

Construction de notre grille d'analyse

OUTIL ALTAI non modifié

Critères d'évaluation	Sous critères	Outil ALTAI Questions et sous-questions
Action humaine et contrôle humain	Droits fondamentaux	Dans les cas d'utilisation susceptibles d'entraîner des effets négatifs sur les droits fondamentaux, avez-vous réalisé une analyse d'impact sur les droits fondamentaux ? Avez-vous déterminé et documenté le recours potentiel à des arbitrages entre les différents principes et droits ?
		<p>Le système d'IA interagit-il avec la prise de décision par un utilisateur final humain (par exemple, en recommandant des mesures ou décisions à prendre, ou en présentant des choix possibles) ?</p> <ul style="list-style-type: none"> Dans de tels cas, existe-t-il un risque que le système d'IA affecte l'autonomie humaine en interférant de manière involontaire avec le processus décisionnel de l'utilisateur final ? Estimez-vous qu'un système d'IA devrait communiquer aux utilisateurs qu'une décision, un contenu, un conseil ou un résultat découlent d'une décision algorithmique ? Lorsque le système d'IA comporte un robot ou système conversationnel, les utilisateurs humains sont-ils informés du fait qu'ils interagissent avec un agent virtuel ?
	Action humaine	Lorsque le système d'IA est intégré dans un processus de travail, avez-vous réfléchi à la répartition des tâches entre le système d'IA et les travailleurs humains pour permettre des interactions constructives ainsi qu'une supervision et un contrôle humains appropriés ?
		Le système d'IA renforce-t-il ou augmente-t-il les capacités humaines ?
		Avez-vous prévu des garanties pour empêcher toute confiance ou dépendance excessives envers le système d'IA dans les processus de travail ?
	Contrôle humain	Avez-vous réfléchi au niveau approprié de contrôle humain pour le système d'IA et le cas d'utilisation en question ?
		Pouvez-vous décrire le niveau de contrôle ou de participation humains, le cas échéant ? Qui est « l'humain aux manettes » et à quel moment y a-t-il intervention humaine, ou avec quels outils ?
		Avez-vous mis en place des mécanismes et des mesures pour garantir un contrôle ou une supervision humains potentiels de cette nature, ou pour veiller à ce que les décisions soient prises sous la responsabilité globale d'êtres humains ?

		Avez-vous pris des mesures pour permettre la réalisation d'audits et résoudre des questions liées à la gouvernance de l'autonomie de l'IA ?
	Dans le cas d'un système d'IA ou d'une utilisation capables d'auto-apprentissage ou autonomes, avez-vous mis en place des mécanismes plus spécifiques de contrôle et de supervision ?	Quel type de mécanismes de détection et de réponse avez-vous mis sur pied pour évaluer le risque que des problèmes surviennent ?
		Avez-vous veillé à la présence d'un « bouton d'arrêt » ou à l'existence d'une procédure pour suspendre, en cas de besoin, une opération en toute sécurité? Cette procédure suspend-elle le processus dans sa totalité, en partie, ou délègue-t-elle le contrôle à un être humain ?
Robustesse technique et sécurité	Résilience aux attaques et sécurité	<p>Avez-vous évalué des formes d'attaques potentielles auxquelles le système d'IA pourrait être vulnérable ?</p> <ul style="list-style-type: none"> • Avez-vous en particulier envisagé différents types et différentes natures de vulnérabilités, comme la pollution des données, l'infrastructure physique, les cyber attaques ? <p>Avez-vous prévu des mesures ou systèmes pour veiller à l'intégrité et à la résilience du système d'IA face à de potentielles attaques ?</p> <p>Avez-vous évalué le comportement de votre système dans des situations ou des environnements imprévus ?</p> <p>Avez-vous envisagé si, et dans quelle mesure, votre système pourrait avoir un double usage ? Le cas échéant, avez-vous pris des mesures préventives appropriées contre un tel cas de figure (y compris, par exemple, ne pas publier la recherche ou ne pas déployer le système)?</p>
	Solutions de secours et sécurité générale	<p>Avez-vous veillé à ce que votre système dispose de suffisamment de solutions de secours pour faire face à d'éventuelles attaques antagonistes ou autres situations imprévues (par exemple, procédures de relais technique ou demande de communication avec un opérateur humain avant d'agir) ?</p> <p>Avez-vous envisagé le niveau de risque posé par le système d'IA dans ce cas d'utilisation spécifique ?</p> <ul style="list-style-type: none"> • Avez-vous mis en place un processus pour mesurer et évaluer les risques et la sécurité ? • Avez-vous fourni les informations nécessaires en cas de risque pour l'intégrité physique humaine ? • Avez-vous réfléchi à une politique d'assurance pour couvrir les dégâts potentiels provoqués par le système d'IA ? • Avez-vous recensé les risques potentiels en matière de sécurité d'autres utilisations prévisibles de la technologie, y compris d'utilisation abusive accidentelle ou malveillante ? Existe-t-il un plan pour atténuer ou gérer ces risques ? <p>Avez-vous évalué s'il est probable que le système d'IA cause des dommages ou préjudices aux utilisateurs ou à des tiers ? Le cas échéant, avez-vous évalué la probabilité, les dommages potentiels, le public concerné et la gravité ?</p> <ul style="list-style-type: none"> • En cas de risque qu'un système d'IA cause des dommages, avez-vous réfléchi à des règles de responsabilité et de protection des consommateurs, et de quelle manière en avez-vous tenu compte ?

		<ul style="list-style-type: none"> • Avez-vous réfléchi à l'incidence potentielle ou au risque en matière de sécurité sur l'environnement ou les animaux ? • Vous êtes-vous demandé, dans le cadre de votre analyse des risques, si des problèmes de sécurité ou de réseau (par exemple, des menaces pesant sur la cybersécurité) pourraient mettre en péril la sécurité ou entraîner des préjudices du fait d'un comportement involontaire du système d'IA ? <p>Avez-vous évalué l'incidence probable d'une défaillance de votre système d'IA entraînant la production de résultats erronés, l'indisponibilité de votre système, ou la production de résultats inacceptables pour la société (par exemple, pratiques discriminatoires) ?</p> <ul style="list-style-type: none"> • Avez-vous mis en place des seuils et une gouvernance pour les scénarios ci-dessus afin de déclencher d'autres plans/solutions de secours ? • Avez-vous défini et testé des solutions de secours ?
	Précision	<p>Avez-vous évalué le niveau de précision et la définition de la précision nécessaires dans le contexte du système d'IA et du cas d'utilisation concerné ?</p> <ul style="list-style-type: none"> • Avez-vous réfléchi à la manière dont la précision est mesurée et assurée ? • Avez-vous mis en place des mesures pour veiller à ce que les données utilisées soient exhaustives et à jour ? • Avez-vous mis en place des mesures pour évaluer si des données supplémentaires sont nécessaires, par exemple pour améliorer la précision et éliminer les biais ? <p>Avez-vous évalué le préjudice que causeraient des prédictions inexactes du système d'IA ?</p> <p>Avez-vous prévu des moyens de mesurer si votre système produit un nombre inacceptable de prédictions inexactes ?</p> <p>En cas de prédictions inexactes, avez-vous mis en place une série d'étapes pour résoudre le problème ?</p>
	Fiabilité et reproductibilité	<p>Avez-vous mis en place une stratégie afin de contrôler le système d'IA et de vous assurer qu'il répond aux objectifs, aux finalités et aux applications prévues ?</p> <ul style="list-style-type: none"> • Avez-vous vérifié si des contextes spécifiques ou conditions particulières doivent être pris en compte pour garantir la reproductibilité ? • Avez-vous mis en place des processus ou méthodes de vérification pour mesurer et garantir les différents aspects de la fiabilité et de la reproductibilité ? • Avez-vous mis en place des processus visant à décrire certains réglages susceptibles d'entraîner une défaillance du système d'IA ?

		<ul style="list-style-type: none"> • Avez-vous clairement documenté et appliqué ces processus aux fins des essais et de la vérification de la fiabilité du système d'IA ? • Avez-vous mis en place un mécanisme ou une communication pour garantir aux utilisateurs (finaux) la fiabilité du système d'IA ?
Respect de la vie privée et gouvernance des données	Respect de la vie privée et protection des données	<p>En fonction du cas d'utilisation, avez-vous mis sur pied un mécanisme permettant à autrui de signaler des problèmes en rapport avec le respect de la vie privée et la protection des données durant les processus suivis par le système d'IA pour la collecte des données (aux fins de l'entraînement et du fonctionnement) et le traitement des données ?</p> <p>Avez-vous évalué le type et la portée des données constituant vos ensembles de données (par exemple, si elles contiennent des données à caractère personnel) ?</p> <p>Avez-vous réfléchi à des manières de mettre au point le système d'IA ou d'entraîner le modèle sans utiliser (ou en utilisant de manière limitée) des données potentiellement sensibles ou à caractère personnel ?</p> <p>Avez-vous intégré des mécanismes de notification et de contrôle concernant les données à caractère personnel en fonction du cas d'utilisation (comme un consentement valable et la possibilité de révoquer le consentement, le cas échéant) ?</p> <p>Avez-vous pris des mesures pour renforcer le respect de la vie privée, par exemple des mesures de cryptage, d'anonymisation et d'agrégation ?</p> <p>Lorsqu'il existe un responsable de la protection des données, avez-vous mobilisé cette personne à un stade précoce dans le processus ?</p>
	Qualité et intégrité des données	<p>Avez-vous aligné votre système sur d'éventuelles normes pertinentes (par exemple, ISO, IEEE) ou des protocoles largement adoptés dans le cadre de votre gestion et de votre gouvernance quotidiennes des données ?</p> <p>Avez-vous mis sur pied des mécanismes de contrôle pour la collecte, le stockage, le traitement et l'utilisation des données ?</p> <p>Avez-vous évalué la mesure dans laquelle vous contrôlez la qualité des sources externes des données utilisées ?</p> <p>Avez-vous mis en place des processus pour garantir la qualité et l'intégrité de vos données ? Avez-vous envisagé d'autres processus ? De quelle manière vérifiez-vous que vos ensembles de données n'ont pas été compromis ou piratés ?</p>
	Accès aux données	<p>Quels protocoles, processus et procédures ont été suivis pour gérer et garantir la gouvernance appropriée des données ?</p> <ul style="list-style-type: none"> • Avez-vous analysé qui peut accéder aux données des utilisateurs et dans quelles circonstances ? • Avez-vous veillé à ce que ces personnes soient qualifiées, qu'elles aient effectivement besoin d'accéder aux données et à ce qu'elles disposent des compétences nécessaires pour comprendre précisément la politique de protection des données ? • Avez-vous prévu un mécanisme de contrôle pour consigner quand, où, comment, par qui et dans quel but les données ont été consultées ?

Transparence	Traçabilité	<p>Avez-vous mis des mesures en place susceptibles de garantir la traçabilité ? Cela pourrait consister à documenter :</p> <ul style="list-style-type: none"> • Les méthodes appliquées aux fins de la conception et de la mise au point du système algorithmique : • Dans le cas d'un système d'IA fondé sur des règles, la méthode de programmation ou la manière dont le modèle a été mis au point devraient être documentées; • Dans le cas d'un système d'IA fondé sur l'apprentissage, la méthode d'entraînement de l'algorithme, y compris quelles données d'entrée ont été collectées et sélectionnées, et dans quelles conditions, devrait être documentée. • Les méthodes appliquées pour tester et valider le système algorithmique : • Dans le cas d'un système d'IA fondé sur des règles, les scénarios ou cas utilisés pour tester et valider devraient être documentés ; • Dans le cas d'un système d'IA fondé sur l'apprentissage, les informations relatives aux données utilisées pour tester et valider devraient être documentées. • Les résultats du système algorithmique : • Les résultats d'un algorithme ou les décisions qu'il prend, ainsi que les éventuelles autres décisions qui résulteraient de différents cas (par exemple, pour d'autres sous-groupes d'utilisateurs) devraient être documentés.
	Explicabilité	<p>Avez-vous évalué la mesure dans laquelle les décisions prises, et donc les résultats obtenus, par le système d'IA peuvent être compris</p> <p>Avez-vous veillé à ce qu'une explication de la raison pour laquelle un système a procédé à un certain choix entraînant un certain résultat puisse être rendue compréhensible pour l'ensemble des utilisateurs qui pourraient souhaiter obtenir une explication ?</p> <p>Avez-vous évalué la mesure dans laquelle la décision du système influence les processus décisionnels de l'organisation ?</p> <p>Avez-vous évalué pourquoi ce système particulier a été déployé dans ce domaine spécifique ?</p> <p>Avez-vous évalué le modèle économique concernant ce système (par exemple, en quoi crée-t-il de la valeur pour l'organisation) ?</p> <p>Avez-vous conçu le système d'IA en ayant dès le départ l'interprétation à l'esprit ?</p> <ul style="list-style-type: none"> • Avez-vous cherché à utiliser le modèle le plus simple et le plus facile à interpréter pour l'application en question ? • Avez-vous évalué si vous êtes en mesure d'analyser les données que vous avez utilisées aux fins de l'entraînement et des essais ? Cela peut-il être modifié et actualisé au fil du temps ?

		Avez-vous évalué si des solutions s’offrent à vous suite à l’entraînement et à la mise au point du modèle pour examiner l’interprétation ou si vous avez accès à la séquence des opérations du modèle ?
	Communication	<p>Avez-vous informé les utilisateurs (finaux) – au moyen d’une clause de non- responsabilité ou de tout autre moyen – qu’ils interagissent avec un système d’IA et pas avec un autre être humain ? Avez-vous indiqué clairement que votre système est doté de l’IA ?</p> <p>Avez-vous mis en place des mécanismes pour informer les utilisateurs des raisons et critères expliquant les résultats du système d’IA ?</p> <ul style="list-style-type: none"> • Les utilisateurs visés en sont-ils informés de manière claire et intelligible ? • Avez-vous établi des processus pour tenir compte des commentaires des utilisateurs et utiliser ces commentaires pour adapter le système ? • Avez-vous également communiqué les risques potentiels ou perçus, tels que les biais ? • En fonction du cas d’utilisation, avez-vous également réfléchi à la communication et à la transparence envers d’autres publics, des tiers ou le grand public ? <p>Avez-vous clairement indiqué la finalité du système d’IA et qui ou ce qui pourrait bénéficier du produit/service ?</p> <ul style="list-style-type: none"> • Les scénarios d’utilisation du produit ont-ils été définis et clairement expliqués, en envisageant également d’autres moyens de communication pour veiller à ce qu’ils soient compréhensibles et appropriés pour le destinataire ? • En fonction du cas d’utilisation, avez-vous réfléchi à la psychologie humaine et aux potentielles limites humaines, comme le risque de confusion, les biais de confirmation ou la fatigue cognitive ? <p>Avez-vous clairement expliqué les caractéristiques, les limites et les éventuelles lacunes du système d’IA :</p> <ul style="list-style-type: none"> • S’agissant de la mise au point : à toute personne chargée de son déploiement pour en faire un produit ou service? • S’agissant du déploiement : à l’utilisateur final ou au consommateur?
Diversité, non-discrimination et équité	Eviter les biais injustes	<p>Avez-vous prévu une stratégie ou un ensemble de procédures pour éviter de créer ou de renforcer des biais injustes dans le système d’IA, en ce qui concerne tant l’utilisation des données d’entrée que la conception de l’algorithme ?</p> <ul style="list-style-type: none"> • Avez-vous évalué et reconnu les éventuelles limites provenant de la composition des ensembles de données utilisés ? • Avez-vous réfléchi à la diversité et à la représentativité des utilisateurs dans les données ? Avez-vous procédé à des essais portant sur des populations spécifiques ou des cas d’utilisation problématiques ?

		<ul style="list-style-type: none"> • Avez-vous recherché et utilisé les outils techniques disponibles pour améliorer votre compréhension des données, du modèle et de la performance ? • Avez-vous mis en place des processus pour tester et contrôler les biais éventuels au cours de la phase de mise au point, de déploiement et d'utilisation du système ? <p>En fonction du cas d'utilisation, avez-vous prévu un mécanisme permettant à autrui de signaler des problèmes liés aux biais, à la discrimination ou aux mauvaises performances du système d'IA ?</p> <ul style="list-style-type: none"> • Avez-vous envisagé des mesures et des moyens de communication clairs pour savoir comment et à qui ces problèmes peuvent être signalés ? • Avez-vous tenu compte non seulement des utilisateurs (finaux) mais également des autres personnes susceptibles d'être indirectement affectées par le système d'IA ? <p>Avez-vous évalué si, dans des conditions identiques, une éventuelle variabilité des décisions est possible ?</p> <ul style="list-style-type: none"> • Le cas échéant, avez-vous réfléchi aux causes probables ? • Concernant la variabilité, avez-vous mis sur pied un mécanisme de mesure ou d'évaluation de l'incidence potentielle de cette variabilité sur les droits fondamentaux ? <p>Avez-vous prévu une définition appropriée de l'«équité» que vous appliquez dans la conception des systèmes d'IA ?</p> <ul style="list-style-type: none"> • Votre définition est-elle couramment utilisée ? Avez-vous envisagé d'autres définitions avant de choisir celle-ci ? • Avez-vous prévu une analyse quantitative ou des indicateurs pour mesurer et tester la définition appliquée de l'équité ? • Avez-vous mis sur pied des mécanismes visant à garantir l'équité dans vos systèmes d'IA ? Avez-vous envisagé d'autres mécanismes potentiels ?
	Accessibilité et conception universelle	<p>Avez-vous veillé à ce que le système d'IA réponde aux besoins d'un large ensemble de préférences et de capacités individuelles ?</p> <ul style="list-style-type: none"> • Avez-vous évalué si le système d'IA peut être utilisé par les personnes présentant des besoins spécifiques ou un handicap ou qui sont exposées au risque d'exclusion ? Comment cet aspect a-t-il été intégré à la conception du système et comment est-il vérifié ? • Avez-vous veillé à ce que les informations relatives au système d'IA soient également accessibles aux utilisateurs de technologies d'assistance ? • Avez-vous mobilisé ou consulté cette communauté d'utilisateurs au cours de la phase de mise au point du système d'IA ?

		<p>Avez-vous tenu compte de l'incidence de votre système d'IA sur le groupe d'utilisateurs potentiels ?</p> <ul style="list-style-type: none"> • L'équipe participant à la mise au point du système d'IA est-elle représentative de votre groupe cible d'utilisateurs ? Est-elle représentative de la population au sens large, compte tenu également d'autres groupes susceptibles d'être indirectement concernés ? • Avez-vous évalué si certaines personnes ou certains groupes pourraient subir de manière disproportionnée des effets négatifs ? • D'autres équipes ou groupes présentant différents parcours professionnels et expériences vous ont-ils fait parvenir des • réactions ?
	Participation des parties prenantes	<p>Avez-vous réfléchi à un mécanisme pour inclure la participation de différentes parties prenantes dans la mise au point et l'utilisation du système d'IA ?</p> <p>Avez-vous préparé la voie à l'introduction du système d'IA au sein de votre organisation en informant et en mobilisant au préalable les travailleurs concernés et leurs représentants ?</p>
Bien-être sociétal et environnemental	IA durable et respectueuse de l'environnement	<p>Avez-vous mis en place des mécanismes pour mesurer l'impact environnemental de la mise au point, du déploiement et de l'utilisation du système d'IA (par exemple, énergie consommée par les centres de données, type d'énergie consommée par les centres de données, etc.) ?</p> <p>Avez-vous prévu des mesures pour réduire l'impact environnemental du cycle de vie de votre système d'IA ?</p>
	Incidence sociale	<p>Lorsque le système d'IA interagit directement avec des êtres humains:</p> <ul style="list-style-type: none"> • Avez-vous évalué si le système d'IA encourage les êtres humains à développer de l'attachement et de l'empathie pour le système ? • Avez-vous veillé à ce que le système d'IA indique clairement que son interaction sociale est simulée et qu'il n'a nullement la capacité de «comprendre» et de «ressentir»? <p>Avez-vous veillé à ce que les incidences sociales du système d'IA soient bien comprises ? Par exemple, vous êtes-vous demandé s'il existe un risque de perte d'emplois et de perte de compétences de la main-d'œuvre ? Quelles mesures ont été prises pour contrer ces risques ?</p>
	Société et démocratie	<p>Avez-vous évalué l'incidence plus large de l'utilisation du système d'IA sur la société, au-delà de l'utilisateur (final) individuel, par exemple les parties prenantes susceptibles d'être indirectement concernées ?</p>
Responsabilité	Audibilité	<p>Avez-vous mis en place des mécanismes pour faciliter l'auditabilité du système par des acteurs internes et/ou indépendants, en veillant pas exemple à la traçabilité et à la journalisation des processus et des résultats du système d'IA ?</p>
	Minimisation et documentation des incidences négatives	<p>Avez-vous réalisé une analyse des risques ou de l'impact du système d'IA qui tienne compte de différentes parties prenantes qui sont directement et indirectement concernées ?</p>

		<p>Avez-vous mis en place des cadres de formation et d'éducation pour définir des pratiques en matière de responsabilité ?</p> <ul style="list-style-type: none"> • Quels travailleurs ou branches de travailleurs sont concernés ? Le sont-ils au-delà de la phase de mise au point ? • Ces formations portent-elles également sur le cadre juridique potentiellement applicable au système d'IA ? • Avez-vous envisagé la mise sur pied d'un « comité d'examen pour l'IA éthique » ou d'un mécanisme similaire pour discuter des pratiques globales en matière de responsabilité et d'éthique, y compris des zones grises potentiellement floues ? <p>Outre les initiatives ou cadres internes destinés à contrôler l'éthique et la responsabilité, des orientations externes ou des processus d'audit ont-ils également été mis en place ?</p> <p>Existe-t-il des processus permettant aux tiers (par exemple, fournisseurs, consommateurs, distributeurs/vendeurs) ou aux travailleurs de signaler de possibles vulnérabilités, risques ou biais dans le système/l'application d'IA ?</p>
	Documentation des arbitrages	<p>Avez-vous mis sur pied un mécanisme permettant de recenser les intérêts et les valeurs pertinents concernés par le système d'IA et les éventuels arbitrages entre eux ?</p> <p>Quel processus utilisez-vous pour prendre des décisions relatives à ces arbitrages ? Avez-vous veillé à ce que les décisions d'arbitrage soient documentées ?</p>
	Voies de recours	<p>Avez-vous mis en place un ensemble approprié de mécanismes permettant un recours en cas de préjudice ou d'effet néfaste ?</p> <p>Avez-vous mis en place des mécanismes pour fournir des informations aux utilisateurs (finaux)/tiers à propos des possibilités de recours ?</p>

OUTIL ALTAI modifié

Chaque dimension a été déclinée en critères et sous-critères, traduits en exigences et besoins opérationnels. Cette transformation a permis de passer d'un cadre théorique normatif à une grille d'évaluation concrète, applicable à l'analyse.

OUTIL ALTAI				
CRITERES	Sous-critères	Questionnements	Transformation des questions en exigences	Besoins
Action humaine et contrôle humain	Droits fondamentaux	Dans les cas d'utilisation susceptibles d'entraîner des effets négatifs sur les droits fondamentaux, avez-vous réalisé une analyse d'impact sur les droits fondamentaux ?	Le système d'IA doit faire l'objet d'une analyse d'impact documentée sur les droits fondamentaux lorsqu'il est susceptible d'avoir des effets significatifs sur les utilisateurs.	Analyse d'impact sur les droits fondamentaux
		Avez-vous déterminé et documenté le recours potentiel à des arbitrages entre les différents principes et droits ?	Les arbitrages entre principes ou droits fondamentaux doivent être identifiés, justifiés et documentés de manière transparente.	Documentation des arbitrages éthiques
		Le système d'IA interagit-il avec la prise de décision par un utilisateur <u>final humain</u> (par exemple, en recommandant des mesures ou décisions à prendre, ou en présentant des choix possibles)?	Le système d'IA doit être conçu pour soutenir la prise de décision humaine, en rendant explicites les recommandations ou options proposées à l'utilisateur.	Interaction avec la prise de décision humaine
		<ul style="list-style-type: none"> Dans de tels cas, existe-t-il un risque que le système d'IA affecte l'autonomie humaine en interférant de manière involontaire avec le processus décisionnel de l'utilisateur final? Estimez-vous qu'un système d'IA devrait communiquer aux utilisateurs qu'une décision, un contenu, un conseil ou un résultat découlent d'une décision algorithmique? Lorsque le système d'IA comporte un robot ou système conversationnel, les utilisateurs humains sont-ils informés du fait qu'ils interagissent avec un agent virtuel? 	Le système d'IA ne doit pas interférer de manière involontaire avec l'autonomie de l'utilisateur ; des mécanismes doivent être prévus pour éviter la dépendance ou la sur confiance.	Préservation de l'autonomie décisionnelle
			Le système d'IA doit informer clairement les utilisateurs lorsqu'un contenu, une décision ou une recommandation est généré par un algorithme.	Signalement des décisions algorithmiques

			Lorsqu'un système conversationnel ou un agent virtuel est utilisé, l'utilisateur doit être informé de manière explicite qu'il interagit avec une entité non humaine.	Identification des agents virtuels
	Action humaine	Lorsque le système d'IA est intégré dans un processus de travail, avez-vous réfléchi à la répartition des tâches entre le système d'IA et les travailleurs humains pour permettre des interactions constructives ainsi qu'une supervision et un contrôle humains appropriés ?	Lorsqu'un système d'IA est intégré dans un processus de travail, la répartition des tâches entre l'IA et les opérateurs humains doit être définie de manière à garantir des interactions constructives, une supervision efficace et un contrôle humain approprié.	Répartition des tâches entre IA et humains
		Le système d'IA renforce-t-il ou augmente-t-il les capacités humaines ?	Le système d'IA doit être conçu pour augmenter ou renforcer les capacités humaines, en soutenant les utilisateurs dans leurs tâches sans les remplacer ni les marginaliser.	Renforcement des capacités humaines
		Avez-vous prévu des garanties pour empêcher toute confiance ou dépendance excessives envers le système d'IA dans les processus de travail ?	Des garanties doivent être mises en place pour éviter toute confiance aveugle ou dépendance excessive des utilisateurs envers le système d'IA, notamment par des mécanismes de vérification, de supervision ou de rappel des limites du système.	Prévention de la dépendance excessive
	Contrôle humain	Avez-vous réfléchi au niveau approprié de contrôle humain pour le système d'IA et le cas d'utilisation en question ?	Le niveau de contrôle humain approprié doit être défini en fonction du cas d'usage du système d'IA, en tenant compte des risques, des contextes d'utilisation et des besoins des utilisateurs.	Définition du niveau de contrôle humain
		Pouvez-vous décrire le niveau de contrôle ou de participation humains, le cas échéant ? Qui est «l'humain aux manettes» et à quel moment y a-t-il intervention humaine, ou avec quels outils ?	Le rôle de l'humain dans le processus décisionnel doit être clairement identifié, en précisant qui intervient, à quel moment, et avec quels outils ou interfaces.	Identification de l'acteur humain responsable
		Avez-vous mis en place des mécanismes et des mesures pour garantir un contrôle ou une supervision humains potentiels de cette nature, ou pour veiller à ce que les décisions soient prises sous la responsabilité globale d'êtres humains ?	Des mécanismes doivent être mis en place pour garantir une supervision humaine effective, permettant de valider, corriger ou interrompre les décisions prises par le système d'IA.	Mécanismes de supervision humaine
		Avez-vous pris des mesures pour permettre la réalisation d'audits et résoudre des questions liées à la gouvernance de l'autonomie de l'IA ?	Le système d'IA doit intégrer des dispositifs permettant la réalisation d'audits, afin de garantir la transparence, la traçabilité et la gouvernance de son autonomie.	Auditabilité et gouvernance de l'autonomie
		Dans le cas d'un système d'IA ou d'une utilisation capables d'auto-apprentissage ou autonomes, avez-vous mis en place des mécanismes plus spécifiques de contrôle et de supervision ? - Quel type de mécanismes de détection et de réponse avez-vous mis sur pied pour évaluer le risque que des problèmes surviennent ?	Les systèmes d'IA dotés de capacités d'auto-apprentissage ou d'autonomie doivent être assortis de mécanismes spécifiques de supervision, de validation et de contrôle des dérives.	Contrôle renforcé pour les systèmes autonomes ou auto-apprenants

		<p>- Avez-vous veillé à la présence d'un « bouton d'arrêt » ou à l'existence d'une procédure pour suspendre, en cas de besoin, une opération en toute sécurité? Cette procédure suspend-elle le processus dans sa totalité, en partie, ou délègue-t-elle le contrôle à un être humain ?</p>	Des mécanismes de détection et de réponse doivent être prévus pour identifier les dysfonctionnements, les comportements inattendus ou les risques liés à l'usage du système.	Détection des anomalies et réponse aux risques
			Le système d'IA doit intégrer un bouton d'arrêt ou une procédure de suspension sécurisée, permettant d'interrompre tout ou partie de son fonctionnement en cas de besoin, avec possibilité de déléguer le contrôle à un opérateur humain.	Présence d'un mécanisme d'arrêt d'urgence
Robustesse technique et sécurité	Résilience aux attaques et sécurité	<p>Avez-vous évalué des formes d'attaques potentielles auxquelles le système d'IA pourrait être vulnérable ?</p> <ul style="list-style-type: none"> Avez-vous en particulier envisagé différents types et différentes natures de vulnérabilités, comme la pollution des données, l'infrastructure physique, les cyber attaques ? 	Le système d'IA doit faire l'objet d'une évaluation approfondie des formes d'attaques potentielles auxquelles il pourrait être vulnérable, incluant les risques liés à la pollution des données, aux infrastructures physiques et aux cyberattaques.	Évaluation des vulnérabilités potentielles
			Les différentes natures de vulnérabilités (techniques, logicielles, physiques, organisationnelles) doivent être identifiées, documentées et prises en compte dans la conception du système.	Identification des types de vulnérabilités
		Avez-vous prévu des mesures ou systèmes pour veiller à l'intégrité et à la résilience du système d'IA face à de potentielles attaques ?	Des mesures techniques et organisationnelles doivent être mises en place pour garantir l'intégrité, la robustesse et la résilience du système d'IA face aux attaques ou aux perturbations.	Mesures de résilience et d'intégrité
	Solutions de secours et sécurité générale	Avez-vous évalué le comportement de votre système dans des situations ou des environnements imprévus ?	Le système d'IA doit être testé dans des conditions imprévues ou extrêmes afin d'évaluer sa stabilité, sa fiabilité et sa capacité à réagir de manière sécurisée.	Comportement en environnement imprévu
		Avez-vous envisagé si, et dans quelle mesure, votre système pourrait avoir un double usage ? Le cas échéant, avez-vous pris des mesures préventives appropriées contre un tel cas de figure (y compris, par exemple, ne pas publier la recherche ou ne pas déployer le système)?	Si le système d'IA présente un risque de double usage (usage détourné, malveillant ou non prévu), des mesures préventives doivent être mises en œuvre, telles que la limitation de diffusion, la restriction d'accès ou la non-publication de certaines fonctionnalités.	Gestion du double usage
		Avez-vous veillé à ce que votre système dispose de suffisamment de solutions de secours pour faire face à d'éventuelles attaques antagonistes ou autres situations imprévues (par exemple, procédures de relais technique ou demande de communication avec un opérateur humain avant d'agir) ?	Le système d'IA doit intégrer des solutions de secours (fallback) permettant de maintenir la continuité du service ou de transférer le contrôle à un opérateur humain en cas de défaillance ou d'attaque.	Existence de solutions de secours

		<p>Avez-vous envisagé le niveau de risque posé par le système d'IA dans ce cas d'utilisation spécifique ?</p> <ul style="list-style-type: none"> • Avez-vous mis en place un processus pour mesurer et évaluer les risques et la sécurité ? • Avez-vous fourni les informations nécessaires en cas de risque pour l'intégrité physique humaine ? • Avez-vous réfléchi à une politique d'assurance pour couvrir les dégâts potentiels provoqués par le système d'IA ? • Avez-vous recensé les risques potentiels en matière de sécurité d'autres utilisations prévisibles de la technologie, y compris d'utilisation abusive accidentelle ou malveillante ? Existe-t-il un plan pour atténuer ou gérer ces risques ? 	Le niveau de risque associé au cas d'usage spécifique du système d'IA doit être évalué, documenté et pris en compte dans la conception et le déploiement.	Évaluation du niveau de risque contextuel
			Un processus structuré doit être mis en place pour identifier, mesurer et évaluer les risques liés à la sécurité, à la fiabilité et à l'usage du système d'IA.	Processus d'évaluation des risques et de la sécurité
			Les risques potentiels pour l'intégrité physique des personnes doivent être identifiés et communiqués aux parties prenantes concernées.	Communication des risques pour l'intégrité physique
			Une politique d'assurance ou de couverture des dommages potentiels causés par le système d'IA doit être envisagée et intégrée dans la stratégie de gestion des risques.	Politique d'assurance
			Les risques liés à des usages abusifs, accidentels ou malveillants du système doivent être recensés, et un plan d'atténuation ou de gestion de ces risques doit être mis en œuvre.	Plan de gestion des usages abusifs ou détournés
		<p>Avez-vous évalué s'il est probable que le système d'IA cause des dommages ou préjudices aux utilisateurs ou à des tiers ? Le cas échéant, avez-vous évalué la probabilité, les dommages potentiels, le public concerné et la gravité ?</p> <ul style="list-style-type: none"> • En cas de risque qu'un système d'IA cause des dommages, avez-vous réfléchi à des règles de responsabilité et de protection des consommateurs, et de quelle manière en avez-vous tenu compte ? • Avez-vous réfléchi à l'incidence potentielle ou au risque en matière de sécurité sur l'environnement ou les animaux ? • Vous êtes-vous demandé, dans le cadre de votre analyse des risques, si des problèmes de sécurité ou de réseau (par exemple, des menaces pesant sur la cybersécurité) pourraient mettre en péril la sécurité ou entraîner des préjudices du fait d'un comportement involontaire du système d'IA ? 	Le système d'IA doit faire l'objet d'une évaluation de la probabilité et de la gravité des dommages qu'il pourrait causer aux utilisateurs ou à des tiers.	Évaluation des dommages potentiels
			Des règles claires de responsabilité doivent être définies pour encadrer les dommages causés par le système, en garantissant la protection des utilisateurs et des consommateurs.	Règles de responsabilité et protection des consommateurs
			L'impact potentiel du système d'IA sur l'environnement ou les animaux doit être pris en compte dans l'analyse des risques.	Évaluation des impacts environnementaux et sur les animaux
			Les menaces liées à la cybersécurité ou aux défaillances réseau doivent être identifiées, et des mesures doivent être prévues pour éviter les comportements involontaires dangereux du système.	Analyse des menaces de cybersécurité
		<p>Avez-vous évalué l'incidence probable d'une défaillance de votre système d'IA entraînant la production de résultats erronés, l'indisponibilité de votre système, ou la production de résultats inacceptables pour la société (par exemple, pratiques discriminatoires) ?</p>	Le système doit être évalué pour anticiper les conséquences de défaillances critiques telles que des résultats erronés, une indisponibilité ou des effets socialement inacceptables (ex. : discrimination).	Évaluation des défaillances critiques
			Des seuils de risque doivent être définis pour activer des plans de secours ou des procédures d'urgence.	Définition de seuils de déclenchement

		<ul style="list-style-type: none"> Avez-vous mis en place des seuils et une gouvernance pour les scénarios ci-dessus afin de déclencher d'autres plans/solutions de secours ? Avez-vous défini et testé des solutions de secours ? 	Les solutions de secours prévues doivent être testées régulièrement pour garantir leur efficacité en cas de besoin.	Test des solutions de secours
	Précision	<p>Avez-vous évalué le niveau de précision et la définition de la précision nécessaires dans le contexte du système d'IA et du cas d'utilisation concerné ?</p> <ul style="list-style-type: none"> Avez-vous réfléchi à la manière dont la précision est mesurée et assurée ? Avez-vous mis en place des mesures pour veiller à ce que les données utilisées soient exhaustives et à jour ? Avez-vous mis en place des mesures pour évaluer si des données supplémentaires sont nécessaires, par exemple pour améliorer la précision et éliminer les biais ? 	Le niveau de précision requis pour le système d'IA doit être défini en fonction du cas d'usage, des objectifs pédagogiques ou opérationnels, et des conséquences potentielles des erreurs.	Définition du niveau de précision attendu
			Des méthodes claires et reproductibles doivent être mises en place pour mesurer et assurer la précision du système d'IA, tout au long de son cycle de vie.	Méthodes de mesure de la précision
			Les données utilisées par le système doivent être exhaustives, pertinentes et régulièrement mises à jour pour garantir la fiabilité des résultats.	Qualité et actualisation des données
			Des mécanismes doivent permettre d'identifier les besoins en données complémentaires afin d'améliorer la précision du système et de réduire les biais.	Évaluation des besoins en données supplémentaires
		Avez-vous évalué le préjudice que causeraient des prédictions inexactes du système d'IA ?	Les conséquences potentielles des prédictions inexactes doivent être évaluées, notamment en termes d'impact sur les utilisateurs, les décisions prises ou les parcours éducatifs.	Évaluation des préjudices liés aux erreurs
		Avez-vous prévu des moyens de mesurer si votre système produit un nombre inacceptable de prédictions inexactes ?	Des indicateurs doivent être mis en place pour mesurer la fréquence et la gravité des prédictions inexactes, et pour identifier les seuils d'acceptabilité.	Suivi du taux d'erreurs
		En cas de prédictions inexactes, avez-vous mis en place une série d'étapes pour résoudre le problème ?	En cas de prédictions inexactes, une procédure de correction ou d'amélioration continue doit être prévue, incluant des étapes de diagnostic, d'ajustement et de validation.	Plan de correction des erreurs
	Fiabilité et reproductibilité	<p>Avez-vous mis en place une stratégie afin de contrôler le système d'IA et de vous assurer qu'il répond aux objectifs, aux finalités et aux applications prévues ?</p> <ul style="list-style-type: none"> Avez-vous vérifié si des contextes spécifiques ou conditions particulières doivent être pris en compte pour garantir la reproductibilité ? Avez-vous mis en place des processus ou méthodes de vérification pour mesurer et garantir les différents aspects de la fiabilité et de la reproductibilité ? Avez-vous mis en place des processus visant à décrire certains réglages susceptibles d'entraîner une défaillance du système d'IA ? Avez-vous clairement documenté et appliqué ces processus aux fins des essais et de la vérification de la fiabilité du système d'IA ? 	Une stratégie claire de contrôle du système d'IA doit être définie afin de garantir qu'il répond aux objectifs, aux finalités et aux cas d'usage prévus.	Stratégie de contrôle du système
			Les contextes d'utilisation particuliers ou les conditions spécifiques susceptibles d'influencer la performance ou la reproductibilité du système doivent être identifiés et intégrés dans la conception et les tests.	Prise en compte des contextes spécifiques
			Des processus de vérification doivent être mis en place pour mesurer et garantir la fiabilité et la reproductibilité du système dans différents contextes d'usage.	Méthodes de vérification de la fiabilité et de la reproductibilité
			Les réglages ou paramètres susceptibles d'entraîner une défaillance du système doivent être identifiés, documentés et surveillés.	Documentation des réglages critiques
			Les processus de test et de vérification de la fiabilité doivent être appliqués de manière rigoureuse et documentée, avec des preuves de conformité.	Application des processus de test

		<ul style="list-style-type: none"> Avez-vous mis en place un mécanisme ou une communication pour garantir aux utilisateurs (finaux) la fiabilité du système d'IA ? 	Des mécanismes de communication doivent être prévus pour informer les utilisateurs finaux du niveau de fiabilité du système, de ses limites et des conditions optimales d'utilisation.	Communication de la fiabilité aux utilisateurs
Respect de la vie privée et gouvernance des données	Respect de la vie privée et protection des données	En fonction du cas d'utilisation, avez-vous mis sur pied un mécanisme permettant à autrui de signaler des problèmes en rapport avec le respect de la vie privée et la protection des données durant les processus suivis par le système d'IA pour la collecte des données (aux fins de l'entraînement et du fonctionnement) et le traitement des données ?	Un mécanisme accessible doit être mis en place pour permettre aux utilisateurs ou parties prenantes de signaler tout problème lié à la protection des données personnelles, que ce soit lors de la collecte, du traitement ou de l'entraînement du système d'IA.	Mécanisme de signalement des atteintes à la vie privée
		Avez-vous évalué le type et la portée des données constituant vos ensembles de données (par exemple, si elles contiennent des données à caractère personnel) ?	Le type de données collectées, y compris leur caractère personnel ou sensible, doit être évalué et documenté afin d'anticiper les risques liés à leur traitement.	Évaluation du type et de la portée des données
		Avez-vous réfléchi à des manières de mettre au point le système d'IA ou d'entraîner le modèle sans utiliser (ou en utilisant de manière limitée) des données potentiellement sensibles ou à caractère personnel ?	Le système d'IA doit être conçu de manière à limiter, voire éviter, l'utilisation de données sensibles ou à caractère personnel, en privilégiant des alternatives techniques respectueuses de la vie privée.	Minimisation de l'usage de données sensibles
		Avez-vous intégré des mécanismes de notification et de contrôle concernant les données à caractère personnel en fonction du cas d'utilisation (comme un consentement valable et la possibilité de révoquer le consentement, le cas échéant) ?	Des mécanismes doivent être intégrés pour garantir un consentement explicite, éclairé et révocable des utilisateurs concernant l'usage de leurs données personnelles.	Mécanismes de consentement et de contrôle
		Avez-vous pris des mesures pour renforcer le respect de la vie privée, par exemple des mesures de cryptage, d'anonymisation et d'agrégation ?	Des mesures techniques telles que le chiffrement, l'anonymisation ou l'agrégation des données doivent être mises en œuvre pour renforcer la confidentialité et la sécurité des données traitées.	Mesures de protection des données
		Lorsqu'il existe un responsable de la protection des données, avez-vous mobilisé cette personne à un stade précoce dans le processus ?	Lorsqu'un délégué à la protection des données (DPO) est désigné, il doit être impliqué dès les premières phases de conception du système afin d'assurer la conformité aux réglementations en vigueur.	Mobilisation du DPO (Data Protection Officer)
	Qualité et des données	Avez-vous aligné votre système sur d'éventuelles normes pertinentes (par exemple, ISO, IEEE) ou des protocoles largement adoptés dans le cadre de votre gestion et de votre gouvernance quotidiennes des données ?	Le système d'IA doit être aligné sur des normes techniques et éthiques reconnues (ex. : ISO, IEEE) ou sur des protocoles largement adoptés en matière de gouvernance des données.	Alignement sur les normes reconnues
		Avez-vous mis sur pied des mécanismes de contrôle pour la collecte, le stockage, le traitement et l'utilisation des données ?	Des mécanismes de contrôle doivent être mis en place pour encadrer la collecte, le stockage, le traitement et l'utilisation des données, en garantissant leur conformité et leur sécurité.	Mécanismes de contrôle des données
		Avez-vous évalué la mesure dans laquelle vous contrôlez la qualité des sources externes des données utilisées ?	La qualité, la fiabilité et la légitimité des sources de données externes utilisées par le système doivent être évaluées et documentées.	Évaluation de la qualité des sources externes

	Accès aux données	Avez-vous mis en place des processus pour garantir la qualité et l'intégrité de vos données ? Avez-vous envisagé d'autres processus ? De quelle manière vérifiez-vous que vos ensembles de données n'ont pas été compromis ou piratés ?	Des processus doivent être mis en œuvre pour assurer la qualité, la cohérence et l'intégrité des données tout au long du cycle de vie du système.	Garantie de la qualité et de l'intégrité des données
			Des procédures doivent permettre de vérifier que les ensembles de données n'ont pas été compromis ou piratés, notamment via des audits, des tests de vulnérabilité ou des systèmes de surveillance.	Vérification de la sécurité des ensembles de données
		Quels protocoles, processus et procédures ont été suivis pour gérer et garantir la gouvernance appropriée des données ?	Les protocoles, processus et procédures mis en place pour assurer la gouvernance des données doivent être clairement définis, documentés et alignés avec les bonnes pratiques en matière de gestion des données.	Documentation des protocoles de gouvernance des données
		<ul style="list-style-type: none"> Avez-vous analysé qui peut accéder aux données des utilisateurs et dans quelles circonstances ? Avez-vous veillé à ce que ces personnes soient qualifiées, qu'elles aient effectivement besoin d'accéder aux données et à ce qu'elles disposent des compétences nécessaires pour comprendre précisément la politique de protection des données ? Avez-vous prévu un mécanisme de contrôle pour consigner quand, où, comment, par qui et dans quel but les données ont été consultées ? 	Une analyse précise des personnes ou entités ayant accès aux données des utilisateurs doit être réalisée, en précisant les conditions d'accès et les finalités associées.	Analyse des accès aux données
			Seules des personnes qualifiées, dûment autorisées et formées à la protection des données doivent pouvoir accéder aux données, en fonction de leur rôle et de la nécessité opérationnelle.	Qualification et légitimité des personnes ayant accès aux données
Transparence	Traçabilité		Un mécanisme de journalisation doit être mis en place pour consigner de manière systématique les accès aux données : qui y accède, quand, comment, à partir de quel terminal, et dans quel but.	Traçabilité des accès aux données
		Avez-vous mis des mesures en place susceptibles de garantir la traçabilité ? Cela pourrait consister à documenter :	Le système d'IA doit intégrer des mesures concrètes permettant d'assurer la traçabilité de son fonctionnement, de sa conception à son déploiement.	Mise en place de mesures de traçabilité
		<ul style="list-style-type: none"> Les méthodes appliquées aux fins de la conception et de la mise au point du système algorithmique : <ul style="list-style-type: none"> Dans le cas d'un système d'IA fondé sur des règles, la méthode de programmation ou la manière dont le modèle a été mis au point devraient être documentées; Dans le cas d'un système d'IA fondé sur l'apprentissage, la méthode d'entraînement de l'algorithme, y compris quelles données d'entrée ont été collectées et sélectionnées, et dans quelles conditions, devrait être documentée. Les méthodes appliquées pour tester et valider le système algorithmique : <ul style="list-style-type: none"> Dans le cas d'un système d'IA fondé sur des règles, les scénarios ou cas utilisés pour tester et valider devraient être documentés ; Dans le cas d'un système d'IA fondé sur l'apprentissage, les informations relatives aux données utilisées pour tester et valider devraient être documentées. Les résultats du système algorithmique : 	Les méthodes utilisées pour la conception et la mise au point du système algorithmique doivent être documentées de manière détaillée, qu'il s'agisse de systèmes fondés sur des règles ou sur l'apprentissage automatique.	Documentation des méthodes de conception
			Pour les systèmes d'IA fondés sur l'apprentissage, la méthode d'entraînement doit être explicitement décrite, incluant les données d'entrée utilisées, leur origine, les critères de sélection et les conditions de collecte.	Documentation de l'entraînement des modèles (IA fondée sur l'apprentissage)
			Les méthodes utilisées pour tester et valider le système doivent être formalisées, incluant les scénarios de test pour les systèmes à base de règles et les jeux de données pour les systèmes d'apprentissage.	Documentation des méthodes de test et de validation

		<ul style="list-style-type: none"> Les résultats d'un algorithme ou les décisions qu'il prend, ainsi que les éventuelles autres décisions qui résulteraient de différents cas (par exemple, pour d'autres sous-groupes d'utilisateurs) devraient être documentés. 	Les résultats produits par le système, ainsi que les décisions prises (y compris les variations selon les profils ou les cas d'usage), doivent être enregistrés, analysables et accessibles pour audit ou justification.	Documentation des résultats et des décisions
	Explicabilité	Avez-vous évalué la mesure dans laquelle les décisions prises, et donc les résultats obtenus, par le système d'IA peuvent être compris	Le système d'IA doit être conçu de manière à ce que les décisions qu'il prend soient compréhensibles par les utilisateurs, en particulier ceux qui sont directement concernés par les résultats.	Compréhensibilité des décisions du système
		Avez-vous veillé à ce qu'une explication de la raison pour laquelle un système a procédé à un certain choix entraînant un certain résultat puisse être rendue compréhensible pour l'ensemble des utilisateurs qui pourraient souhaiter obtenir une explication ?	Une explication claire et accessible doit être fournie pour chaque décision ou résultat généré par le système, afin de permettre aux utilisateurs de comprendre pourquoi une réponse ou une action a été produite.	Explication des choix algorithmiques
		Avez-vous évalué la mesure dans laquelle la décision du système influence les processus décisionnels de l'organisation ?	L'impact des décisions du système d'IA sur les processus décisionnels humains ou organisationnels doit être évalué, notamment en termes de délégation, d'automatisation ou de transformation des pratiques.	Évaluation de l'influence sur les processus décisionnels
		Avez-vous évalué pourquoi ce système particulier a été déployé dans ce domaine spécifique ?	Le choix de déployer le système dans un domaine donné doit être justifié par une analyse de pertinence, tenant compte des besoins, des risques et des bénéfices attendus.	Justification du déploiement dans un domaine spécifique
		Avez-vous évalué le modèle économique concernant ce système (par exemple, en quoi crée-t-il de la valeur pour l'organisation) ?	Le modèle économique du système d'IA doit être explicité, en précisant comment il crée de la valeur pour l'organisation et quels sont les enjeux éthiques ou stratégiques associés.	Évaluation du modèle économique
		Avez-vous conçu le système d'IA en ayant dès le départ l'interprétation à l'esprit ?	Le système doit être conçu dès le départ avec l'objectif d'être interprétable, en privilégiant des modèles simples ou des techniques d'explication adaptées à l'usage.	Conception orientée vers l'interprétabilité
		<ul style="list-style-type: none"> Avez-vous cherché à utiliser le modèle le plus simple et le plus facile à interpréter pour l'application en question ? Avez-vous évalué si vous êtes en mesure d'analyser les données que vous avez utilisées aux fins de l'entraînement et des essais ? Cela peut-il être modifié et actualisé au fil du temps ? 	<p>Lorsque cela est possible, des modèles simples et facilement interprétables doivent être privilégiés, en particulier dans les contextes à fort enjeu éthique ou pédagogique.</p> <p>Les données utilisées pour entraîner et tester le système doivent être analysables, modifiables et mises à jour régulièrement pour garantir la transparence et l'adaptabilité du système.</p>	<p>Choix de modèles interprétables</p> <p>Analyse des données d'entraînement et de test</p>
		Avez-vous évalué si des solutions s'offrent à vous suite à l'entraînement et à la mise au point du modèle pour examiner l'interprétation ou si vous avez accès à la séquence des opérations du modèle ?	Des solutions doivent être envisagées pour permettre l'examen de la logique interne du modèle, que ce soit par des outils d'explicabilité, des visualisations ou des audits techniques.	Accès à la logique interne du modèle
	Communication	Avez-vous informé les utilisateurs (finaux) – au moyen d'une clause de non- responsabilité ou de tout autre moyen – qu'ils interagissent avec un système d'IA et pas avec un autre être humain ? Avez-vous indiqué clairement que votre système est doté de l'IA ?	Le système doit informer explicitement les utilisateurs qu'ils interagissent avec une intelligence artificielle, et non avec un être humain, via des mentions visibles, des icônes ou des messages contextuels.	Identification claire du système comme étant une IA
		Avez-vous mis en place des mécanismes pour informer les utilisateurs des raisons et critères expliquant les résultats du système d'IA ?	Les utilisateurs doivent être informés de manière claire et intelligible des critères et logiques qui sous-tendent les résultats ou recommandations générés par le système d'IA.	Explication des critères de décision
		<ul style="list-style-type: none"> Les utilisateurs visés en sont-ils informés de manière claire et intelligible ? 	Les explications fournies doivent être compréhensibles pour tous les profils d'utilisateurs, en tenant compte de leur niveau de littératie numérique ou de leur rôle (étudiant, enseignant, administrateur...).	Accessibilité de l'information pour tous les profils d'utilisateurs

		<ul style="list-style-type: none"> Avez-vous établi des processus pour tenir compte des commentaires des utilisateurs et utiliser ces commentaires pour adapter le système ? Avez-vous également communiqué les risques potentiels ou perçus, tels que les biais ? En fonction du cas d'utilisation, avez-vous également réfléchi à la communication et à la transparence envers d'autres publics, des tiers ou le grand public ? 	Des mécanismes doivent être mis en place pour recueillir les commentaires des utilisateurs et adapter le système en fonction de ces retours (ex. : signalement d'erreurs, suggestions d'amélioration).	Prise en compte des retours utilisateurs
		<p>Avez-vous clairement indiqué la finalité du système d'IA et qui ou ce qui pourrait bénéficier du produit/service ?</p> <ul style="list-style-type: none"> Les scénarios d'utilisation du produit ont-ils été définis et clairement expliqués, en envisageant également d'autres moyens de communication pour veiller à ce qu'ils soient compréhensibles et appropriés pour le destinataire ? En fonction du cas d'utilisation, avez-vous réfléchi à la psychologie humaine et aux potentielles limites humaines, comme le risque de confusion, les biais de confirmation ou la fatigue cognitive ? 	Les risques potentiels liés à l'utilisation du système (biais, erreurs, limitations) doivent être communiqués de manière transparente aux utilisateurs.	Communication des risques et biais
			En fonction du cas d'usage, la communication sur le fonctionnement et les limites du système doit être adaptée à d'autres publics concernés (ex. : parents, décideurs, grand public).	Communication élargie selon les publics
			La finalité du système d'IA, ses objectifs et les bénéficiaires visés doivent être clairement indiqués dans la documentation ou l'interface utilisateur.	Indication de la finalité du système
			Les scénarios d'utilisation du système doivent être définis, illustrés et expliqués de manière à être compréhensibles pour les utilisateurs finaux, avec des exemples concrets.	Définition et explication des scénarios d'usage
			La conception du système doit tenir compte des biais cognitifs, de la fatigue mentale ou du risque de confusion, en adaptant l'interface et les messages à ces contraintes.	Prise en compte des limites cognitives humaines
		<p>Avez-vous clairement expliqué les caractéristiques, les limites et les éventuelles lacunes du système d'IA :</p> <ul style="list-style-type: none"> S'agissant de la mise au point : à toute personne chargée de son déploiement pour en faire un produit ou service? S'agissant du déploiement : à l'utilisateur final ou au consommateur? 	Les caractéristiques, limites et lacunes du système doivent être expliquées à la fois aux personnes en charge de son déploiement (ex. : administrateurs) et aux utilisateurs finaux (ex. : étudiants, enseignants).	Communication différenciée selon les rôles
Diversité, non-discrimination et équité	Eviter les biais injustes	<p>Avez-vous prévu une stratégie ou un ensemble de procédures pour éviter de créer ou de renforcer des biais injustes dans le système d'IA, en ce qui concerne tant l'utilisation des données d'entrée que la conception de l'algorithme ?</p> <ul style="list-style-type: none"> Avez-vous évalué et reconnu les éventuelles limites provenant de la composition des ensembles de données utilisés ? Avez-vous réfléchi à la diversité et à la représentativité des utilisateurs dans les données ? Avez-vous procédé à des essais portant sur des populations spécifiques ou des cas d'utilisation problématiques ? Avez-vous recherché et utilisé les outils techniques disponibles pour améliorer votre compréhension des données, du modèle et de la performance ? 	Une stratégie claire et documentée doit être mise en place pour éviter la création ou le renforcement de biais injustes, tant dans les données d'entrée que dans la conception des algorithmes.	Stratégie de prévention des biais
			Les limites des ensembles de données utilisés (ex. : manque de diversité, représentativité insuffisante) doivent être identifiées, reconnues et prises en compte dans la conception du système.	Évaluation des limites des ensembles de données
			La diversité et la représentativité des utilisateurs doivent être intégrées dans les données et les tests, notamment via des essais sur des populations spécifiques ou des cas d'usage sensibles.	Prise en compte de la diversité des utilisateurs
			Des outils techniques doivent être mobilisés pour analyser les biais potentiels dans les données, les modèles et les performances du système.	Utilisation d'outils d'analyse des biais

		<ul style="list-style-type: none"> Avez-vous mis en place des processus pour tester et contrôler les biais éventuels au cours de la phase de mise au point, de déploiement et d'utilisation du système ? 	Des processus doivent être mis en place pour tester, détecter et corriger les biais à chaque étape du cycle de vie du système : conception, déploiement, utilisation.	Processus de détection et de contrôle des biais
		En fonction du cas d'utilisation, avez-vous prévu un mécanisme permettant à autrui de signaler des problèmes liés aux biais, à la discrimination ou aux mauvaises performances du système d'IA ?	Un mécanisme clair et accessible doit permettre aux utilisateurs ou parties prenantes de signaler les biais, discriminations ou mauvaises performances du système.	Mécanisme de signalement des biais
		<ul style="list-style-type: none"> Avez-vous envisagé des mesures et des moyens de communication clairs pour savoir comment et à qui ces problèmes peuvent être signalés ? 	Des moyens de communication doivent être prévus pour informer les utilisateurs sur la manière de signaler un biais, à qui s'adresser et comment le système y répond.	Communication sur les biais et les recours
		<ul style="list-style-type: none"> Avez-vous tenu compte non seulement des utilisateurs (finaux) mais également des autres personnes susceptibles d'être indirectement affectées par le système d'IA ? 	Le système doit être conçu en tenant compte non seulement des utilisateurs directs, mais aussi des personnes ou groupes susceptibles d'être affectés indirectement.	Prise en compte des effets indirects
		Avez-vous évalué si, dans des conditions identiques, une éventuelle variabilité des décisions est possible ?	Le système doit être évalué pour détecter toute variabilité injustifiée dans les décisions rendues dans des conditions identiques.	Évaluation de la variabilité des décisions
		<ul style="list-style-type: none"> Le cas échéant, avez-vous réfléchi aux causes probables ? 	En cas de variabilité, les causes doivent être identifiées, analysées et documentées.	Analyse des causes de variabilité
		<ul style="list-style-type: none"> Concernant la variabilité, avez-vous mis sur pied un mécanisme de mesure ou d'évaluation de l'incidence potentielle de cette variabilité sur les droits fondamentaux ? 	Un mécanisme doit permettre d'évaluer l'impact potentiel de la variabilité des décisions sur les droits fondamentaux des utilisateurs.	Évaluation de l'impact de la variabilité sur les droits fondamentaux
		Avez-vous prévu une définition appropriée de l'«équité» que vous appliquez dans la conception des systèmes d'IA ?	Une définition claire et justifiée de l'équité doit être adoptée pour guider la conception du système, en tenant compte des standards existants et des alternatives possibles.	Définition de l'équité appliquée
		<ul style="list-style-type: none"> Votre définition est-elle couramment utilisée ? Avez-vous envisagé d'autres définitions avant de choisir celle-ci ? 	Des indicateurs quantitatifs ou qualitatifs doivent être mis en place pour mesurer et tester l'équité du système.	Mesure de l'équité
		<ul style="list-style-type: none"> Avez-vous prévu une analyse quantitative ou des indicateurs pour mesurer et tester la définition appliquée de l'équité ? 	Des mécanismes concrets doivent être intégrés pour garantir l'équité dans les décisions du système, et des alternatives doivent être envisagées si nécessaire.	Mécanismes de garantie de l'équité
		<ul style="list-style-type: none"> Avez-vous mis sur pied des mécanismes visant à garantir l'équité dans vos systèmes d'IA ? Avez-vous envisagé d'autres mécanismes potentiels ? 		
	Accessibilité et conception universelle	Avez-vous veillé à ce que le système d'IA réponde aux besoins d'un large ensemble de préférences et de capacités individuelles ?	Le système d'IA doit être conçu pour répondre aux besoins d'un large éventail de préférences, de profils cognitifs et de capacités individuelles, afin de garantir une accessibilité universelle.	Adaptation aux préférences et capacités individuelles

		<ul style="list-style-type: none"> • Avez-vous évalué si le système d'IA peut être utilisé par les personnes présentant des besoins spécifiques ou un handicap ou qui sont exposées au risque d'exclusion ? Comment cet aspect a-t-il été intégré à la conception du système et comment est-il vérifié ? • Avez-vous veillé à ce que les informations relatives au système d'IA soient également accessibles aux utilisateurs de technologies d'assistance ? • Avez-vous mobilisé ou consulté cette communauté d'utilisateurs au cours de la phase de mise au point du système d'IA ? 	Le système doit être utilisable par des personnes en situation de handicap ou à risque d'exclusion, avec des fonctionnalités adaptées et des vérifications de compatibilité intégrées dès la conception.	Prise en compte des besoins spécifiques et du handicap
			Les informations et interfaces du système d'IA doivent être accessibles via des technologies d'assistance (lecteurs d'écran, navigation vocale, etc.), conformément aux standards d'accessibilité.	Compatibilité avec les technologies d'assistance
			Les utilisateurs ayant des besoins spécifiques doivent être consultés ou impliqués dans la phase de conception ou de test du système, afin de garantir que leurs attentes sont prises en compte.	Consultation des communautés concernées
		Avez-vous tenu compte de l'incidence de votre système d'IA sur le groupe d'utilisateurs potentiels ?	L'impact du système sur les différents groupes d'utilisateurs potentiels doit être évalué, notamment en ce qui concerne les risques d'exclusion ou de traitement inéquitable.	Évaluation de l'impact sur les groupes d'utilisateurs
		<ul style="list-style-type: none"> • L'équipe participant à la mise au point du système d'IA est-elle représentative de votre groupe cible d'utilisateurs ? Est-elle représentative de la population au sens large, compte tenu également d'autres groupes susceptibles d'être indirectement concernés ? • Avez-vous évalué si certaines personnes ou certains groupes pourraient subir de manière disproportionnée des effets négatifs ? • D'autres équipes ou groupes présentant différents parcours professionnels et expériences vous ont-ils fait parvenir des • réactions ? 	L'équipe en charge du développement du système doit être représentative de la diversité des utilisateurs finaux, ou inclure des profils variés pour anticiper les effets indirects.	Diversité de l'équipe de conception
			Le système doit être évalué pour identifier si certains groupes pourraient subir des effets négatifs disproportionnés, et des mesures doivent être prises pour les atténuer.	Évaluation des effets disproportionnés
			Les retours d'équipes ou de groupes externes ayant des parcours professionnels ou des expériences variées doivent être pris en compte pour enrichir la conception et anticiper les angles morts.	Prise en compte des retours de groupes externes
	Participation des parties prenantes	Avez-vous réfléchi à un mécanisme pour inclure la participation de différentes parties prenantes dans la mise au point et l'utilisation du système d'IA ?	Un mécanisme structuré doit être mis en place pour inclure la participation active de différentes parties prenantes (utilisateurs finaux, experts, représentants institutionnels, etc.) dans la conception, le développement et l'évaluation du système d'IA.	Participation des parties prenantes

		Avez-vous préparé la voie à l'introduction du système d'IA au sein de votre organisation en informant et en mobilisant au préalable les travailleurs concernés et leurs représentants ?	L'introduction du système d'IA dans une organisation doit être précédée d'une phase d'information, de sensibilisation et de mobilisation des travailleurs concernés et de leurs représentants, afin de garantir leur adhésion, leur compréhension et leur capacité à interagir avec le système.	Mobilisation des acteurs internes
Bien-être sociétal et environnemental	IA durable et respectueuse de l'environnement	Avez-vous mis en place des mécanismes pour mesurer l'impact environnemental de la mise au point, du déploiement et de l'utilisation du système d'IA (par exemple, énergie consommée par les centres de données, type d'énergie consommée par les centres de données, etc.)?	Des mécanismes doivent être mis en place pour mesurer l'impact environnemental du système d'IA tout au long de son cycle de vie, incluant la consommation énergétique des centres de données et le type d'énergie utilisée.	Mesure de l'impact environnemental
		Avez-vous prévu des mesures pour réduire l'impact environnemental du cycle de vie de votre système d'IA ?	Des mesures concrètes doivent être prévues pour réduire l'impact environnemental du système d'IA, notamment en optimisant les ressources, en limitant les calculs intensifs et en favorisant des infrastructures durables.	Réduction de l'empreinte écologique
	Incidence sociale	Lorsque le système d'IA interagit directement avec des êtres humains: <ul style="list-style-type: none"> Avez-vous évalué si le système d'IA encourage les êtres humains à développer de l'attachement et de l'empathie pour le système ? Avez-vous veillé à ce que le système d'IA indique clairement que son interaction sociale est simulée et qu'il n'a nullement la capacité de «comprendre» et de «ressentir»? 	Lorsque le système interagit directement avec des humains, une évaluation doit être menée pour déterminer s'il suscite un attachement ou une empathie non souhaitée, pouvant influencer les comportements ou les décisions.	Évaluation de l'attachement émotionnel
			Le système d'IA doit indiquer de manière explicite que ses interactions sociales sont simulées et qu'il ne possède ni conscience, ni émotions, ni capacité de compréhension humaine.	Clarté sur la nature simulée de l'interaction
		Avez-vous veillé à ce que les incidences sociales du système d'IA soient bien comprises ? Par exemple, vous êtes-vous demandé s'il existe un risque de perte d'emplois et de perte de compétences de la main-d'œuvre ? Quelles mesures ont été prises pour contrer ces risques ?	Les incidences sociales du système d'IA doivent être évaluées, notamment les risques de perte d'emploi, de déqualification ou de transformation des compétences, et des mesures doivent être prises pour les anticiper ou les compenser.	Analyse des incidences sociales
	Société et démocratie	Avez-vous évalué l'incidence plus large de l'utilisation du système d'IA sur la société, au-delà de l'utilisateur (final) individuel, par exemple les parties prenantes susceptibles d'être indirectement concernées ?	L'impact du système d'IA doit être évalué au-delà de l'utilisateur final, en tenant compte des effets indirects sur d'autres parties prenantes, groupes sociaux ou dynamiques collectives.	Évaluation de l'impact sociétal élargi
Responsabilité	Audibilité	Avez-vous mis en place des mécanismes pour faciliter l'auditabilité du système par des acteurs internes et/ou indépendants, en veillant pas exemple à la traçabilité et à la journalisation des processus et des résultats du système d'IA ?	Le système d'IA doit intégrer des mécanismes permettant son auditabilité par des acteurs internes ou externes indépendants. Cela inclut la mise en place de dispositifs de traçabilité et de journalisation des processus, des décisions et des résultats produits par le système, afin de garantir la transparence, la responsabilité et la conformité réglementaire.	Auditabilité du système

	Minimisation et documentation des incidences négatives	Avez-vous réalisé une analyse des risques ou de l'impact du système d'IA qui tienne compte de différentes parties prenantes qui sont directement et indirectement concernées ?	Une analyse des risques ou de l'impact du système d'IA doit être réalisée en tenant compte des parties prenantes directement et indirectement concernées, incluant les utilisateurs finaux, les opérateurs, les institutions et les tiers affectés.	Analyse des risques multi-acteurs
		Avez-vous mis en place des cadres de formation et d'éducation pour définir des pratiques en matière de responsabilité ?	Des cadres de formation et de sensibilisation doivent être mis en place pour les travailleurs impliqués dans la conception, le déploiement ou l'usage du système, afin de promouvoir des pratiques responsables.	Mise en place de formations à la responsabilité
		<ul style="list-style-type: none"> Quels travailleurs ou branches de travailleurs sont concernés ? Le sont-ils au-delà de la phase de mise au point ? 	Les formations doivent être adaptées aux différents profils de travailleurs concernés, y compris ceux intervenant après la phase de mise au point (ex. : maintenance, supervision, accompagnement pédagogique).	Identification des publics concernés par la formation
		<ul style="list-style-type: none"> Ces formations portent-elles également sur le cadre juridique potentiellement applicable au système d'IA ? 	Les formations doivent inclure une dimension juridique, en expliquant les obligations réglementaires applicables au système d'IA (ex. : RGPD, IA Act, droit à l'explication).	Intégration du cadre juridique dans les formations
		<ul style="list-style-type: none"> Avez-vous envisagé la mise sur pied d'un « comité d'examen pour l'IA éthique » ou d'un mécanisme similaire pour discuter des pratiques globales en matière de responsabilité et d'éthique, y compris des zones grises potentiellement floues ? 	Un comité d'examen éthique ou un dispositif équivalent doit être instauré pour discuter des enjeux de responsabilité, des dilemmes éthiques et des zones grises identifiées dans le développement ou l'usage du système.	Mise en place d'un comité d'éthique ou d'un mécanisme équivalent
		Outre les initiatives ou cadres internes destinés à contrôler l'éthique et la responsabilité, des orientations externes ou des processus d'audit ont-ils également été mis en place ?	En complément des dispositifs internes, des audits externes ou des référentiels indépendants doivent être mobilisés pour garantir la conformité éthique et réglementaire du système.	Recours à des audits ou orientations externes
		Existe-t-il des processus permettant aux tiers (par exemple, fournisseurs, consommateurs, distributeurs/vendeurs) ou aux travailleurs de signaler de possibles vulnérabilités, risques ou biais dans le système/l'application d'IA ?	Un processus clair et accessible doit permettre aux tiers (fournisseurs, utilisateurs, distributeurs, etc.) de signaler les vulnérabilités, biais ou risques identifiés dans le système d'IA.	Mécanisme de signalement des vulnérabilités
	Documentation des arbitrages	Avez-vous mis sur pied un mécanisme permettant de recenser les intérêts et les valeurs pertinents concernés par le système d'IA et les éventuels arbitrages entre eux ?	Un mécanisme structuré doit être mis en place pour identifier les valeurs humaines, les intérêts des parties prenantes et les principes éthiques concernés par le système d'IA, dès les premières phases de conception.	Mécanisme de recensement des valeurs et intérêts
		Quel processus utilisez-vous pour prendre des décisions relatives à ces arbitrages ? Avez-vous veillé à ce que les décisions d'arbitrage soient documentées ?	Un processus clair et justifiable doit être défini pour prendre des décisions en cas de conflit entre valeurs ou intérêts (ex. : performance vs. équité, transparence vs. sécurité).	Processus de décision pour les arbitrages
			Les arbitrages réalisés entre différentes valeurs ou exigences doivent être documentés de manière transparente, en précisant les justifications, les alternatives envisagées et les impacts anticipés.	Documentation des arbitrages
	Voies de recours	Avez-vous mis en place un ensemble approprié de mécanismes permettant un recours en cas de préjudice ou d'effet néfaste ?	Un ensemble de mécanismes accessibles et appropriés doit être mis en place pour permettre aux utilisateurs ou parties concernées d'exercer un recours en cas de préjudice, d'erreur ou d'effet néfaste causé par le système d'IA.	Mécanismes de recours en cas de préjudice

		Avez-vous mis en place des mécanismes pour fournir des informations aux utilisateurs (finaux)/tiers à propos des possibilités de recours ?	Le système doit fournir aux utilisateurs finaux et aux tiers des informations claires, compréhensibles et facilement accessibles sur les possibilités de recours, les procédures à suivre et les interlocuteurs à contacter.	Information sur les voies de recours
--	--	--	--	--------------------------------------

Notre grille d'analyse basé sur notre cadre théorique

Nous souhaitons confronté les critères que nous avons observé dans notre cadre théorique avec les critères définis dans l'outil ALTAI afin d'avoir une nouvelle grille

Variables (+ objectifs)	Besoins /critères	Exigences	But	Approches de notre cadre théorique utilisé
<u>Accès à l'information</u> <i>(garantir que les utilisateurs disposent des informations nécessaires pour comprendre, utiliser et interagir avec le système)</i>	Transparence envers les déployeurs	Fourniture d'une notice d'information claire sur le fonctionnement, les risques, les limites et les conditions d'utilisation du système.	Permettre aux utilisateurs d'accéder à des informations essentielles avant l'usage	Approche réglementaire
	Contrôle humain	Interface permettant une supervision humaine continue et une intervention manuelle en cas de besoin.	Assure un accès en temps réel à l'état du système et permet une interaction humaine.	Approche réglementaire
	Design centré utilisateur (VSD)	Intégration d'experts en valeurs dans les équipes de conception.	Favorise la prise en compte de perspectives humaines dans la conception, ce qui améliorer la qualité de l'information transmise.	Approche éthique by design
	Design centré utilisateur (VSD)	Prise en compte du contexte cognitif et social d'usage.	Permet d'adapter l'information aux utilisateurs selon leur contexte.	Approche éthique by design
	Design centré utilisateur (VSD)	Conception d'interfaces favorisant la compréhension et l'appropriation.	Vise à rendre l'information accessible et compréhensible.	Approche éthique by design
	Accès à la documentation	Fournir une documentation claire, complète et facilement accessible intégrer directement dans l'interface utilisateur.	Permettre à l'utilisateur de comprendre le fonctionnement du système sans devoir chercher des ressources externes.	Approche transparence interactionnelle
	Visualisation pédagogique	Intégrer des rapports, graphiques, tutoriels, infobulles, etc. pour présenter les objectifs pédagogiques de manière visuelle et compréhensible.	Aider l'utilisateur à comprendre les finalités éducatives du système et à s'approprier les outils.	Approche transparence interactionnelle
	Simplicité d'usage	Réduire la complexité de l'interface et organiser les fonctionnalités de manière intuitive.	Faciliter la prise en main du système, même pour les utilisateurs novices.	Approche transparence interactionnelle
	Feedback formateur	Fournir un retour d'information clair et utile avec pour objectif d'aider l'utilisateur à progresser dans ses compétences.	Renforcer l'apprentissage par l'usage du système.	Approche transparence interactionnelle
	Accessibilité cognitive	Eviter la surcharge mentale en adaptant la densité d'information à la capacité de traitement de l'utilisateur.	Maintenir une expérience fluide et non anxiogène.	Approche transparence interactionnelle

<u>Traçabilité des décisions</u> <i>(Pourvoir retracer les décisions prises par le système, leurs causes et les responsabilités associées)</i>	Journalisation des événements	Enregistrement systématique des événements critiques pour assurer la traçabilité.	Permettre de suivre les événements clés pour comprendre les décisions.	Approche réglementaire
	Documentation technique	Plans d'atténuation des risques et d'intervention en cas de défaillance.	Documenter les réponses prévus aux incidents, ce qui contribue à la traçabilité.	Approche réglementaire
	Responsabilité sociale	Partage de la responsabilité entre concepteurs, utilisateurs et systèmes.	Clarifier les responsabilités dans la chaîne de décision.	Approche éthique by design
	Gouvernance éthique	Cartographie des parties prenantes.	Identifier les acteurs impliqués dans les décisions et leurs rôles.	Approche éthique by design
	Recours d'utilisateur	Permettre à l'utilisateur de signaler une erreur ou un biais en offrant un mécanisme de révision ou de contestation.	Garantir une transparence sur les responsabilités et renforcer la confiance.	Approche transparence interactionnelle
	Communication sur les biais	Informar l'utilisateur des limites ou biais potentiels du système en rendant visible les zones d'incertitudes.	Eviter les malentendus et permettre une interprétation critique des résultats.	Approche transparence interactionnelle
	Ethique by design	Intégrer les principes éthiques dès la conception de l'interface et prendre en compte les impacts sociaux et éducatifs.	Assurer une utilisation responsable et équitable du système.	Approche transparence interactionnelle
	Contrôle utilisateur	Offrir la possibilité d'accepter, refuser ou ajuster les suggestions de l'IA et permettre une personnalisation des paramètres.	Maintenir l'autonomie de l'utilisateur face aux décisions algorithmiques.	Approche transparence interactionnelle
<u>Explicabilité des résultats</u> <i>(Permettre de comprendre comment et pourquoi un système a produit un résultat donné)</i>	Documentation technique	Mécanismes d'audit et d'explicabilité.	Avoir des outils dédiés à rendre les décisions compréhensibles.	Approche réglementaire
	Documentation technique	Description des algorithmes, sources de données, méthodes de traitement (ex. : anonymisation, réduction des biais).	Fournir les éléments techniques nécessaires à l'explication des résultats.	Approche réglementaire
	Responsabilité sociale	Analyse des effets négatifs (biais, perte d'autonomie, surveillance).	Permettre de comprendre les impacts potentiels des résultats.	Approche éthique by design
	Méthodologie éthique intégrée	Identification explicite des valeurs humaines fondamentales.	Aider à interpréter les résultats à la lumière des valeurs intégrées.	Approche éthique by design
	Méthodologie éthique intégrée	Co-construction de solutions techniques respectueuses de ces valeurs.	Renforcer la lisibilité des choix techniques.	Approche éthique by design
	Méthodologie éthique intégrée	Acceptation de compromis fonctionnels pour respecter l'éthique.	Expliquer les arbitrages faits dans les résultats.	Approche éthique by design
	Compréhension utilisateur	Expliquer les décisions prises par l'IA de manière intelligible et rendre visibles les étapes du raisonnement algorithmique.	Favoriser une appropriation critique des résultats par l'utilisateur.	Approche transparence interactionnelle
	Explication contextualisée	Adapter les explications au profil et au contexte d'usage de l'utilisateur et utiliser un langage accessible.	Rendre les résultats compréhensibles pour tous les types d'utilisateurs.	Approche transparence interactionnelle
	Explicabilité graduée	Proposer plusieurs niveaux de détail dans les explications et permettre à l'utilisateur de choisir la profondeur d'analyse.	Offrir une transparence adaptée aux besoins et compétences de chacun.	Approche transparence interactionnelle

	Transparence située	Contextualiser les informations selon l'usage et le profil utilisateur et aligner les explications avec les objectifs pédagogiques.	Eviter une transparence générique et favoriser une compréhension ciblée.	Approche transparence interactionnelle
<u>Documentation technique et réglementaire</u> <i>(Assurer la conformité aux normes et fournir une base technique complète)</i>	Inscription au registre public européen	Informations obligatoires : nom du système, fonction, entreprise responsable, résultats des tests de conformité, documentation technique, secteur d'application.	Regrouper un ensemble d'éléments réglementaires et technique requis.	Approche réglementaire
	Documentation technique	Preuves des conformité aux normes de sécurité et d'éthique.	Avoir les éléments nécessaire pour répondre aux exigences réglementaires.	Approche réglementaire
	Gouvernance éthique	Elicitation des valeurs et traduction en exigences techniques.	Faire le lien entre valeurs et documentation technique.	Approche éthique by design
	Gouvernance éthique	Evaluation continue via audits et retours d'expérience.	Faire partie des processus de documentation et d'amélioration continue.	Approche éthique by design
	Responsabilité sociales	Anticipation des médiations (morales, comportementales, cognitives)	Avoir le nécessaire pour avoir une documentation sur les impacts prévus.	Approche éthique by design
	Responsabilité sociales	Intégration de médiations souhaitables (durabilité, transparence, surveillance)	Refléter ces choix de conception de manière documentés.	Approche éthique by design
	Documentation technique intégrée	Fournir une documentation conforme aux normes (ex. : ISO, RGPD) et intégrer les politiques de confidentialité et les conditions d'usage.	Garantir la conformité réglementaire et la transparence technique.	Approche transparence interactionnelle

Confrontation de la grille ALTAIL modifié avec notre grille d'analyse

Nous avons adopté un code visuel afin de visualiser lorsqu'il y a une concordance entre les critères :  critères similaire.

De plus nous mettrons sur un fond vert clair les critères de transparence (traçabilité, explicabilité, communication) proposé par l'outil ALTAI.

Critères ALTAIL (24 critères de transparence/129 critères)	Critères		
	Approche réglementaire	Approche éthique by design	Approche interactionnelle
Analyse d'impact sur les droits fondamentaux			
Documentation des arbitrages éthiques			
Interaction avec la prise de décision humaine			
Préservation de l'autonomie décisionnelle			
Signalement des décisions algorithmiques			
Identification des agents virtuels			
Répartition des tâches entre IA et humains			
Renforcement des capacités humaines			
Prévention de la dépendance excessive			
Définition du niveau de contrôle humain			
Identification de l'acteur humain responsable			
Mécanismes de supervision humaine			
Auditabilité et gouvernance de l'autonomie			
Contrôle renforcé pour les systèmes autonomes ou auto-apprenants			
Détection des anomalies et réponse aux risques			
Présence d'un mécanisme d'arrêt d'urgence			
Évaluation des vulnérabilités potentielles			
Identification des types de vulnérabilités			
Mesures de résilience et d'intégrité			
Comportement en environnement imprévu			
Gestion du double usage			
Existence de solutions de secours			
Évaluation du niveau de risque contextuel			


Processus d'évaluation des risques et de la sécurité			
Communication des risques pour l'intégrité physique			
Politique d'assurance			
Plan de gestion des usages abusifs ou détournés			
Évaluation des dommages potentiels			
Règles de responsabilité et protection des consommateurs			
Évaluation des impacts environnementaux et sur les animaux			
Analyse des menaces de cybersécurité			
Évaluation des défaillances critiques			
Définition de seuils de déclenchement			
Test des solutions de secours			
Définition du niveau de précision attendu			
Méthodes de mesure de la précision			
Qualité et actualisation des données			
Évaluation des besoins en données supplémentaires			
Évaluation des préjudices liés aux erreurs			
Suivi du taux d'erreurs			
Plan de correction des erreurs			
Stratégie de contrôle du système			
Prise en compte des contextes spécifiques			
Méthodes de vérification de la fiabilité et de la reproductibilité			
Documentation des réglages critiques			
Application des processus de test			
Communication de la fiabilité aux utilisateurs			
Mécanisme de signalement des atteintes à la vie privée			
Évaluation du type et de la portée des données			
Minimisation de l'usage de données sensibles			
Mécanismes de consentement et de contrôle			
Mesures de protection des données			
Mobilisation du DPO (Data Protection Officer)			

Alignement sur les normes reconnues			
Mécanismes de contrôle des données			
Évaluation de la qualité des sources externes			
Garantie de la qualité et de l'intégrité des données			
Vérification de la sécurité des ensembles de données			
Documentation des protocoles de gouvernance des données			
Analyse des accès aux données			
Qualification et légitimité des personnes ayant accès aux données			
Traçabilité des accès aux données			
Mise en place de mesures de traçabilité			
Documentation des méthodes de conception			
Documentation de l'entraînement des modèles (IA fondée sur l'apprentissage)			
Documentation des méthodes de test et de validation			
Documentation des résultats et des décisions			
Compréhensibilité des décisions du système			
Explication des choix algorithmiques			
Évaluation de l'influence sur les processus décisionnels			
Justification du déploiement dans un domaine spécifique			
Évaluation du modèle économique			
Conception orientée vers l'interprétabilité			
Choix de modèles interprétables			
Analyse des données d'entraînement et de test			
Accès à la logique interne du modèle			
Identification claire du système comme étant une IA			
Explication des critères de décision			
Accessibilité de l'information pour tous les profils d'utilisateurs			
Prise en compte des retours utilisateurs			
Communication des risques et biais			
Communication élargie selon les publics			

Indication de la finalité du système			
Définition et explication des scénarios d'usage			
Prise en compte des limites cognitives humaines			
Communication différenciée selon les rôles			
Stratégie de prévention des biais			
Évaluation des limites des ensembles de données			
Prise en compte de la diversité des utilisateurs			
Utilisation d'outils d'analyse des biais			
Processus de détection et de contrôle des biais			
Mécanisme de signalement des biais			
Communication sur les biais et les recours			
Prise en compte des effets indirects			
Évaluation de la variabilité des décisions			
Analyse des causes de variabilité			
Évaluation de l'impact de la variabilité sur les droits fondamentaux			
Définition de l'équité appliquée			
Mesure de l'équité			
Mécanismes de garantie de l'équité			
Adaptation aux préférences et capacités individuelles			
Prise en compte des besoins spécifiques et du handicap			
Compatibilité avec les technologies d'assistance			
Consultation des communautés concernées			
Évaluation de l'impact sur les groupes d'utilisateurs			
Diversité de l'équipe de conception			
Évaluation des effets disproportionnés			
Prise en compte des retours de groupes externes			
Participation des parties prenantes			

Mobilisation des acteurs internes			
Mesure de l'impact environnemental			
Réduction de l'empreinte écologique			
Évaluation de l'attachement émotionnel			
Clarté sur la nature simulée de l'interaction			
Analyse des incidences sociales			
Évaluation de l'impact sociétal élargi			
Auditabilité du système			
Analyse des risques multi-acteurs			
Mise en place de formations à la responsabilité			
Identification des publics concernés par la formation			
Intégration du cadre juridique dans les formations			
Mise en place d'un comité d'éthique ou d'un mécanisme équivalent			
Recours à des audits ou orientations externes			
Mécanisme de signalement des vulnérabilités			
Mécanisme de recensement des valeurs et intérêts			
Processus de décision pour les arbitrages			
Documentation des arbitrages			
Mécanismes de recours en cas de préjudice			
Information sur les voies de recours			

Confrontation de la grille ALTAIL modifié avec les critères de transparence des entreprises

Nous avons adopté un code visuel afin de visualiser lorsqu'il y a une concordance entre les critères :  critères similaire.

De plus nous mettrons sur un fond vert clair les critères de transparence (traçabilité, explicabilité, communication) proposé par l'outil ALTAI.

Critères ALTAIL (24 critères de transparence/129 critères)	Critères						
	Entreprises Salesforce	Entreprise Grammarly	Entreprise Turnitin	Copilot (lien)	Gemini (lien 1) (lien 2)	Claude AI (Anthropic) (lien)	ChatGPT
Analyse d'impact sur les droits fondamentaux							
Documentation des arbitrages éthiques							
Interaction avec la prise de décision humaine							
Préservation de l'autonomie décisionnelle							
Signalement des décisions algorithmiques							
Identification des agents virtuels							
Répartition des tâches entre IA et humains							
Renforcement des capacités humaines							
Prévention de la dépendance excessive							
Définition du niveau de contrôle humain							
Identification de l'acteur humain responsable							
Mécanismes de supervision humaine							
Auditabilité et gouvernance de l'autonomie							
Contrôle renforcé pour les systèmes autonomes ou auto-apprenants							
Détection des anomalies et réponse aux risques							
Présence d'un mécanisme d'arrêt d'urgence							
Évaluation des vulnérabilités potentielles							
Identification des types de vulnérabilités							
Mesures de résilience et d'intégrité							
Comportement en environnement imprévu							
Gestion du double usage							
Existence de solutions de secours							

Évaluation du niveau de risque contextuel							
Processus d'évaluation des risques et de la sécurité							
Communication des risques pour l'intégrité physique							
Politique d'assurance							
Plan de gestion des usages abusifs ou détournés							
Évaluation des dommages potentiels							
Règles de responsabilité et protection des consommateurs							
Évaluation des impacts environnementaux et sur les animaux							
Analyse des menaces de cybersécurité							
Évaluation des défaillances critiques							
Définition de seuils de déclenchement							
Test des solutions de secours							
Définition du niveau de précision attendu							
Méthodes de mesure de la précision							
Qualité et actualisation des données							
Évaluation des besoins en données supplémentaires							
Évaluation des préjudices liés aux erreurs							
Suivi du taux d'erreurs							
Plan de correction des erreurs							
Stratégie de contrôle du système							
Prise en compte des contextes spécifiques							
Méthodes de vérification de la fiabilité et de la reproductibilité							
Documentation des réglages critiques							
Application des processus de test							
Communication de la fiabilité aux utilisateurs							
Mécanisme de signalement des atteintes à la vie privée							
Évaluation du type et de la portée des données							
Minimisation de l'usage de données sensibles							

Mécanismes de consentement et de contrôle							
Mesures de protection des données							
Mobilisation du DPO (Data Protection Officer)							
Alignement sur les normes reconnues							
Mécanismes de contrôle des données							
Évaluation de la qualité des sources externes							
Garantie de la qualité et de l'intégrité des données							
Vérification de la sécurité des ensembles de données							
Documentation des protocoles de gouvernance des données							
Analyse des accès aux données							
Qualification et légitimité des personnes ayant accès aux données							
Traçabilité des accès aux données							
Mise en place de mesures de traçabilité							
Documentation des méthodes de conception							
Documentation de l'entraînement des modèles (IA fondée sur l'apprentissage)							
Documentation des méthodes de test et de validation							
Documentation des résultats et des décisions							
Compréhensibilité des décisions du système							
Explication des choix algorithmiques							
Évaluation de l'influence sur les processus décisionnels							
Justification du déploiement dans un domaine spécifique							
Évaluation du modèle économique							
Conception orientée vers l'interprétabilité							
Choix de modèles interprétables							
Analyse des données d'entraînement et de test							
Accès à la logique interne du modèle							
Identification claire du système comme étant une IA							
Explication des critères de décision							
Accessibilité de l'information pour tous les profils d'utilisateurs							
Prise en compte des retours utilisateurs							

Communication des risques et biais							
Communication élargie selon les publics							
Indication de la finalité du système							
Définition et explication des scénarios d'usage							
Prise en compte des limites cognitives humaines							
Communication différenciée selon les rôles							
Stratégie de prévention des biais							
Évaluation des limites des ensembles de données							
Prise en compte de la diversité des utilisateurs							
Utilisation d'outils d'analyse des biais							
Processus de détection et de contrôle des biais							
Mécanisme de signalement des biais							
Communication sur les biais et les recours							
Prise en compte des effets indirects							
Évaluation de la variabilité des décisions							
Analyse des causes de variabilité							
Évaluation de l'impact de la variabilité sur les droits fondamentaux							
Définition de l'équité appliquée							
Mesure de l'équité							
Mécanismes de garantie de l'équité							
Adaptation aux préférences et capacités individuelles							
Prise en compte des besoins spécifiques et du handicap							
Compatibilité avec les technologies d'assistance							
Consultation des communautés concernées							
Évaluation de l'impact sur les groupes d'utilisateurs							
Diversité de l'équipe de conception							

Évaluation des effets disproportionnés							
Prise en compte des retours de groupes externes							
Participation des parties prenantes							
Mobilisation des acteurs internes							
Mesure de l'impact environnemental							
Réduction de l'empreinte écologique							
Évaluation de l'attachement émotionnel							
Clarté sur la nature simulée de l'interaction							
Analyse des incidences sociales							
Évaluation de l'impact sociétal élargi							
Auditabilité du système							
Analyse des risques multi-acteurs							
Mise en place de formations à la responsabilité							
Identification des publics concernés par la formation							
Intégration du cadre juridique dans les formations							
Mise en place d'un comité d'éthique ou d'un mécanisme équivalent							
Recours à des audits ou orientations externes							
Mécanisme de signalement des vulnérabilités							
Mécanisme de recensement des valeurs et intérêts							
Processus de décision pour les arbitrages							
Documentation des arbitrages							
Mécanismes de recours en cas de préjudice							
Information sur les voies de recours							

