiques à l'IA générative

Critères DNS partiellement couverts par le label

Critères DNS totalement couverts par le label

Figure 15 – Comparaison des critères spécifiques DNS et ceux du label Ekitia.

Il est à noter que les thématiques du label non couvertes par le guide DNS portent en particulier sur le principe #6 de **réciprocité** (7% de recouvrement par le projet de référentiel), le principe #3 de **qualité des données et sécurité de l'information** (34% de recouvrement) et le principe #2 **concernant la science des données et l'individu** (39% de recouvrement). (Cf. Figure 16).

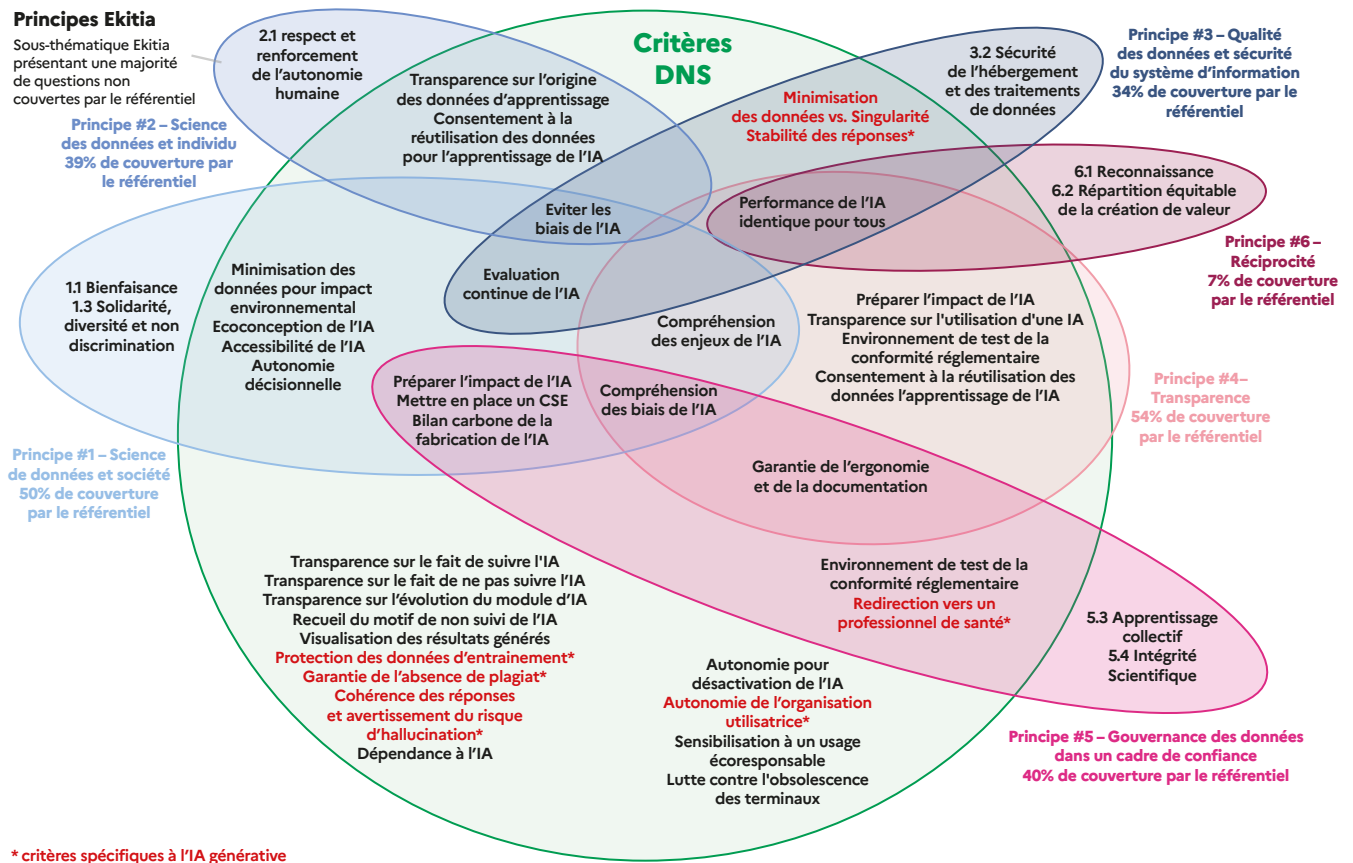


Figure 16 – Projection des critères DNS dans les principes du label Ekitia.

5.7 Limites de l'analyse

La méthodologie de comparaison proposée présente certaines limites :

- L'association entre les critères DNS et les critères d'un label reste un exercice subjectif. Ces pourcentages doivent donc être davantage pris comme des indicateurs qualitatifs du niveau de similarité avec le label plutôt que comme des pourcentages de correspondance stricte.

- Le contexte, la granularité de l'évaluation, les objectifs et destinataires des labels étant parfois différents de ceux du guide DNS, une comparaison trait pour trait n'est pas toujours pertinente. Comme expliqué précédemment, c'est une des raisons qui explique le faible pourcentage de recouvrement du projet de référentiel par certains labels (Labelia, Ethical AI).

Malgré ces limites, cette analyse permet de faciliter la mise en pratique du projet de guide DNS par les fournisseurs de solutions, tout en les encourageant à inscrire leur démarche éthique dans un parcours de labellisation.

Annexe 2 : Règlements et normes associés aux critères du guide d'implémentation de l'éthique dans les SIA en santé et propositions d'éléments d'auto-évaluation

Afin de faciliter leur mise en œuvre, les critères proposés dans ce guide sont présentés suivant l'ordre chronologique des phases de développement d'un SIA, et comprennent, le cas échéant, la référence aux règlements ou normes existant et contenant un principe équivalent (par exemple l'article correspondant du RGPD ou du RIA). Des explications et des éléments d'autoévaluation sont proposés pour chaque critère.

Ainsi, pour chaque critère, sont présentés dans le tableau suivant :

- **Le numéro permettant d'identifier le critère**
- **Le libellé court**
- **Le ou les règlement, référentiel ou norme associé(e.s) et l'article correspondant (le cas échéant)**
- **Des éléments d'autoévaluation**

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
1	Mettre en place un Comité Scientifique et Éthique	N.A.	Afin de s'assurer que le SIA est conçu dans le respect des principes éthiques, le fournisseur met en place un comité scientifique et éthique (CSE) représentatif de la population concernée par le SIA comportant ainsi des utilisateurs du SIA et des personnes pour lesquelles le SIA sera utilisé. Ce comité est mobilisé aux différentes étapes d'élaboration du SIA, à commencer par l'étape de cadrage pour évaluer l'éthique de la finalité du SIA, puis au moment du recueil des données nécessaires à l'apprentissage pour évaluer l'éthique du recueil des données et l'éthique du choix de celles qui seront utilisées, la construction de l'algorithme en vérifiant la séparation entre les populations d'apprentissage et d'évaluation, et le niveau de performance de l'algorithme, ainsi que, le cas échéant, la qualité ergonomique des interfaces homme-machines permettant de vérifier que le SIA est effectivement facilement utilisable par tous les publics.
2	Transparence sur l'origine des données d'apprentissage du SIA	RIA Art. 10 - Données et gouvernance des données Art 47 - Déclaration de conformité de l'UE RGPD Art.6 - Licéité du traitement Art.9 - Traitement portant sur des catégories particulières de données à caractère personnel Art. 13 - Informations à fournir lorsque des données à caractère personnel sont collectées auprès de la personne concernée Loi Informatique et Libertés Art. 44 Art. 66	Lorsque le fournisseur d'un SIA en santé utilise des données de santé issues du soin pour entraîner l'algorithme du SIA, il devrait, au-delà du respect de la base légale prévue par le RGPD, apporter la preuve que les données d'entraînement qu'il utilise ont été obtenues de sorte que les patients dont les données sont utilisées ont été informés de l'utilisation secondaire de leurs données de santé et qu'ils ont compris cette information. Pour rappel, concernant la ré-utilisation secondaire des données de santé, le RGPD s'appuie sur l'opt out traduit en français par un droit d'opposition (traduction officielle). Le Règlement sur l'Espace Européen des Données de Santé (EEDS) s'appuie de la même manière sur l'opt out traduit en français par un droit de non-refus (traduction officielle). Dans les deux cas, il est obligatoire d'avoir informé les patients et d'avoir une base légale de traitement des données. Dans un cadre éthique de l'IA, et concernant la réutilisation secondaire des données de santé, on recommande de solliciter le consentement des patients. Le fournisseur du SIA pourrait utiliser toute méthode lui permettant d'apporter la preuve que l'information sur l'utilisation secondaire de leurs données de santé a été donnée aux patients, qu'ils l'ont bien compris, et, selon le cas, la preuve du consentement.

Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
3 Minimisation des données vs singularité des patients	RIA Art. 10 - Données et gouvernance des données 10.2.e - Les ensembles de données de formation, de validation et d'essai sont soumis à des pratiques de gouvernance et de gestion des données adaptées à la finalité du système d'IA à haut risque. Ces pratiques concernent en particulier : (e) une évaluation de la disponibilité, de la quantité et de l'adéquation des ensembles de données nécessaires ;	La collecte et le traitement des données à caractère personnel sont interdits. Il existe cependant des dérogations au traitement des données à caractère personnel, mais, en accord avec le principe de minimisation des données énoncé dans le RGPD et repris dans le RIA, seules les données strictement nécessaires au traitement devront être utilisées. Aussi, une réflexion doit être conduite par le fournisseur du SIA pour s'assurer que seules les données nécessaires seront utilisées pour l'entraînement, la production et l'exploitation du SIA, tout en s'assurant qu'elles sont suffisantes pour que le SIA soit performant de la même manière pour toutes les personnes de la population cible du SIA. Il s'agit donc pour le fournisseur de démontrer qu'un travail a été mené pour déterminer les données nécessaires et suffisantes, et limiter le recours à ces seules données. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour attester de ce processus de réflexion interne et les conclusions auxquelles il a permis d'aboutir quant à la minimisation des données dans le respect de la singularité : • décrire comment s'est opéré le choix des données finalement sélectionnées avec mention des données finalement non retenues en précisant le volume des données comparées. Par exemple préciser, pour chaque type de données, quelles étaient les données estimées comme requises en première intention, la pertinence de la donnée par rapport à l'objectif de prise en compte de la singularité des patients, et la quantité de données qui ont pu être retirées suite à cette analyse. • produire les comptes rendus des études réalisées en amont du développement du SIA sur ce sujet • produire les comptes rendus des travaux du CSE consacrés à ce sujet. Il est bien sûr possible de conserver des données pour lesquelles la pertinence pour l'apprentissage est hypothétique. Elles devront être identifiées.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
4	Minimisation des données vs impact environnemental	N.A.	Les SIA ont un impact sur l'environnement lié à l'utilisation de données massives pour leur entraînement. Afin de réduire cet impact, il est important de s'interroger sur les données à utiliser pour cet entraînement et à ne retenir que celles qui sont vraiment utiles dans une démarche de sobriété numérique. L'enjeu est ainsi que le fournisseur du SIA puisse établir qu'il a conduit cette réflexion qui consiste à minimiser les données d'apprentissage du SIA afin de ne garder que celles qui sont réellement nécessaires. A cette fin, le fournisseur du SIA pourrait produire, à titre d'exemple, les éléments suivants : • décrire les travaux réalisés pour caractériser la population cible du SIA et les données nécessaires à l'apprentissage • décrire comment s'est opéré le choix des données finalement sélectionnées avec mention des données finalement non retenues en précisant le volume des données comparées. Par exemple préciser, pour chaque type de données, la réflexion conduite sur la réalité du besoin, la détection de redondances, et la quantité de données qui ont pu être retirées suite à cette analyse. • produire les comptes rendus des sessions du CSE ayant traité de cette problématique de minimisation des données, ou des travaux de réflexion conduits avec des experts sur ce sujet. Il est bien sûr possible de conserver des données pour lesquelles la pertinence pour l'apprentissage est hypothétique. Elles devront être identifiées.
5	Réduire et éliminer les biais du SIA	RIA Art. 10 - Données et gouvernance des données 10.2, f et g en particulier - Les ensembles de données de formation, de validation et d'essai sont soumis à des pratiques de gouvernance et de gestion des données adaptées à la finalité du système d'IA à haut risque. Ces pratiques concernent en particulier (f) l'examen en vue d'éventuels biais susceptibles d'affecter la santé et la sécurité des personnes, d'avoir une incidence négative sur les droits fondamentaux ou d'entraîner une discrimination interdite par le droit de l'Union, en particulier lorsque les données de sortie influencent les données d'entrée pour des opérations futures ; (g) des mesures appropriées pour détecter, prévenir et atténuer les éventuels biais identifiés conformément au point f) ;	Les SIA comportent de nombreux biais : biais de représentativité quand les données d'entraînement ne représentent pas la diversité des patients, biais de sélection quand les échantillons de données sont choisis selon des critères qui excluent certaines populations, biais algorithmiques quand les décisions sont influencées par des paramètres ou méthodes de modélisation inadéquats, biais d'interprétation quand les résultats sont mal compris ou surinterprétés par les professionnels de santé, biais socio-économiques qui reproduisent les inégalités sociales dans l'accès ou la qualité des soins, biais historiques quand les données d'apprentissage sont issues de pratiques médicales passées contenant elles-mêmes des discriminations, etc. Il convient de minimiser, idéalement d'éliminer, ces biais afin de garantir un niveau de performance identique au sein de la population cible du SIA. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il a mis en œuvre les mécanismes appropriés pour réduire et, idéalement, éliminer les biais de l'IA : • analyse préalable des biais et des corrections à mettre en œuvre pour les réduire dans le cadre du développement du SIA, notamment adaptation du périmètre de la population cible compte tenu des données collectées pour l'entraînement de l'algorithme • description des mécanismes de réduction / élimination des biais mis en œuvre • comptes rendus des travaux du CSE consacrés à ce sujet • tests effectués pour évaluer les biais résiduels du SIA ou pour démontrer que le SIA n'est pas biaisé.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
6	Explicabilité du SIA	N.A.	Lorsqu'un SIA en santé propose une aide à la décision, il est important que l'utilisateur puisse obtenir des explications relatives à la proposition générée par l'IA afin de disposer de tous les éléments dont il a besoin pour prendre une décision éclairée sur la base des éléments fournis par l'outil et des éléments dont il dispose par ailleurs. L'adhésion des professionnels de santé aux propositions de l'IA, parce qu'ils restent responsables de leurs décisions, ne peut s'envisager que si les propositions de l'IA leur sont expliquées. Lorsque la proposition générée par l'IA n'est pas explicable, il est tout aussi important, par transparence vis à vis des utilisateurs du SIA, de le dire aux utilisateurs et de leur fournir des informations sur le modèle d'IA utilisé en indiquant pourquoi les propositions générées ne peuvent pas être expliquées. Au-delà de donner des explications aux propositions d'un SIA en santé, il convient également de vérifier que les explications proposées sont correctes et jugées satisfaisantes, c'est-à-dire, compréhensibles et convaincantes par les utilisateurs. Pour cela, le fournisseur du SIA en santé pourrait produire les éléments suivants pour montrer que l'interface du système permet de demander des explications et que ces explications sont pertinentes, ou que le SIA ne peut pas donner d'explications et pourquoi. 1) L'interface permet de demander des explications : • Production des copies d'écran montrant que la fonctionnalité est présente et que l'utilisateur a la possibilité d'accéder en un clic à une explication pour toute proposition générée par l'IA • Les preuves pourraient être complétées par les résultats des tests effectués auprès d'un panel d'utilisateurs montrant qu'ils ont bien identifié la possibilité d'afficher une explication 2) Les explications sont jugées pertinentes : • Production des copies d'écran montrant des exemples d'explications de réponses fournies par le système lorsqu'une explication est demandée • Les preuves pourraient être complétées par la présentation des moyens mis en œuvre pour vérifier que les explications fournies ont été jugées satisfaisantes du point de vue de l'utilisateur, en incluant par exemple l'analyse d'un ensemble de cas d'usage par le CSE ou un panel d'utilisateurs, ou la possibilité pour l'utilisateur de faire un retour facilement lorsqu'une explication ne lui donne pas satisfaction et les éléments collectés via ces retours. 3) Si la proposition du SIA n'est pas explicable, production des éléments de justification de la non explicabilité : • les éléments de fonctionnement du SIA, tels que les méthodes utilisées pour entraîner l'algorithme justifiant pourquoi la proposition générée n'est pas explicable, mais permettant à l'utilisateur de comprendre comment elle a été générée • les moyens mis en œuvre par le fournisseur du SIA pour s'assurer que les éléments d'explication apportés dans ce cas sont satisfaisants du point de vue de l'utilisateur, en incluant par exemple l'analyse d'un ensemble de cas d'usage par le CSE ou un panel d'utilisateurs.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
7	Empêcher l'IA générative de restituer des données à caractère personnel qui auraient été utilisées pour son entraînement (G)	N.A.	Des études scientifiques ont montré que lorsqu'une IA générative était entraînée à partir de ressources incluant des données à caractère personnel (par exemple des cas cliniques nominatifs, des dossiers patients, ...), il était possible d'obtenir dans les réponses de l'IA générative aux questions qui lui étaient posées, des extraits de ces ressources comportant donc des données à caractère personnel. L'objectif de ce critère est d'alerter sur cette possibilité de façon à ce que le fournisseur du SIA se donne les moyens de rendre ceci impossible. Il est attendu que le fournisseur du SIA explicite les moyens mis en œuvre pour éviter toute divulgation d'informations à caractère personnel dans les propositions générées par le SIA. Cela pourrait, par exemple, passer par un apprentissage effectué sans utiliser de données à caractère personnel ou avec des données à caractère personnel ayant été anonymisées au préalable, ou bien si l'apprentissage est fait sur des ressources incluant des données à caractère personnel, que des mécanismes de filtrage soient intégrés en aval pour détecter et empêcher la divulgation de données à caractère personnel dans les réponses générées. Le fournisseur pourrait fournir les éléments précisant dans quel cas de figure il se trouve (apprentissage effectué sans utiliser de données à caractère personnel ou avec données à caractère personnel anonymisées, ou mécanismes de filtrage en post-traitement), ainsi que des exemples concrets montrant comment le SIA identifie et filtre les informations sensibles le cas échéant.
8	Garantir l'absence de plagiat avec les IA génératives (G)	N.A.	Une IA générative peut être amenée à reprendre à l'identique des extraits de textes sur lesquelles elle a effectué son apprentissage. Si des ressources telles que des textes d'auteurs ont été utilisées lors de son apprentissage, le fournisseur du SIA devrait se donner les moyens de garantir l'absence de plagiat en empêchant que les réponses à des questions posées ne contiennent pas des recopies de tout ou partie d'un texte d'auteur sans indiquer explicitement qu'il s'agit d'une citation. Cela peut passer tout d'abord par l'absence de textes d'auteurs dans les données d'apprentissage. Si des textes d'auteurs sont utilisés, ils doivent être caractérisés en tant que tels et des mécanismes de détection et de filtrage ou de citation avec référencement des sources doivent être mis en œuvre lors de la génération de propositions par l'IA. Le fournisseur du SIA pourrait préciser les données utilisées pour l'apprentissage pour montrer, selon les cas : • qu'elles ne contiennent pas de texte d'auteurs, • ou que les textes d'auteurs sont identifiés en tant que tels, et que des filtres sont utilisés en aval pour éviter qu'ils ne soient repris, tout ou partie, dans les propositions générées par l'IA, ou qu'ils le soient mais sous forme de citations avec référencement des sources. Il pourrait également fournir les résultats des tests effectués pour montrer l'efficacité des algorithmes de détection des contenus issus de textes d'auteurs, avec des exemples de cas où les réponses incluent des citations correctement référencées. Il pourra aussi fournir les comptes rendus des séances de travail sur le sujet incluant par exemple l'analyse d'un ensemble de cas d'usage par le CSE ou par un panel d'utilisateurs.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
9	Autonomie pour la désactivation d'un SIA modulaire	RIA - Art. 14.4.e Art. 14.4 : Surveillance humaine Aux fins de la mise en œuvre des paragraphes 1, 2 et 3, le système d'IA à haut risque est fourni au responsable du déploiement de telle sorte que les personnes physiques auxquelles la surveillance humaine est confiée soient habilitées, selon ce qui est approprié et proportionné : (e) intervenir dans le fonctionnement du système d'IA à haut risque ou interrompre le système au moyen d'un bouton «stop» ou d'une procédure similaire permettant au système de s'arrêter dans un état sûr.	Dans le cas d'un SIA modulaire en santé, l'utilisateur, qu'il soit patient ou professionnel de santé, doit pouvoir désactiver de manière autonome le SIA sans porter atteinte au fonctionnement des autres applications du SI. Afin d'apporter la preuve de ce critère, le fournisseur pourrait, à titre d'exemple, fournir les éléments suivants pour montrer que le SIA modulaire est effectivement facilement désactivable par tout utilisateur qui peut donc le faire par lui-même : - le mode d'emploi ou les règles d'utilisation mentionnant les modalités de désactivation du SIA par l'utilisateur, précisant les conditions pour cette désactivation et les conséquences sur l'utilisation ultérieure du système - les supports de formation ou d'information explicitant les modalités de désactivation du SIA - les copies d'écran ou vidéo de cas d'usage où l'utilisateur désactive le SIA - les comptes rendus des séances du CSE dédiées à tester l'utilisabilité de la désactivation du SIA.
10	Écoconception du SIA et lutte contre l'obsolescence des équipements	RGESN https://ecoresponsable numerique.gouv.fr/publications/referentiel-general-ecoconception	Le fournisseur d'un SIA devrait mettre en place une démarche d'écoconception du SIA, en s'appuyant sur les bonnes pratiques énoncées dans le référentiel général d'écoconception des services numériques (RGESN) de la DINUM. Il pourrait fournir les éléments suivants pour démontrer qu'il a respecté, autant que possible, les bonnes pratiques du référentiel : - les résultats de l'autoévaluation réalisée par le fournisseur afin d'évaluer la conformité du SIA au RGESN, indiquant la liste des bonnes pratiques du RGESN qui sont vérifiées par le SIA et la liste de celles qui ne le sont pas en indiquant pourquoi - l'analyse des bonnes pratiques du RGESN non vérifiées, effectuée en interne par le fournisseur, et les actions d'amélioration qu'il a définies dans le cadre d'une démarche d'amélioration de la conformité au RGESN - les comptes rendus des travaux du CSE relatifs à l'écoconception et à l'utilisation écoresponsable de l'IA - le descriptif des prérequis matériels de la solution précisant les options retenues pour les optimiser, avec l'assurance qu'elle fonctionne sur des équipements autres que ceux de dernière génération, afin de limiter l'obsolescence des terminaux utilisateurs.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
11	Bilan carbone de la fabrication du SIA	Spec AFNOR 2134	Le fournisseur d'un SIA en santé devrait réaliser un bilan carbone de sa solution afin d'être en capacité de minimiser son impact carbone. Il peut pour cela suivre la spec AFNOR 2134 qui établit une méthodologie d'évaluation des impacts environnementaux avec une analyse du cycle de vie, avec la proposition de 31 fiches de bonnes pratiques et des recommandations pour communiquer avec justesse sur le caractère frugal d'un service d'IA. Le fournisseur pourrait ainsi, à titre d'exemple, fournir les éléments suivants pour attester de la réalisation du bilan carbone de sa solution : • le document de synthèse du bilan carbone réalisé • l'attestation de bilan carbone réalisé par un tiers • l'impact carbone du SIA en intégrant toutes les phases du cycle de vie du SIA. Lorsque le fournisseur du SIA s'appuie sur un modèle d'IA sous-jacent fourni par un tiers (par exemple de type «AI as a Service»), il devrait intégrer le bilan carbone de l'élaboration du modèle d'IA sous-jacent dans les externalités de son bilan carbone. La publication des bilans carbone des modèles d'IA devrait d'ailleurs permettre au fournisseur du SIA de choisir le modèle d'IA ayant le plus petit impact.
12	Transparence sur l'interaction avec une IA	RIA - Art. 50.1 Art. 50 : Obligations de transparence pour les fournisseurs et les déployeurs de certains systèmes d'IA Art. 50.1 - Les fournisseurs veillent à ce que les systèmes d'IA destinés à interagir directement avec des personnes physiques soient conçus et développés de manière à ce que les personnes physiques concernées soient informées qu'elles interagissent avec un système d'IA, à moins que cela ne soit évident du point de vue d'une personne physique raisonnablement bien informée, observatrice et avisée, compte tenu des circonstances et du contexte d'utilisation.	Tel que mentionné dans le RIA, l'utilisateur d'un SIA en santé devrait être informé qu'il interagit avec une IA au moment de l'utilisation du SIA. Le fournisseur pourrait, à titre d'exemple, fournir les éléments suivants pour démontrer la transparence du SIA sur l'interaction de l'utilisateur avec une IA : • les supports d'information ou de formation transmis à l'utilisateur avant son utilisation du SIA, précisant de manière adaptée au public visé le fait qu'il interagit avec une IA • dans le cas d'un SIA disposant d'une interface utilisateur, des copies d'écran montrant les messages d'avertissement affichés à l'utilisateur l'informant qu'il interagit avec une IA.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
13	Vérifier la bonne compréhension de l'interaction avec une IA	N.A.	Selon le public concerné et le contexte d'utilisation du SIA, il se peut que l'utilisateur n'ait pas clairement compris qu'il interagissait avec une IA, malgré le respect du critère précédent relatif à la transparence sur l'interaction avec une IA. Il convient ainsi que le fournisseur du SIA se donne les moyens de vérifier la bonne compréhension de cette information par le public cible d'utilisateurs, qu'il s'agisse de professionnels de santé ou de patients, en tenant compte de la diversité des profils des utilisateurs et des messages leur indiquant qu'ils interagissent avec une IA. Pour cela, le fournisseur du SIA pourrait produire, à titre d'exemple, les éléments suivants pour démontrer qu'il s'est assuré que les informations communiquées lors de l'utilisation du SIA sur le fait d'être en interaction avec une IA ont été comprises par tout type d'utilisateur : • les comptes rendus des réunions du CSE ou d'un panel représentatif des utilisateurs du SIA consacrées à l'analyse de l'interface utilisateur afin d'évaluer si l'information relative à l'interaction avec une IA est effectivement visible, compréhensible et adaptée à la diversité des publics cibles. • les comptes rendus des réunions du CSE ou d'un panel représentatif des utilisateurs du SIA consacrées à l'étude de leur bonne compréhension du fait qu'ils interagissent avec une IA. • la caractérisation des profils d'utilisateurs et des mesures mises en œuvre pour tenir compte de leurs éventuelles difficultés à accéder ou comprendre l'information. • les résultats de tests utilisateurs montrant que la majorité des profils représentatifs identifient correctement leur interaction avec une IA.
14	Marquer les résultats de l'IA	RIA - Art.50.2 et 50.4 Art. 50 : Obligations de transparence pour les fournisseurs et les déployeurs de certains systèmes d'IA Art. 50.2 - Les fournisseurs de systèmes d'IA, y compris les systèmes d'IA à usage général, qui génèrent des contenus synthétiques audio, image, vidéo ou texte, veillent à ce que les résultats du système d'IA soient marqués dans un format lisible par machine et détectables comme étant générés ou manipulés artificiellement. [...] Art. 50.4 - Les personnes qui déploient un système d'IA qui génère ou manipule un contenu image, audio ou vidéo constituant une contrefaçon profonde doivent indiquer que le contenu a été généré ou manipulé artificiellement. [...]	Conformément au RIA, le fournisseur d'un SIA en santé doit permettre d'identifier que des résultats ont été produits par une IA. A cette fin, le fournisseur du SIA pourrait fournir, à titre d'exemple, les éléments suivants pour montrer qu'il a mis en place des mécanismes permettant de rendre visibles les résultats qui ont été produits par l'IA : • les règles de conception des interfaces mentionnant l'affichage spécifique des résultats générés par l'IA • des copies d'écran montrant que les résultats générés par IA sont bien affichés de manière visible, permettant de les identifier parmi les autres éléments affichés • les résultats de tests utilisateurs qui montrent que l'utilisateur identifie facilement et correctement les résultats produits par l'IA • les comptes rendus des travaux du CSE consacrés à ce sujet.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
15	Traçabilité de la désactivation du SIA	RIA Art. 12 - Enregistrement Art. 12.1. Les systèmes d'IA à haut risque permettent, techniquement, l'enregistrement automatique des événements (journaux) tout au long de la durée de vie du système. Art. 12.3.a - Pour les systèmes d'IA à haut risque visés à l'annexe III, point 1 a), les fonctionnalités de journalisation fournissent, au minimum : a) l'enregistrement de la période de chaque utilisation du système (date et heure de début et de fin pour chaque utilisation)	Conformément au RIA, toute désactivation de l'IA doit pouvoir être tracée afin de pouvoir accéder ultérieurement à l'information de la désactivation de l'IA, par qui, dans quelles circonstances, et pour combien de temps. Le fournisseur pourrait fournir, à titre d'exemple, les éléments suivants pour montrer qu'il a mis en place les mécanismes permettant de tracer la désactivation de l'IA : - le descriptif des mécanismes mis en place - des exemples de journal de désactivation permettant de visualiser les éléments sur les désactivations effectuées sur une période donnée
16	Garantir l'autonomie décisionnelle de l'utilisateur	RIA - Art. 14.4.d Art. 14 : Surveillance humaine Art. 14.4. Aux fins de la mise en œuvre des paragraphes 1, 2 et 3, le système d'IA à haut risque est fourni au responsable du déploiement de telle sorte que les personnes physiques auxquelles la surveillance humaine est confiée soient habilitées, selon ce qui est approprié et proportionné : (d) de décider, dans une situation donnée, de ne pas utiliser le système d'IA à haut risque ou de ne pas tenir compte des résultats du système d'IA à haut risque, de les ignorer ou de les inverser	Dans le cas d'un SIA d'aide à la décision, qui dispose d'une interface utilisateur, le SIA fait des propositions que l'utilisateur reste toujours libre de suivre ou pas. Aussi, le SIA est conçu de sorte que l'utilisateur conserve toujours son autonomie décisionnelle, conformément au RIA. Le fournisseur pourrait fournir, à titre d'exemple, les éléments suivants pour montrer que le SIA garantit l'autonomie décisionnelle de l'utilisateur : - les interfaces montrant que la proposition de l'IA est bien formulée de sorte à être comprise comme une proposition, pouvant être adoptée ou rejetée - des copies d'écran ou vidéos de cas d'usage où l'utilisateur choisit de ne pas utiliser l'IA ou de ne pas suivre sa recommandation, montrant que l'utilisation de l'outil est fluide de la même manière dans ces deux cas de figure - les résultats de tests utilisateurs montrant que l'utilisateur est bien autonome dans sa prise de décision - les comptes rendus des travaux du CSE à ce sujet.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
17	Transparence sur le fait de suivre les propositions du SIA	N.A.	Dans le cadre du suivi d'un patient, le fait qu'une décision diagnostique ou thérapeutique ait été prise par un professionnel de santé en utilisant un SIA, en suivant la proposition générée par le SIA, et que ceci ait été communiqué au patient devrait pouvoir être conservé dans le dossier du patient dans un objectif de traçabilité. L'enjeu est que le fournisseur du SIA intègre des mécanismes permettant de tracer qu'une décision diagnostique ou thérapeutique ait été prise en suivant les propositions du SIA. Le fournisseur du SIA pourrait produire, à titre d'exemple, les éléments suivants pour montrer que de tels mécanismes sont en place pour tracer ces informations : - des captures d'écran issues de tests illustrant des situations où une décision médicale a été prise en suivant la proposition du SIA, et montrant que cette information peut être conservée, affichée a posteriori et intégrée dans le dossier du patient - des exemples de journal d'audit anonymisés démontrant comment l'utilisation du SIA est tracée pour chaque patient (incluant le recours à l'IA et le fait d'avoir suivi ou non la proposition du SIA), et comment cette information est accessible et exploitable a posteriori.
18	Transparence sur le fait de ne pas suivre les propositions du SIA et sur le motif de non suivi	N.A.	Dans le cadre du suivi d'un patient, le fait qu'une décision diagnostique ou thérapeutique ait été prise par un professionnel de santé en utilisant un SIA, mais en ne suivant pas la proposition générée par le SIA, et que ceci ait été communiqué au patient devrait pouvoir être conservé dans le dossier du patient dans un objectif de traçabilité. L'enjeu est que le fournisseur du SIA intègre des mécanismes permettant de tracer qu'une décision diagnostique ou thérapeutique ait été prise en ne suivant pas les propositions du SIA, avec le motif de non suivi. Le fournisseur du SIA pourrait produire, à titre d'exemple, les éléments suivants pour montrer que de tels mécanismes sont en place pour tracer ces informations : - des captures d'écran issues de tests illustrant des situations où une décision médicale a été prise en ne suivant pas la proposition du SIA, et montrant qu'un motif de non suivi doit nécessairement être renseigné, que ces informations, à savoir la décision prise et le motif de non suivi, peuvent être conservées, affichées a posteriori et intégrées dans le dossier du patient - des exemples de journal d'audit anonymisés démontrant comment l'utilisation du SIA est tracée pour chaque patient, (incluant le recours à l'IA et le fait d'avoir suivi ou non la proposition du SIA, et d'avoir renseigné un motif en cas de non suivi), et comment cette information est accessible et exploitable a posteriori.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
19	Transparence sur le fait de ne pas avoir utilisé le SIA	N.A.	Dans certains cas, le professionnel de santé peut choisir de ne pas recourir à l'IA, et ce choix devrait pouvoir être communiqué au patient et conservé dans son dossier dans un objectif de traçabilité. L'enjeu est que le fournisseur du SIA intègre des mécanismes permettant de tracer qu'une décision diagnostique ou thérapeutique ait été prise sans utilisation de l'IA. Le fournisseur du SIA pourrait produire, à titre d'exemple, les éléments suivants pour montrer que de tels mécanismes sont en place pour tracer cette information : - des captures d'écran issues de tests illustrant des situations où une décision médicale a été prise sans utiliser l'IA, que cette information peut être conservée, affichée a posteriori et intégrée dans le dossier du patient - des exemples de journal d'audit anonymisés montrant comment la non-utilisation de l'IA est enregistrée pour chaque patient, et comment cette information est accessible et exploitable a posteriori.
20	Conformité RGAA des SIA avec une interface utilisateurs	RGAA - Référentiel Général d'Amélioration de l'Accessibilité (https://accessibilite.numérique.gouv.fr) RIA Art. 16 - Obligation des fournisseurs de systèmes d'IA à haut risque Art. 9.9 – Prise en compte des personnes vulnérables	Dans le cas d'un SIA en santé avec une interface homme-machine, en conformité avec le RGAA, et en accord avec le RIA, l'interface doit garantir que le SIA est accessible à tout type d'utilisateur. Dans ce cadre, le fournisseur d'un SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il est en conformité avec le RGAA : - les résultats de l'autoévaluation réalisée par le fournisseur afin d'évaluer la conformité du SIA au RGAA, indiquant la liste des critères du RGAA qui sont vérifiés par le SIA et la liste de ceux qui ne le sont pas en indiquant pourquoi - les mécanismes mis en place pour garantir l'accessibilité (règles, normes, ou chartes suivies pour la conception des interfaces) - les copies d'écran ou vidéos montrant des cas d'usage avec des problématiques d'accessibilité - les résultats de tests utilisateurs réalisés avec des publics ayant des problématiques d'accessibilité diverses - les comptes rendus des travaux du CSE sur l'ergonomie des interfaces et l'accessibilité du SIA à tous les publics.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
21	Intelligibilité des interfaces des SIA comportant une interface utilisateurs	N.A.	Dans le cas d'un SIA en santé avec une interface homme-machine, la manière dont sont conçues les interfaces influence directement la manière dont l'outil va être utilisé et pourrait induire des usages vertueux ou, au contraire, engendrer un risque de dérive ou d'usage inapproprié dont il convient de se prémunir. L'enjeu est que le fournisseur du SIA se donne les moyens de s'assurer que les interfaces du SIA sont appropriées pour garantir un bon usage du SIA par tout type d'utilisateur. Pour ce faire, le fournisseur du SIA pourrait, par exemple, fournir les éléments suivants pour montrer qu'il a mis en place une démarche d'évaluation et d'amélioration de l'utilisabilité du SIA par tout type d'utilisateur : • les moyens mis en œuvre dans le cadre d'un processus de co-conception itératif pour tester des prototypes d'interfaces auprès de panels représentatifs d'utilisateurs et ajuster les interfaces sur la base des résultats des tests • les évaluations réalisées auprès des utilisateurs du SIA sur l'ergonomie et l'intelligibilité des interfaces, au travers, par exemple, de tests utilisateurs, d'enquêtes ou de travaux du CSE.
22	Cohérence des réponses dans la répétition d'un SIA générative (G)	N.A.	Un système d'IA générative interrogé à plusieurs reprises sur un sujet donné génère des réponses qui peuvent sensiblement varier. L'enjeu est de s'assurer que le SIA génère des propositions qui sont effectivement cohérentes entre elles s'il est interrogé plusieurs fois pour un cas similaire voire sur la même question. A cette fin, le fournisseur du SIA pourrait produire, à titre d'exemple, les éléments suivants montrant qu'il s'est donné les moyens de vérifier la cohérence des réponses dans la répétition : • les rapports de tests effectués pour prouver la cohérence des réponses dans la répétition, tels que des tests réalisés en parallèle sur plusieurs dispositifs utilisant le SIA et effectuant une requête identique concernant un même patient, ou réalisés à quelques minutes d'intervalle sur un même dispositif • des captures d'écran ou vidéos de l'interface montrant la date, les requêtes effectuées et les réponses affichées.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
23	Qualité des réponses d'un SIA générative (G)	N.A.	Une limite fréquemment observée des SIA générative est le fait de proposer, à une question pour laquelle le SIA ne trouve pas de réponse, une réponse absurde («hallucination»), plutôt que de dire qu'il ne sait pas répondre. Le taux d'hallucinations des IA génératives varie généralement entre 3 % et 25 %, mais peut monter jusqu'à 50 % dans des contextes sensibles ou mal contrôlés (références scientifiques, résumés d'articles, réponses médicales). L'enjeu est ici que le SIA générative soit programmé pour éviter de telles hallucinations. Le fournisseur du SIA générative pourrait, par exemple, fournir les éléments suivants pour montrer qu'il a mis en oeuvre les mesures nécessaires pour garantir la qualité des réponses fournies, éviter que le SIA ne produise des hallucinations et avertir l'utilisateur sur ce risque : • description des méthodes utilisées pour détecter et signaler les hallucinations ou incohérences dans les réponses de l'IA • description de la démarche de validation des réponses générées, incluant les tests de qualité et de fiabilité mis en oeuvre, tels que l'évaluation par des métriques de précision, la revue par des experts humains, des mécanismes de détection et de correction des hallucinations avant présentation à l'utilisateur, ou encore la capacité du système à indiquer qu'il ne sait pas répondre lorsqu'un risque d'hallucination est identifié. Cette démarche peut également inclure l'utilisation d'un autre LLM pour vérifier la qualité des réponses produites par le SIA. • description des systèmes d'avertissement et d'information sur les risques résiduels d'hallucination, avec par exemple des captures d'écran des messages affichés à l'intention de l'utilisateur. • description du système mis en place pour recueillir des retours d'utilisateurs en cas de suspicion d'hallucination afin d'améliorer la solution en continu, et éléments permettant de montrer que le système mis en place est approprié du point de vue des utilisateurs, comme des enquêtes utilisateurs, ou des tests effectués auprès de panels d'utilisateurs, ou un rapport des signalements collectés.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
24	Garantir que le SIA générative incite le patient utilisateur à consulter un professionnel de santé en cas d'alerte (G)	MDR Annexe I point 22 : Protection contre les risques émanant des dispositifs destinés par le fabricant à des profanes point 23.4 w : pour les dispositifs destinés à être utilisés par des profanes, les circonstances dans lesquelles l'utilisateur devrait consulter un professionnel de la santé	Dans le cas d'un système d'IA générative à destination du patient, il est essentiel que le système incite le patient à consulter un professionnel de santé lorsque les questions posées par le patient permettent de détecter une situation d'urgence sanitaire pour le patient. A cette fin, le fournisseur du SIA générative pourrait, à titre d'exemple, fournir les éléments suivants pour montrer que le SIA générative incite effectivement à consulter un professionnel de santé en cas d'alerte : - la preuve de la détection des situations d'urgence pour lesquelles l'incitation à consulter est générée sur une base de cas simulés avec des copies d'écran ou vidéos montrant que le patient est invité à consulter - les messages associés à toute génération d'alerte sur l'interface utilisateur et les explications de l'état d'urgence - les résultats de tests utilisateurs montrant que l'utilisateur a compris la situation et qu'il lui est recommandé de consulter - les comptes rendus des travaux réalisés par le CSE à ce sujet. Les propositions générées par l'IA peuvent être soumises à des biais qui peuvent conduire à une inéquité dans le traitement des patients selon leurs profils. Une façon de répondre à cette limite est de définir, en toute transparence, la population cible d'un SIA, c'est-à-dire, la population sur laquelle l'algorithme a été entraîné et au sein de laquelle la performance est garantie par le fournisseur du SIA. L'enjeu est alors de s'assurer que le SIA génère des propositions avec le même niveau de performance pour toutes les personnes au sein de la population cible. La solution est d'avoir une performance moyenne à la hauteur de ce que le fournisseur du SIA annonce avec de très faibles écarts types. Pour cela, le fournisseur du SIA pourrait, à titre d'exemple, produire les éléments suivants pour montrer qu'il a mis en place des mécanismes permettant d'offrir le même niveau de performance de l'IA pour tous les profils de patients de la population cible : - décrire la méthode utilisée pour caractériser le périmètre de la population cible et la performance du SIA sur ce périmètre - produire les résultats de tests d'équité montrant que sur un ensemble de cas simulés, mais représentatifs, de la population cible, le système fournit des propositions dont la distribution de la validité est homogène - produire des études de validation indépendantes confirmant l'absence de biais systématiques dans les décisions générées par le système (collège de garantie humaine) - décrire le processus de surveillance continue mis en place pour détecter et corriger les biais émergents du SIA.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
25	Performance du SIA identique pour tous les publics couverts par la population cible	N.A.	Le déploiement d'un SIA en santé au niveau d'une structure de soin va générer des impacts importants sur l'organisation des soins, les pratiques cliniques, les relations entre les professionnels du soin et avec les patients. Ces impacts doivent être anticipés afin d'être, le cas échéant, modérés. Le fournisseur du SIA, de par sa connaissance approfondie de son système, et les retours d'expériences de ses différents clients/utilisateurs, dispose d'éléments particulièrement pertinents à apporter à tout utilisateur / structure dans le cadre de la préparation du déploiement d'un SIA. L'enjeu est que le fournisseur du SIA propose systématiquement un accompagnement à toute entité souhaitant déployer le SIA afin de conduire conjointement une analyse de risques. Le refus de cet accompagnement par les utilisateurs / la structure doit être tracé. A cette fin, le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il propose systématiquement un programme d'accompagnement au déploiement, que ce programme répond aux attentes des utilisateurs et que le refus de cet accompagnement par les utilisateurs / la structure est tracé : • la description du programme d'accompagnement mis en place, précisant les sessions et les profils des personnes y participant • des exemples d'analyses de risque effectuées en amont du déploiement, la réflexion conduite avec la structure sur la base de ces exemples, l'identification des impacts propres à la structure en termes de risque sur la qualité et la continuité des soins, les nouvelles règles de coordination établies entre les professionnels, la mesure de la satisfaction des patients, ainsi que les pistes de réponse ayant été collectivement définies afin de réduire les impacts négatifs du déploiement du SIA • les retours des bénéficiaires du programme d'accompagnement collectés dans un questionnaire de satisfaction à l'issue du programme • le cas échéant, l'éventuelle confirmation par la structure qu'elle ne souhaite pas bénéficier de l'accompagnement du fournisseur pour cette analyse d'impact en amont du déploiement.
26	Accompagner les utilisateurs/la structure à l'impact du déploiement d'un SIA	N.A.	Afin de satisfaire aux exigences du RIA, les ressources de formation aux enjeux de l'IA devraient être co-construites avec les utilisateurs du SIA. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il a conçu ces ressources avec un panel représentatif des utilisateurs du SIA : • la composition du panel d'utilisateurs impliqués dans la co-construction des supports de formation • le plan des ressources co-construites afin de valider que le contenu couvre bien l'ensemble des enjeux de l'IA • les comptes rendus des travaux effectués par le CSE sur le sujet

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
27	Co-construire avec les utilisateurs les ressources de formation aux enjeux de l'IA	RIA Art. 4 – Maîtrise de l'IA Art. 13 – transparence et fourniture d'information au déployeur	Afin de satisfaire aux exigences du RIA, les ressources de formation aux enjeux de l'IA devraient être co-construites avec les utilisateurs du SIA. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il a conçu ces ressources avec un panel représentatif des utilisateurs du SIA : - la composition du panel d'utilisateurs impliqués dans la co-construction des supports de formation - le plan des ressources co-construites afin de valider que le contenu couvre bien l'ensemble des enjeux de l'IA - les comptes rendus des travaux effectués par le CSE sur le sujet
28	Former les utilisateurs aux enjeux de l'IA (biais, limites de performance)	MDR Annexe I point 4.c Annexe I point 23.4 (points b,c,e,g,h,i,s) information de la notice par rapport à l'utilisation du DM et les risques associés Annexe I point 23.4 (j) formation requise ou qualification de l'utilisateur ou autre.	Comme pour tout dispositif médical, et conformément au MDR, les utilisateurs d'un SIA de santé doivent être formés aux enjeux spécifiques de l'utilisation du dispositif, ici aux enjeux spécifiques de l'IA. Il est de la responsabilité du fournisseur d'un SIA de former les utilisateurs mais aussi les déployeurs aux enjeux spécifiques liés au recours à l'IA, en particulier les biais et les limites de performance. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il a formé les utilisateurs et déployeurs à ces enjeux : - le programme de formation qui accompagne le déploiement du SIA, avec les éléments sur les limites de performances et les biais de l'IA - le détail des sessions de formation réalisées auprès des utilisateurs, ainsi que les retours des utilisateurs sur les formations reçues - les comptes rendus des sessions du CSE traitant de cette formation.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
29	Vérifier la bonne compréhension par les utilisateurs des enjeux de l'IA en particulier les limites de performance	N.A.	L'utilisateur doit avoir compris les enjeux spécifiques de l'IA, en particulier les limites de performances du SIA pour pouvoir apprécier la valeur du SIA et notamment le fait de suivre les propositions faites par le SIA de façon éclairée. L'enjeu est que le fournisseur du SIA se donne les moyens de s'assurer que les utilisateurs, qu'ils soient patients ou professionnels de santé, se sont bien appropriés les enjeux de l'IA, ont conscience de ses limites, et sont en capacité d'adhérer de façon éclairée aux propositions du SIA. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il s'est donné les moyens de vérifier la bonne compréhension des enjeux de l'IA par les utilisateurs : • les résultats de questionnaires d'évaluations réalisés à l'issue des sessions de formation, sensibilisation ou ateliers réalisés auprès des utilisateurs, qui auraient pour objectif d'évaluer la bonne compréhension des enjeux de l'IA, en particulier ses limites de performance • la réalisation régulière d'enquêtes utilisateurs ou sondages pour interroger les utilisateurs sur leur compréhension des enjeux de l'IA, sur leur confiance envers les propositions du SIA à la lumière de ces enjeux, ainsi que les implications sur leurs pratiques et leur adhésion aux propositions du SIA. • les résultats de ces sondages montrant le niveau de compréhension des utilisateurs.
30	Former à l'utilisation du SIA en amont de son introduction	N.A.	Au vu des enjeux du déploiement d'un SIA, les utilisateurs doivent être correctement formés par le fournisseur à l'utilisation du SIA en amont de son introduction. Il est de la responsabilité du fournisseur de former les futurs utilisateurs en amont du déploiement, quels que soient leurs profils, en proposant des formations adaptées à chaque profil d'utilisateur. Ces formations pourront être réalisées par le fournisseur lui-même, sous traitées, ou proposées en autonomie sur la base de supports de formation fournis par le fournisseur, destinés à tous les personnels concernés par le déploiement du SIA (responsable du déploiement dans la structure, utilisateurs). Le fournisseur du SIA pourrait fournir, à titre d'exemple, les éléments suivants pour montrer qu'il met en place les démarches et outils nécessaires à la formation des futurs utilisateurs : • Si le fournisseur forme directement les utilisateurs, lui-même ou avec le support d'un sous-traitant, les supports de formation ainsi qu'un rapport sur les sessions de formation réalisées (pouvant inclure les profils et nombre de participants, photos / captures d'écran des sessions, feedbacks des utilisateurs à l'issue de la formation) • Si la formation s'effectue en autonomie, les programmes et supports de formations, les attestations de participation des utilisateurs et les évaluations à l'issue de la formation, ainsi que tout élément attestant de la mise en place de temps d'échange avec le fournisseur pour accompagner les utilisateurs dans leur formation (comme par exemple des sessions de questions/réponses).

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
31	Permettre un accès facile et intuitif à la documentation du SIA	RIA Art. 13 : Transparence et information des entreprises de déploiement Art. 13.2 - Les systèmes d'IA à haut risque sont accompagnés d'un mode d'emploi dans un format numérique approprié ou autre, comprenant des informations concises, complètes, correctes et claires qui sont pertinentes, accessibles et compréhensibles pour les personnes chargées du déploiement.	En conformité avec les exigences du RIA, la documentation du SIA doit être facilement accessible. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il met en œuvre des mécanismes lui permettant de garantir l'accessibilité de la documentation relative au SIA : • les modalités permettant d'accéder à la documentation du SIA, que ce soit en version papier ou numérique, pour tout type d'utilisateur • des copies d'écran ou vidéos montrant qu'il est facile et intuitif d'accéder à une version numérique de la documentation du SIA depuis l'interface du SIA • les comptes rendus des sessions du CSE traitant de ce sujet.
32	Transparence sur les types d'IA employés	RIA Art. 11 – Documentation technique Art. 4 – Maîtrise de l'IA	En conformité avec le RIA, le ou les types d'IA utilisés devraient être précisés dans la documentation du SIA. Cette information est également explicitée lors des formations. Le fournisseur du SIA pourrait, à titre d'exemple, fournir : • la documentation du SIA, qui contient la description détaillée du type d'IA développée • les comptes rendus des sessions de formation où figure la transmission de l'information sur le type de l'IA développée • les questionnaires d'évaluation de l'impact des formations aux enjeux de l'IA qui questionnent sur le type de l'IA développée • le pourcentage des utilisateurs ayant répondu correctement à la question sur le type d'IA utilisée.
33	Transparence sur la performance du SIA	RIA - Art.13.3 et 15.3 Art. 13 - Transparence et fourniture d'informations aux déployeurs Art. 13.3 - Le mode d'emploi contient au moins les informations suivantes : (b) les caractéristiques, les capacités et les limites de performance du système d'IA à haut risque, y compris : (v) le cas échéant, ses performances à l'égard de personnes ou de groupes de personnes spécifiques sur lesquels le système est destiné à être utilisé ; Art. 15 – Notion d'exactitude et de robustesse Art. 15.3 - Les niveaux de précision et les paramètres de précision pertinents des systèmes d'IA à haut risque doivent être indiqués dans les instructions d'utilisation qui les accompagnent.	En conformité avec le RIA, le niveau de performance de l'IA pour le public cible doit être connu du déployeur et des utilisateurs. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il satisfait à cette exigence de transparence sur la performance de l'IA : • les éléments de la documentation du SIA mentionnant explicitement les performances du SIA à l'égard des personnes ou des groupes de personnes spécifiques sur lesquels le système est destiné à être utilisé • les questionnaires d'évaluation de l'impact des formations aux enjeux de l'IA qui questionnent sur la performance de l'IA et les caractéristiques de la population cible • le pourcentage des utilisateurs ayant répondu correctement à la question sur la performance de l'IA et les caractéristiques de la population cible. • les Datasheets et Model Cards détaillant les objectifs, les jeux de données, les biais identifiés, les limites, les traitements, la population cible et les performances attendues.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
34	Détecter la dépendance au SIA	RIA Art. 14 : Surveillance humaine Art. 14.4. Aux fins de la mise en œuvre des paragraphes 1, 2 et 3, le système d'IA à haut risque est fourni au responsable du déploiement de telle sorte que les personnes physiques auxquelles la surveillance humaine est confiée soient habilitées, selon ce qui est approprié et proportionné : (b) de rester conscients de la tendance possible à se fier automatiquement ou de manière excessive aux résultats produits par un système d'IA à haut risque (biais d'automatisation), en particulier pour les systèmes d'IA à haut risque utilisés pour fournir des informations ou des recommandations en vue de décisions à prendre par des personnes physiques	L'IA comporte des risques de dépendance pour son utilisateur, qui peuvent avoir des conséquences particulièrement néfastes dans le domaine de la santé si, par exemple, un professionnel de santé perd la capacité de prendre une décision de manière autonome, et suit systématiquement les propositions du SIA. Ceci correspond au biais d'automatisation. L'enjeu est que le SIA intègre une fonctionnalité permettant de détecter, sur la base d'une analyse fine du comportement d'un utilisateur, s'il y a des éléments mettant en lumière un risque de dépendance à l'IA. Ainsi, le fournisseur du SIA pourrait montrer qu'il a mis en place les mécanismes permettant la détection de toute situation où l'utilisateur se trouve à risque de dépendance à l'IA, en fournissant, à titre d'exemple, les éléments suivants : • les outils de suivi du comportement des utilisateurs vis à vis du SIA, tels que des systèmes permettant d'analyser la fréquence et la nature des interactions des utilisateurs avec le SIA. • des rapports réguliers sur les modalités d'utilisation du SIA de chaque utilisateur, contenant des éléments quantitatifs relatifs à son profil d'utilisation : recours à l'IA, % de suivi et de non suivi de la proposition de l'IA avec justification du non suivi • le critère de définition d'une situation de dépendance d'un utilisateur, par exemple, lorsque le taux de suivi des propositions du SIA est supérieur à la performance du SIA.
35	Alerter en cas de détection d'une dépendance au SIA	N.A.	En cas de détection d'un risque de dépendance à l'IA, l'utilisateur doit en être informé, ainsi que l'administrateur, pour pouvoir mettre en place les mesures correctrices appropriées. L'enjeu est que le SIA puisse générer automatiquement des alertes en cas de détection d'un risque de dépendance à l'IA, à destination de l'utilisateur et de l'administrateur. Le fournisseur du SIA pourrait fournir, à titre d'exemple, les éléments suivants pour montrer qu'il a mis en place les mécanismes adéquats de génération et de gestion des alertes en cas de dépendance à l'IA : • la description technique des systèmes d'alerte intégrés dans le SIA, qui notifient les utilisateurs et les administrateurs lorsqu'une dépendance potentielle est détectée, par exemple lorsqu'un seuil prédéfini de recours à l'IA avec suivi de la recommandation est dépassé, sur la base des éléments quantitatifs collectés sur le profil d'utilisateur (cf critère 36). Ce seuil doit être inférieur au niveau de performance affiché de l'IA. • des captures d'écran ou documentation montrant le mode de fonctionnement de ces alertes en action • les protocoles de gestion des alertes, détaillant les étapes à suivre lorsqu'une alerte de dépendance est déclenchée, incluant des interventions spécifiques • la documentation des interventions réalisées suite aux alertes • la description des services de support technique et d'assistance proposés pour aider les utilisateurs à gérer les alertes de dépendance et à maintenir leur autonomie décisionnelle.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
36	Garder un esprit critique et faire confiance à son expertise clinique	MDR S'il s'agit d'un risque identifié par le fabricant alors : Annexe I point 4.c Annexe I point 22.2 profanes S'il n'y a pas de risque identifié par le fabricant alors il n'y aura pas de formation.	Comme pour tout dispositif médical, le fournisseur d'un SIA doit mettre en place des mesures de maîtrise des risques incluant l'information et la formation des utilisateurs afin de leur permettre de conserver un esprit critique lorsqu'ils utilisent le SIA. Il convient de rappeler à l'utilisateur que la performance d'un SIA n'est jamais de 100%, et qu'il doit faire confiance à son expertise clinique pour prendre une décision. Il est de la responsabilité du fournisseur du SIA d'insister sur cet aspect dans les formations délivrées aux utilisateurs. Le fournisseur pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il délivre une formation appropriée : • les supports de formation aux utilisateurs, incluant une sensibilisation à l'esprit critique dans le cadre de l'utilisation de l'IA • les retours des utilisateurs sur les formations dispensées.
37	Rappeler la responsabilité de l'utilisateur sur sa décision	N.A.	Le SIA est conçu pour aider les utilisateurs à décider, mais c'est toujours l'utilisateur qui décide et assume la responsabilité de sa décision. Il est crucial que l'utilisateur ne perde pas de vue qu'il est responsable de ses décisions, qu'il ait recours ou pas au SIA, qu'il suive ou pas ses propositions. Le SIA doit donc inclure des fonctionnalités qui rappellent à l'utilisateur, qu'il soit patient ou professionnel de santé, que les suggestions de l'IA sont simplement des propositions et que la décision finale lui appartient. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il a mis en place les mécanismes nécessaires pour que l'utilisateur soit régulièrement rappelé à sa responsabilité quant à la prise de décision, et qu'il ait à sa disposition les éléments nécessaires pour comprendre ce que cela signifie dans le cadre de l'utilisation du SIA : • - les messages de rappel intégrés dans l'interface utilisateur du SIA, indiquant clairement que l'utilisateur conserve la responsabilité de ses décisions, avec des captures d'écran ou de la documentation montrant ces messages de rappel en action • - les preuves de diffusion aux utilisateurs d'informations adaptées sur leur responsabilité et les limites des propositions de l'IA, telles que les programmes des formations et de sensibilisation concernant la responsabilité des utilisateurs dans l'utilisation des propositions de l'IA, les rapports de participation à ces programmes et les feedbacks des utilisateurs sur ces sessions, les guides et manuels expliquant la responsabilité des utilisateurs et les limites des propositions de l'IA.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
38	Sensibilisation à un usage écoresponsable	N.A.	Le recours à l'IA est consommateur de ressources, notamment d'énergie et d'eau. Aussi est-il important de sensibiliser les utilisateurs à l'impact environnemental et aux bonnes pratiques à adopter pour réduire leur empreinte écologique, afin d'encourager une sobriété numérique partagée. L'enjeu est que le fournisseur du SIA s'engage activement dans la sensibilisation de ses utilisateurs à un usage écoresponsable du SIA, c'est-à-dire, à un usage pas forcément systématique mais qui réponde à un besoin. A cette fin, le fournisseur du SIA pourrait produire, à titre d'exemple, les éléments suivants pour montrer son engagement en ce sens : • détail du contenu des sessions de formation spécifique sur les bonnes pratiques d'utilisation écoresponsable de l'IA • les copies d'écran et vidéos permettant d'établir l'accessibilité de la documentation sur le sujet • les rapports de participation aux sessions de formation et feedbacks des utilisateurs sur ces sessions • des preuves de distribution et d'utilisation des supports pédagogiques par les utilisateurs • les supports pédagogiques (vidéos, infographies, brochures) expliquant les impacts environnementaux de l'utilisation de l'IA et les moyens de les minimiser, mis à disposition de l'utilisateur du SIA par le fournisseur.
39	Souveraineté des données de l'organisation utilisatrice	HDS V2	Les données de santé étant des données sensibles, l'organisation utilisatrice d'un SIA doit être assurée de pouvoir conserver la souveraineté sur ses données. L'enjeu est de garantir la possibilité d'exécuter le SIA de telle sorte que les données partagées par l'utilisateur, en particulier les données de santé, ne soient pas soumises à des réglementations extra-européennes. Le fournisseur du SIA devrait dans ce cadre réaliser une cartographie des dépendances technologiques critiques, et proposer des solutions permettant à l'utilisateur de conserver la pleine souveraineté sur ses données. A cette fin, le fournisseur du SIA pourrait, à titre d'exemple, produire les éléments suivants, pour montrer qu'il propose des solutions techniques permettant à l'utilisateur de conserver la pleine souveraineté sur ses données : • - soit des solutions de cloud souverain, en précisant les spécifications techniques attestant de l'exécution du SIA sur un cloud souverain qui garantisse que les données restent sous juridiction contrôlée • - soit des solutions d'utilisation en local, permettant, dans le cas des LLM, aux structures utilisatrices d'être autonomes pour installer et exécuter le module d'IA sur leurs infrastructures afin d'interdire les flux de données extraterritoriaux.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
40	Évaluer le SIA en continu	RIA Art. 9 - Système de gestion des risques Art. 14 – Surveillance Humaine	Conformément au RIA, le fournisseur d'un SIA met en place des mécanismes permettant d'évaluer régulièrement les résultats de l'IA. Le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il a mis en place les moyens adéquats pour évaluer régulièrement les résultats du SIA : - la description des systèmes de gestion des risques et de surveillance humaine mis en place - les rapports d'évaluation périodique.
41	Mettre en place une démarche d'amélioration continue du SIA	N.A.	Les retours sur l'utilisation d'un SIA en conditions réelles permettent d'identifier des pistes d'amélioration ou de possibles dérives à corriger, qu'il convient de prendre en compte dans le cadre d'une démarche organisée d'amélioration continue. L'enjeu est que le fournisseur s'engage à mettre en place une démarche d'amélioration continue du SIA, incluant l'évaluation régulière des propositions du SIA, un dispositif accessible et unifié de signalement des problèmes, et la mise en place de plan d'actions. Ce dispositif devra être accompagné d'un protocole de remédiation rapide, transparent, et intelligible, à destination des professionnels comme des patients. A cette fin, le fournisseur du SIA pourrait, à titre d'exemple, fournir les éléments suivants pour montrer qu'il s'inscrit dans une démarche d'amélioration continue : - les rapports d'évaluation périodique des performances du SIA, et les résultats des évaluations montrant les améliorations apportées et les ajustements nécessaires - les plans d'actions spécifiques mis en place suite aux évaluations du SIA, incluant des mesures correctives et des améliorations continues - les méthodologies et outils utilisés pour détecter les dérives dans les systèmes d'IA, comme les biais et les erreurs, avec les rapports sur les dérives détectées et les actions correctrices mises en place - les protocoles de contrôle humain intégrés dans le fonctionnement du SIA.

	Libellé court	Règlement, référentiel ou norme	Propositions d'éléments d'autoévaluation
42	Transparence sur la mise à jour d'un SIA et ses conséquences	N.A.	Une mise à jour du SIA peut entraîner des modifications importantes pour l'utilisateur, que ce soit dans sa manière d'interagir avec le SIA ou dans les propositions générées par l'IA, ce qui peut avoir des impacts sur la façon d'intégrer le recours au SIA dans la pratique de soin. Il est donc important que les mises à jour et leurs impacts sur l'utilisation du SIA soient clairement explicités à l'utilisateur. L'enjeu est de s'assurer que toute mise à jour est clairement notifiée à l'utilisateur, avec des explications lisibles et transparentes quant à ses impacts potentiels sur son utilisation et le comportement de l'utilisateur. A cette fin, le fournisseur du SIA pourrait, à titre d'exemple, produire les éléments suivants pour montrer qu'il met en œuvre les moyens adéquats pour informer l'utilisateur d'une mise à jour du SIA et de ses conséquences sur son utilisation : - la documentation sur les notifications envoyées aux utilisateurs concernant les mises à jour du SIA, avec des exemples de messages de notification expliquant les modifications et leurs impacts potentiels - l'organisation de sessions d'information pour expliquer les mises à jour ayant des impacts importants et leurs conséquences sur l'utilisation du SIA - des preuves de distribution et d'accès facile aux versions mises à jour de la documentation pour l'utilisateur, tels que guides ou manuels d'utilisation
43	Vérifier la bonne compréhension des conséquences d'une mise à jour d'un SIA	N.A.	Une mise à jour du SIA pouvant entraîner des conséquences importantes pour l'utilisateur, il est important de vérifier que les informations qui lui ont été communiquées quant à la mise à jour et à ses impacts ont été bien reçues et comprises, et ce pour tout type d'utilisateur, qu'il soit patient ou professionnel de santé. L'enjeu est de s'assurer qu'en cas de mise à jour du SIA, l'utilisateur a bien compris les impacts sur son utilisation et les décisions qu'il peut être amené à prendre en l'utilisant. A cette fin, le fournisseur du SIA pourrait, à titre d'exemple, produire les éléments suivants pour montrer qu'il a bien vérifié la bonne compréhension des messages adressés à l'utilisateur quant aux conséquences d'une mise à jour : - des questionnaires et évaluations pour vérifier la compréhension des utilisateurs sur les impacts des mises à jour, avec les résultats des évaluations montrant le niveau de compréhension des utilisateurs suite à la mise à jour - l'inclusion dans les sessions d'information spécifiques post mise à jour d'un questionnaire ou d'exercices visant à vérifier leur bonne compréhension des conséquences de la mise à jour du SIA, les rapports de participation et feedbacks des utilisateurs sur ces sessions - la mise en place de services de support technique et éducatif pour répondre aux questions des utilisateurs concernant les mises à jour, avec les rapports sur les interactions avec le support technique et les questions fréquemment posées suite à une mise à jour - des mécanismes de suivi pour vérifier la compréhension continue des utilisateurs après les mises à jour, comme des check-ins réguliers.

