

Transparence et automatisation des systèmes

Nom de la recommandation	Détails
Renforcer la traçabilité technique et documentaire	<ul style="list-style-type: none"> • Créer des cartes de modèles pour chaque système d'IA, incluant : nom, fonction, entreprise responsable, secteur d'application, résultats des tests, documentation technique. • Documenter les méthodes de conception, de test et de validation dans des onglets accessibles (ex. : indicateurs, données d'entraînement, analyses quantitatives). • Mettre en œuvre une journalisation automatisée des événements significatifs (décisions critiques, anomalies, interventions humaines).
Identifier clairement les systèmes à haut risque	<ul style="list-style-type: none"> • Mentionner explicitement dans la documentation et l'interface si le système est classé comme "à haut risque". • Enregistrer le système dans un registre public (Art. 51, AI Act), par exemple via des normes ISO.
Garantir l'explicabilité des décisions	<ul style="list-style-type: none"> • Intégrer des interfaces explicatives dynamiques permettant aux utilisateurs de comprendre les décisions prises. • Poser des questions clés dans l'interface, comme : « Avez-vous évalué la mesure dans laquelle la décision du système influence les processus décisionnels ? » • Documenter les objectifs explicites et implicites du système, en lien avec les finalités pédagogiques.
Informier clairement sur l'interaction homme-machine	<ul style="list-style-type: none"> • Afficher un message d'accueil clair dans les interfaces conversationnelles : « Je suis [nom du système], une IA conçue pour [objectif]. » • Permettre une supervision humaine et une possibilité de recours à un humain si l'IA ne répond pas aux attentes. • Éviter toute confusion avec un humain en intégrant des icônes, labels ou clauses de non-responsabilité visibles.
Adopter une architecture hybride	<ul style="list-style-type: none"> • Combiner règles explicites et apprentissage automatique pour faciliter la traçabilité et le contrôle humain. • Définir des seuils de confiance au-delà desquels une intervention humaine est requise. • Utiliser les règles comme mécanismes de validation des prédictions basées sur les données.
Rendre la documentation accessible à tous les profils d'utilisateurs	<ul style="list-style-type: none"> • Adapter le vocabulaire et les formats (textes, graphiques dynamiques, scénarios d'usage) selon les publics (étudiants, enseignants, administrateurs). • Organiser des audits réguliers pour mettre à jour la documentation en fonction des usages réels et des retours d'expérience.

Transparence et autonomisation des systèmes

Nom de la recommandation	Détails
Mettre en place des mécanismes d'ajustement basés sur les retours d'expérience	<p>Développer un système de retour d'expérience intégré pour ajuster le fonctionnement du système d'IA en fonction de son impact réel :</p> <ul style="list-style-type: none"> • Intégrer des rapports interprétables afin d'affiner les suggestions de l'IA afin d'affiner les recommandations générées par le système, en le rendant plus compréhensibles pour l'utilisateur ; interpréter les suggestions dans des rapports explicites, facilitant ainsi une appropriation critique des résultats; coconstruire des solutions techniques respectueuses des valeurs humaines même si cela implique de renoncer à certaines fonctionnalités ou d'adapter les objectifs initiaux du projet. • Organiser des ateliers participatifs avec les parties prenantes pour identifier les impacts sociaux et techniques. • Mettre en œuvre des audits réguliers et des protocoles d'évaluation éthique.
Documenter les méthodes de conception et d'entraînement de manière contextualisée	<p>Fournir une documentation technique complète et accessible, adaptée au type de système :</p> <ul style="list-style-type: none"> • Pour les IA basées sur les données : décrire les données d'entraînement, les conditions de collecte, les algorithmes utilisés, et les stratégies de test. • Pour les IA basées sur les règles : expliciter les règles heuristiques, le moteur d'inférence, et les logiques de décisions. • Adapter la documentation aux types d'utilisateurs et aux objectifs pédagogiques.
Prévoir un plan d'intervention en cas de défaillance ou d'erreur	<p>Inclure dans la documentation technique :</p> <ul style="list-style-type: none"> • Une stratégie pour éviter les biais, garantir l'équité et protéger les droits fondamentaux. • Un plan d'intervention clair en cas de dysfonctionnement du système ou de dérive algorithmique. • Des protocoles d'audit réguliers pour vérifier la conformité et ajuster le système en fonction de son impact réel.
Interface homme-machine avec supervision humaine	<p>Permettre aux utilisateurs de comprendre lorsqu'ils utilisent un système doté d'intelligence artificielle :</p> <ul style="list-style-type: none"> • Répondre à la question : « Avez-vous informé les utilisateurs finaux (au moins par une clause de non-responsabilité ou tout autre moyen) qu'ils interagissent avec un système d'IA et non avec un être humain ? » • Les utilisateurs doivent être informés ou avertis clairement qu'ils interagissent avec une IA. Par exemple, un chatbot doit indiquer explicitement qu'il n'est pas humain, afin d'éviter toute confusion.

Transparence et capacité d'auto-apprentissage des systèmes

Nom de la recommandation	Détails
Intégrer des mécanismes de supervision humaine graduée	<p>Concevoir les interfaces homme-machine de manière à permettre une supervision humaine proportionnelle au niveau d'autonomie du système :</p> <ul style="list-style-type: none"> • Ajouter des boutons d'arrêt, de correction ou de désactivation des fonctionnalités automatisées. • Définir des seuils de confiance au-delà desquels une intervention humaine est requise. • Permettre à l'utilisateur de visualiser et modifier les décisions prises par le système.
Garantir la qualité et la traçabilité des données d'entraînement	<p>Dès la phase de conception, collecter et étiqueter les données de manière rigoureuse :</p> <ul style="list-style-type: none"> • Fournir un document certifiant la qualité des données de formation. • Décrire les algorithmes utilisés pour aboutir à une décision. • Assurer la traçabilité complète du processus de texte, avec des preuves de conformité aux normes de sécurité et d'éthique.
Documenter les évolutions du système liées à l'apprentissage et l'adaptation	<p>Mettre en place une documentation dynamique qui décrit non seulement ce que le système fait, mais aussi comment il évolue :</p> <ul style="list-style-type: none"> • Expliquer les zones où l'auto-apprentissage est autorisé ou interdit. • Décrire les arbitrages entre performance et éthique, notamment les choix de conception qui ont un impact sur la transparence. • Intégrer des outils comme les matrices multicritères, les scénarios d'usage ou simulations éthiques pour anticiper les tensions et les compromis.
Assurer la traçabilité des décisions et des résultats du système	<p>Mettre en œuvre une journalisation automatisée et sécurisée des décisions critiques :</p> <ul style="list-style-type: none"> • Conserver un historique des décisions accessible en cas de suspicion de non-conformité. • Rendre cette documentation consultable par les autorités et les utilisateurs concernés. • Intégrer des mécanismes de signalement des anomalies et des interventions humaines pour garantir un contrôle éthique.

Transparence et objectifs

Nom de la recommandation	Détails
Privilégier une explicabilité forte plutôt qu'une précision maximale	<p>Concevoir le système en mettant l'accent sur la compréhension des résultats par les utilisateurs :</p> <ul style="list-style-type: none"> • Documenter les finalités réelles du système, y compris les impacts sociaux et pédagogiques. • Répondre à la question : « Pourquoi ce système particulier a-t-il été déployé dans ce domaine spécifique ? ». • Intégrer des interfaces explicatives qui justifient les recommandations avec des données et des raisonnements clairs.
Cartographier les différents niveaux d'objectifs du système	<p>Identifier et rendre visibles les objectifs :</p> <ul style="list-style-type: none"> • Objectifs système (ce que l'entreprise dit que le système fait). • Objectifs dans le code (ce que le système fait réellement). • Objectifs perçus par les utilisateurs (ce qu'ils pensent que le système fait). • Objectifs implicites des concepteurs (ce qu'ils veulent vraiment que le système fasse). <p>Cette cartographie permet de réduire l'opacité éthique et d'aligner les intentions des concepteurs avec les usages réels.</p>
Concevoir le système avec l'interprétation à l'esprit dès la phase de développement	<p>Dès la conception, intégrer des mécanismes permettant de justifier les choix algorithmiques :</p> <ul style="list-style-type: none"> • Répondre à la question : « avez-vous conçu le système d'IA en ayant dès le départ l'interprétation à l'esprit ? ». • S'assurer que l'objectif système est explicable par l'objectif dans le code. • Utiliser des structures de données internes (ex: CRM) pour produire des résultats interprétables.
Adapter la notice d'information aux profils des utilisateurs	<p>Rendre la documentation accessible, visible et compréhensible :</p> <ul style="list-style-type: none"> • Intégrer des graphismes dynamiques ou des descriptions textuelles des processus de décision. • Adapter le vocabulaire et les formats selon les publics (étudiants, enseignants, administrateurs). • Créer un pont informationnel entre les concepteurs et les utilisateurs pour garantir une transparence effective. Par exemple, en créant une carte de votre modèle accessible.

Transparence et capacité d'inférence

Nom de la recommandation	Détails
Rendre les inférences compréhensibles et justifiables	Documenter clairement le processus d'inférence pour éviter l'opacité algorithmique : <ul style="list-style-type: none">• Décrire comment les données observées sont transformées en résultats.• Fournir une explication claire et compréhensible des décisions prises par le système.• Ne pas se limiter aux résultats générés : expliquer les logiques internes et les chaînes de raisonnement.
Mettre en place une journalisation complète et sécurisée	Automatiser l'enregistrement des événements significatifs liés au fonctionnement du système : <ul style="list-style-type: none">• Journaliser les entrées et sorties, les décisions critiques, les anomalies et les interventions humaines.• Conserver ces données pour une durée proportionnée à l'usage, afin de faciliter les audits.• Garantir la sécurité et l'intégrité des enregistrements.
Informers les utilisateurs sur les risques et les biais potentiels	Communiquer de manière transparente sur les limites du modèle : <ul style="list-style-type: none">• Expliquer les risques associés aux inférences, notamment les biais ou les erreurs possibles.• Fournir des bonnes pratiques d'utilisation pour éviter les malentendus.• Intégrer une clause de transparence dans l'interface utilisateur.
Intégrer les retours des utilisateurs dans l'évolution du système	Mettre en place des processus d'adaptation basés sur les retours : <ul style="list-style-type: none">• Permettre aux utilisateurs de formuler des commentaires sur les décisions du système.• Utiliser ces retours pour ajuster les modèles et améliorer la précision et la transparence.• Organiser des ateliers d'inclusion pour identifier les impacts sociaux et techniques.

Transparence et typologie des résultats générés

Nom de la recommandation	Détails
Sensibiliser les utilisateurs à la diversité des résultats générés	<p>Informer les utilisateurs sur les types de résultats produits et les logiques sous-jacentes :</p> <ul style="list-style-type: none">• Expliquer clairement comment et pourquoi un résultat a été généré.• Mettre en place des interfaces explicatives qui accompagnent chaque sortie avec une justification compréhensible.• Intégrer des indicateurs de fiabilité ou de niveau de confiance pour aider à l'interprétation.
Adapter la documentation aux différents profils d'utilisateurs	<p>Rendre la documentation accessible, contextualisée et appropriée :</p> <ul style="list-style-type: none">• Répondre à la question : « Avez-vous veillé à ce qu'une explication de la raison pour laquelle un système a procédé à un certain choix puisse être rendue compréhensible pour l'ensemble des utilisateurs ? ».• Utiliser des moyens de communication variés : textes simplifiés, graphiques dynamiques, vidéos explicatives, etc.• Proposer des scénarios d'usage clairs pour chaque finalité du système.
Réduire les biais cognitifs et les risques d'interprétation erronée	<p>Concevoir les interfaces en tenant compte de la psychologie des utilisateurs :</p> <ul style="list-style-type: none">• Anticiper les biais de confirmation, la fatigue cognitive, et les risques de confusion.• Tester les interfaces avec des groupes utilisateurs diversifiés pour identifier les points de friction.• Intégrer des mécanismes de correction ou de clarification lorsque des résultats ambigus sont générés.

Transparence et interaction avec les environnements physiques et virtuels

Nom de la recommandation	Détails
Anticiper les effets cognitifs, comportementaux et moraux des systèmes d'IA	<p>Intégrer une analyse des médiations technologiques dès la phase de conception :</p> <ul style="list-style-type: none">• Identifier les effets indésirables que le système peut avoir sur les comportements, les perceptions ou les décisions éthiques des utilisateurs.• Evaluer les risques de manipulation ou d'influence morale liés à l'intégration du système dans l'environnement numérique.• Mettre en œuvre des mécanismes de régulation éthique pour encadrer ces effets.
Intégrer des médiations souhaitables dans la conception du système	<p>Favoriser une moralisation active du système :</p> <ul style="list-style-type: none">• Définir des principes éthiques explicites (exemples : responsabilité, inclusion, transparences).• Recueillir les retours d'expérience des parties prenantes pour ajuster les pratiques et les politiques.• Intégrer des fonctionnalités qui renforcent la responsabilité de l'utilisateur et du concepteur.
Mobiliser des experts en valeur dans les équipes de développement	<p>Constituer des comités consultatifs interdisciplinaires :</p> <ul style="list-style-type: none">• Inclure des experts en droits humains, en éthique appliquée et en sciences sociales dans les phases de conception et d'évaluation.• Adopter une approche de Value-Based Engineering pour aligner les choix techniques avec les valeurs sociétales.• Documenter les décisions de conception éthiques et les rendre accessibles aux utilisateurs et aux régulateurs.

Transparence et degrés de risque

Nom de la recommandation	Détails
Evaluer et déclarer le niveau de risque du système	<p>Avant tout déploiement, le concepteur doit :</p> <ul style="list-style-type: none">• Identifier si le système relève de la catégorie « à haut risque » selon les critères de l'AI Act (Art.6).• Informer le fournisseur d'IA de l'obligation d'enregistrement dans la base de données européenne (Art.51, AI Act).• Ne pas utiliser le système proposé par le fournisseur de modèle d'IA si aucune traçabilité publique n'est disponible.
Mettre en œuvre une documentation technique complète et accessible	<ul style="list-style-type: none">• Documenter chaque étape de conception, développement et validation du système.• Création d'une notice d'information détaillée sur le fonctionnement, les risques et les limites du système.• Assurer la traçabilité complète du processus de test, avec des preuves de conformité aux normes de sécurité et d'éthique.• Rendre cette documentation consultable par les autorités et les utilisateurs concernés.
Assurer la journalisation des événements critiques	<ul style="list-style-type: none">• Mettre en place une journalisation automatisée et sécurisée des décisions critiques, anomalies et interventions humaines.• Utiliser des cartes de modèles pour centraliser les informations.• Décrire comment les données observées sont transformées en résultats.• Conserver un historique des décisions accessible en cas de suspicion de non-conformité• Conserver les données pour une durée proportionnée à l'usage, facilitant les audits et les recours.
Garantir une interface homme-machine avec supervision humaine	<ul style="list-style-type: none">• Intégrer des mécanismes de contrôle renforcés : bouton d'arrêt, possibilité de correction, seuils de confiance.• Permettre à l'utilisateur de désactiver ou ajuster certaines fonctionnalités selon le niveau de risque.• L'interface de l'utilisateur doit informer clairement les utilisateurs qu'ils interagissent avec une IA.• Permettre une supervision humaine et une possibilité de recours à un humain si l'IA ne répond pas aux attentes.• Mentionner explicitement dans la documentation et l'interface si le système est classé comme « à haut risque ».• Assurer que l'utilisateur reste l'arbitre final, même dans un système hautement autonome.• Fournir une explication claire et compréhensible des décisions prises par le système.
Adapter la transparence au pouvoir de nuisance du système	<p>Plus le système est à risque, plus la transparence doit être renforcée.</p> <p>Ne pas se contenter de dispositifs contractuels privés, ils doivent être complétés par des engagements éthiques pérennes.</p> <p>Intégrer une transparence sur les motivations commerciales et les conflits d'intérêts.</p> <p>La transparence ne concerne pas seulement le déployeur de système d'IA mais également les fournisseurs de modèles d'IA ou les fournisseurs d'IA. Le concepteur doit obligatoirement communiquer avec les fournisseurs sur les modèles d'IA qui seront vendus ou loués.</p>