

TABLEAU DE BORD

TRANSPARENCE DES SIA/RECOMMANDATIONS

Niveau d'automatisation du système

Fortement automatisé : systèmes basés sur des algorithmes d'apprentissage automatique

Amélioration de la transparence en concevant des systèmes hybrides (IA basée sur les règles et les données).

Plus un SIA est automatisé, plus il est crucial **d'évaluer si les utilisateurs comprennent les décisions** prises (*recommandation : garantir l'explicabilité des décisions*)

Plus un système est autonome, plus il doit intégrer **des mécanismes de supervision humaine et d'ajustement éthique** (*recommandation : adopter une architecture hybride*).

Plus les résultats générés sont variés, plus il est nécessaire de les **rendre compréhensibles pour éviter les interprétations** erronées (*recommandation : rendre la documentation accessible à tous les profils d'utilisateurs*).

Plus le niveau de risque est élevé, plus les **exigences de transparence doivent être renforcées** (*recommandation : identifier clairement les systèmes à haut risque*).

Plus le système interagit avec l'environnement de l'utilisateur, plus il influence ses comportements et décisions, ce qui impose une **responsabilité partagée** (*recommandation : informer clairement sur l'interaction homme-machine*).

Plus les décisions sont critiques, plus leur traçabilité doit être assurée par une **documentation technique rigoureuse et une journalisation sécurisée** (*recommandation : renforcer la traçabilité technique et documentaire*).

Degré d'autonomisation du système

Fortement autonome : fonctionnement sans supervision humaine constante

Augmenter la transparence en rendant nécessaire la validation humaine à chaque étape.

Plus un système d'IA est autonome, plus il doit intégrer des mécanismes d'ajustement fondés sur les **retours d'expérience pour garantir son acceptabilité sociale** (*recommandation : mettre en place des mécanismes d'ajustement basés sur les retours d'expérience*).

Plus le système repose sur des données ou des règles, plus sa **documentation doit être adaptée à sa nature** pour assurer une traçabilité complète (*recommandation : documenter les méthodes de conception et d'entraînement de manière contextualisée*).

Plus les décisions du système ont un impact, plus il est essentiel de **conserver leur historique et de les rendre consultables** en cas de doute ou d'audit (*recommandation : prévoir un plan d'intervention en cas de défaillance ou d'erreur*).

Plus le système est utilisé dans des contextes sensibles, plus **l'interface doit permettre une intervention humaine** claire et active (*recommandation : interface homme-machine avec supervision humaine*).

Capacité d'auto-apprentissage

Fortement capable d'auto-apprentissage : système évolutif après déploiement

Augmenter la transparence en expliquant les capacités d'adaptation dans une documentation.

Plus un système devient performant par auto-apprentissage, plus il doit **intégrer des mécanismes de supervision humaine** pour éviter une perte de contrôle éthique (*recommandation : intégrer des mécanismes de supervision humaine graduée*).

Plus le système évolue avec l'usage, plus il doit **ajuster en fonction des retours d'expériences** pour garantir sa pertinence et sa conformité aux valeurs humaines (*recommandation : garantir la qualité et la traçabilité des données d'entraînement*).

Plus le système apprend à partir des données, plus ses **méthodes de conception et d'entraînement doivent être documentées** de manière contextualisée et transparente (*recommandation : documenter les évolutions du système liées à l'apprentissage et l'adaptation*).

Plus les décisions du système sont influencées par des processus adaptatifs, plus leur traçabilité doit être assurée par une **journalisation sécurisée et accessible** (*recommandation : assurer la traçabilité des décisions et des résultats du système*).

Objectifs explicites et implicites

Capacité du système à intégrer des objectifs à partir d'information contextuelles

Augmenter la transparence en ayant une documentation technique intégrant les objectifs pédagogiques.

Plus un système est utilisé dans un contexte éducatif, plus il doit privilégier une explicabilité forte pour permettre aux utilisateurs de **comprendre les résultats et les impacts pédagogiques** (*recommandation : privilégier une explicabilité forte plutôt qu'une précision maximale*).

Plus les objectifs du système sont multiples, plus il est essentiel de les **cartographier pour éviter les confusions et aligner les intentions des concepteurs** avec les usages réels (*recommandation : cartographier les différents niveaux d'objectifs du système*).

Plus le système est complexe, plus il doit être conçu dès le départ avec **l'interprétation à l'esprit** pour garantir la transparence des logiques internes (*recommandation : concevoir le système avec l'interprétation à l'esprit dès la phase de développement*).

Plus les profils d'utilisateurs sont variés, plus **la notice d'information** doit être adaptée pour assurer une compréhension claire et accessible à tous (*recommandation : adapter la notice d'information aux profils des utilisateurs*).

TABLEAU DE BORD

TRANSPARENCE DES SIA/RECOMMANDATIONS

Capacité d'inférence

Bonne capacité d'inférence : système générant des recommandations ou prédictions à partir de données d'entrée.

Améliorer la transparence en n'ayant pas seulement que des résultats statistiques ou préprogrammées.

Typologie des résultats générés

Génération de prédictions, recommandations pédagogiques personnalisées, création de contenu, etc.

Améliorer la transparence en traçant et en communiquant sur l'influence sur l'environnements physiques ou virtuels auprès des utilisateurs.

Interaction avec l'environnement

Influence sur les environnements physiques et virtuels

Amélioration de la transparence en permettant qu'il n'y ait pas d'impact négatif direct sur les environnements.

Degré de risque du système

Systèmes pouvant être différents niveaux de risques : haut risque, risque limité, risque inacceptable.

Améliorer la transparence en respectant la réglementation de l'AI Act selon le niveau de risque du système d'IA. Pour les systèmes à haut risque, voir ci-dessous.

Plus un système repose sur des inférences, plus il doit **rendre explicite le processus de transformation des données en résultats** pour éviter l'opacité algorithmique (*recommandation : rendre les inférences compréhensibles et justifiables*).

Plus les décisions du système sont critiques, plus leur **enregistrement doit être automatisé, sécurisé et conservé** pour garantir la traçabilité et faciliter les audits (*recommandation : mettre en place une journalisation complète et sécurisée*).

Plus le système est susceptible de produire des biais ou des erreurs, plus il doit **informer les utilisateurs sur les risques** associés et les bonnes pratiques d'utilisation (*recommandation : informer les utilisateurs sur les risques et les biais potentiels*).

Plus le système est utilisé dans des environnements variés, plus il doit **intégrer les retours des utilisateurs** pour adapter ses modèles et améliorer sa transparence (*recommandation : intégrer les retours des utilisateurs dans l'évolution du système*).

Plus un système génère des résultats variés, plus il doit **sensibiliser les utilisateurs à leur diversité** pour éviter les interprétations erronées (*recommandation : sensibiliser les utilisateurs à la diversité des résultats générés*).

Plus les profils d'utilisateurs sont hétérogènes, plus la **documentation doit être adaptée et contextualisée** pour garantir une compréhension équitable (*recommandation : adapter la documentation aux différents profils d'utilisateurs*).

Plus les résultats du système sont complexes ou ambigus, plus **l'interface doit être conçue en tenant compte des biais cognitifs et des limites humaines** (*recommandation : réduire les biais cognitifs et les risques d'interprétation erronée*).

Plus un système interagit avec l'environnement numérique de l'utilisateur, plus il doit **anticiper les effets cognitifs, comportementaux et moraux** qu'il peut induire (*recommandation : anticiper les effets cognitifs, comportementaux et moraux des systèmes d'IA*).

Plus le système influence les usages, plus il doit **intégrer des médiations** souhaitables dans sa conception pour renforcer la responsabilité et la transparence (*recommandation : intégrer des médiations souhaitables dans la conception du système*).

Plus les enjeux éthiques sont élevés, plus il est essentiel de **mobiliser des experts en valeurs** dans les équipes de développement pour garantir une conception alignée avec les principes sociétaux (*recommandation : mobiliser des experts en valeurs dans les équipes de développement*).

Plus un système présente un potentiel de nuisance élevé, plus il doit être **évalué et déclaré comme "à haut risque"** avant tout déploiement (*recommandation : évaluer et déclarer le niveau de risque du système*).

Plus le système est critique, plus sa conception, son développement et sa validation doivent être **documentés de manière rigoureuse et accessible** (*recommandation : mettre en œuvre une documentation technique complète et accessible*).

Plus les décisions du système sont sensibles, plus leur **enregistrement doit être sécurisé, traçable et proportionné à l'usage** pour garantir la responsabilité (*recommandation : assurer la journalisation des événements critiques*).

Plus le niveau de risque est élevé, plus **l'interface doit permettre une supervision humaine renforcée** pour garantir que l'utilisateur reste l'arbitre final (*recommandation : garantir une interface homme-machine avec supervision humaine*).

Plus le pouvoir de nuisance du système est important, plus la **transparence doit être renforcée** au-delà des obligations contractuelles, en incluant les motivations commerciales et les conflits d'intérêts (*recommandation : adapter la transparence au pouvoir de nuisance du système*).