



GoodAlgo

Algorithm Data Ethic

Glossaire

Guide explicatif des évaluations





TABLE DES MATIERES

1.	Evaluation systématique	3
2.	Evaluation contextuelle.....	8
3.	Evaluation algorithmique	9





1. EVALUATION SYSTEMIQUE

Cette évaluation vous permet de vous positionner et d'identifier les zones de valorisation à améliorer, selon six angles différents :

Famille d'éthique

- L'Éthique de la donnée
- L'Éthique de l'Algorithme
- L'Éthique des systèmes
- L'Éthique des pratiques
- L'Éthique de l'environnement
- L'Éthique des décisions
- L'Éthique de la condition humaine

Étape d'un projet d'IA

- Idéation
- Contexte/objectifs
- Conception
- Méthodologie technique
- Analyse/Modélisation/Développement
- Recette/Résultat
- Déploiement – Maintenance

Principe et engagement éthique

- L'Autonomie
- La Bienfaisance
- L'Environnement durable
- La Garantie humaine
- L'Humanité digitale
- La Justice
- La Non-malfaisance
- La Qualité
- La Responsabilité
- La Transparence



Conformité aux exigences à l'AI ACT

- Les Codes de bonnes pratiques
- le Contrôle humain
- les Données d'entraînement et Gouvernance
- la Précision, Robustesse & Cybersécurité
- le Système de gestion des risques
- la Tenue du registre des activités
- la Transparence et information des utilisateurs

Risques

- L'Action humaine et Contrôle humain
- Le Bien-être sociétal et environnemental
- La Diversité, Non-discrimination et Équité
- La Non-malfaisance
- La Responsabilité
- La Robustesse technique et Sécurité
- La Transparence
- La Vie privée et la Gouvernance des données

Transparence systémique

- La Loyauté
- La Responsabilité
- L'Explicabilité
- La Garantie humaine
- L'Interprétabilité
- La Justice



Action humaine et Contrôle humain : L'idée que les actions et les décisions, en particulier dans le domaine technologique, devraient être guidées et contrôlées par des individus plutôt que par des systèmes automatisés.

Analyse / Modélisation / Développement : Collecte et traitement des données, conception des modèles d'IA et mise en œuvre des solutions techniques.

Autonomie : La capacité d'agir, de prendre des décisions et de diriger sa propre vie sans être excessivement influencé ou contrôlé par des facteurs externes.

Bien-être sociétal et environnemental : La promotion du bien-être collectif tant pour les individus que pour l'environnement, cherchant un équilibre entre progrès social et conservation de la nature.

Bienfaisance : L'action de faire le bien, d'agir de manière bénéfique envers autrui, généralement dans le but de promouvoir le bien-être.

Code de bonnes pratiques : Ensemble de recommandations volontaires visant à assurer le développement et l'utilisation responsables des systèmes d'IA, en complément des exigences réglementaires.

Conception : Élaboration des spécifications fonctionnelles et techniques, ainsi que des critères d'évaluation éthique et de performance du système.

Contexte / Objectifs : Définition du cadre du projet, des enjeux, des contraintes et des objectifs attendus en lien avec les parties prenantes.

Contrôle humain : Mécanismes garantissant l'intervention humaine pour superviser, corriger ou interrompre le fonctionnement du système afin de préserver l'autonomie et la sécurité des utilisateurs.

Déploiement / Maintenance : Mise en production du système et suivi continu pour garantir sa fiabilité, son amélioration et sa conformité dans le temps.

Diversité : La présence de différentes caractéristiques, perspectives, expériences et compétences au sein d'un groupe ou d'une communauté.

Données d'entraînement et Gouvernance : Encadrement de la collecte, de la gestion et de la qualité des données pour assurer la fiabilité, la représentativité et l'éthique du système d'IA.

Environnement durable : Un état dans lequel les ressources naturelles sont utilisées de manière à satisfaire les besoins actuels sans compromettre la capacité des générations futures à satisfaire leurs propres besoins.



Éthique de la condition humaine : Respect de la dignité et du bien-être des individus, en veillant à l'équité, à l'inclusion et à l'impact des systèmes sur leur perception de soi et leur développement.

Éthique de l'algorithme : Supervision, qualité d'usage et processus associés à l'exploitation des données numériques tout au long de leur cycle de vie.

Éthique de la donnée : Encadrement et qualification des données traitées par l'algorithme.

Éthique de l'environnement : Analyse des diverses interactions et impacts en fonction de l'écosystème ambiant via des paramètres contextuels.

Éthique des décisions : Garantie humaine et répercussions sur la responsabilité, le libre-arbitre et l'autonomie de la personne dans sa prise de décision.

Éthique des pratiques : Accompagnement et respect des finalités, des résultats et des usages du traitement de données numériques.

Éthique des systèmes : Ergonomie et fonctionnalité de l'architecture technologique, en adéquation avec les compétences et les pratiques de l'utilisateur.

Explicabilité : Cela se rapporte à la capacité d'expliquer le fonctionnement interne d'un modèle ou d'un système d'IA de manière compréhensible pour les utilisateurs. Une IA explicite fournit des explications sur la façon dont elle a pris une décision particulière, ce qui permet aux utilisateurs de comprendre les raisonnements derrière les résultats générés.

Garantie humaine : L'assurance que les technologies et les systèmes respectent et préservent les droits fondamentaux de l'homme, par de la supervision ou du contrôle humain.

Humanité digitale : La préservation des valeurs et des droits humains dans le contexte numérique, en particulier à mesure que la technologie devient de plus en plus intégrée dans la société.

Idéation : Phase initiale où les besoins et opportunités sont explorés pour identifier des cas d'usage pertinents de l'IA.

Interprétabilité : Il s'agit de la capacité à comprendre et à interpréter les résultats produits par un modèle ou un algorithme d'IA. Un modèle interprétable permet aux utilisateurs de saisir les relations entre les entrées et les sorties, ce qui contribue à la confiance dans le système.

Justice : L'équité, l'égalité et le respect des droits dans la distribution des avantages et des charges au sein d'une société.



GoodAlgo – GLOSSAIRE

Loyauté : Engagement à minimiser les impacts négatifs du système d'IA en identifiant, évaluant et atténuant les risques pouvant affecter les utilisateurs, la société ou l'environnement, tout en garantissant l'équité et la non-discrimination.

Méthodologie technique : Choix des approches algorithmiques, des sources de données et des outils technologiques adaptés au projet.

Non-discrimination et Équité : Le traitement juste et égal de toutes les personnes, sans discrimination, favorisant l'équité dans l'accès aux opportunités et aux ressources.

Non-malfaisance : Le principe éthique qui stipule qu'il est impératif de ne pas causer de tort intentionnellement, et de minimiser les dommages potentiels dans toute action entreprise.

Précision, robustesse et cybersécurité : Exigences visant à garantir des performances fiables, à minimiser les biais et à protéger le système contre les attaques ou dysfonctionnements.

Qualité : Le degré de perfection ou d'excellence dans un produit, un service ou un processus, conformément aux normes établies.

Recette / Résultat : Vérification de la conformité aux attentes, tests de performance et d'éthique, validation des résultats obtenus.

Responsabilité : L'obligation de répondre de ses actions et de prendre en charge les conséquences éventuelles, notamment en matière éthique.

Robustesse technique et Sécurité : La capacité d'un système à fonctionner de manière fiable et sécurisée dans différentes conditions, tout en protégeant les données et les utilisateurs contre les menaces potentielles.

Tenue du registre des activités : Obligation de documenter les actions, décisions et évolutions du système pour assurer traçabilité, auditabilité et conformité réglementaire.

Transparence : La clarté et l'ouverture dans la communication et les processus, permettant aux parties prenantes de comprendre les actions et les décisions. Les notions d'explicabilité et d'interprétabilité sont étroitement liées à la transparence dans le domaine de l'IA.

Transparence et information des utilisateurs : Communication claire sur les capacités, limites et implications du système d'IA afin de garantir une utilisation éclairée et responsable.

Vie privée et Gouvernance des données : La protection des informations personnelles des individus et la mise en place de règles et de pratiques pour assurer une utilisation éthique des données numériques.



2. EVALUATION CONTEXTUELLE

Cette évaluation vous permet de vous positionner et d'identifier les zones de valorisation à améliorer, selon quatre cinq différents :

Contexte

- Contexte et finalités
- Process du SIA
- Contexte de collecte
- Monitoring & Suivi
- Environnement durable

Contexte & Finalités : l'examen des objectifs et de l'utilisation prévue du système d'IA. Les finalités doivent être claires, éthiques et alignées avec les besoins des utilisateurs.

Process de l'IA : les étapes et les méthodes utilisées pour développer, tester et déployer l'algorithme. Le processus doit être bien structuré, rigoureux et transparent.

Contexte de collecte : dans quelles conditions les données ont été collectées. La collecte des données doit être réalisée de manière éthique, précise et représentative de la population cible.

Monitoring / suivi : la qualité et la fréquence des actions de suivi et de surveillance après le déploiement de l'IA. Le système doit être régulièrement surveillé, mis à jour et ajusté pour assurer son bon fonctionnement.

L'environnement durable : les pratiques et les initiatives mises en œuvre pour réduire l'impact écologique de l'IA. Ces actions doivent être responsables, mesurables et alignées avec des objectifs de durabilité.



3. EVALUATION ALGORITHMIQUE

Cette évaluation vous permet de vous positionner et d'identifier les zones de valorisation à améliorer, selon quatre angle différents :

Éthique de la donnée

- Biais historique
- Fiabilité
- Représentation

Éthique de l'algorithme

- Pertinence
- Sensibilité
- Robustesse
- Interprétabilité
- Discrimination

Suivi de l'algorithme

- Déploiement
- Robustesse dans le temps

Documentation

- Documentation générale
- Documentation de la donnée
- Documentation de l'algorithme

Sécurité du code et de la donnée

- Déploiement
- Accès
- Données
- Code



Éthique de la donnée

Biais d'agrégation : Analyse des effets introduits par le traitement ou la combinaison des données, susceptibles d'entraîner des biais ou des erreurs.

Biais de représentation : Évaluation de l'équilibre et de la diversité des données pour garantir une couverture équitable des groupes représentés.

Biais historique : Identification des éventuelles distorsions liées à l'origine des données pour limiter les biais liés au passé.

Fiabilité des données d'entrée : Analyse de la qualité et de la pertinence des données utilisées pour garantir une base solide et fiable.

Éthique de l'algorithme

Biais de mesure : Identification des biais introduits par les méthodes de calcul et de traitement dans les décisions algorithmiques.

Discrimination : Analyse des impacts potentiellement inéquitables du modèle sur différents groupes d'utilisateurs.

Interprétabilité : Vérification de la clarté des processus décisionnels, afin que les parties prenantes puissent comprendre et justifier les résultats.

Pertinence : Vérification de l'adéquation entre les objectifs du modèle et ses mécanismes décisionnels.

Sensibilité : Évaluation de la réponse du modèle aux variations des données d'entrée pour s'assurer qu'il reste stable dans des contextes variés.

Robustesse : Analyse de la capacité du modèle à maintenir ses performances face à des données bruitées ou à des situations imprévues.

Suivi de l'algorithme

Biais de déploiement : Étude des impacts que l'environnement de déploiement peut avoir sur les performances et l'équité du modèle.

Robustesse 'dans le temps' : Évaluation de la capacité de l'algorithme à rester performant et pertinent à mesure que les données ou les contextes évoluent.



Documentation

Documentation générale : Vérification de la clarté et de la complétude des informations relatives à l'algorithme et à son cycle de vie.

Documentation sur l'éthique de la donnée : Mise en avant des pratiques et considérations éthiques appliquées à la gestion des données.

Documentation sur l'éthique de l'algorithme : Explication des choix, des hypothèses et des limites liés au modèle.

Sécurité du code et de la donnée

Accès : Protection contre les usages non autorisés.

Code : Vérification de l'absence de vulnérabilités.

Déploiement : Garantie d'un environnement d'utilisation sécurisé.

Données : Préservation de la confidentialité et de l'intégrité des informations.

