

Ecolab – CVRH "L'IA dans le pole ministériel"

- Qu'est-ce qu'une démarche éthique?
- Quels sont les principes de référence en cas de doute sur le recours à l'IA?
- Quelle est la portée de la démarche éthique ?
- Ressources utiles pour aller plus loin

Qu'est-ce qu'une démarche éthique?

- Démarche volontaire
- Incertitude sur le bien-fondé d'une action (recours à l'IA)
- Questionnement de l'action par rapport à des valeurs que l'on s'impose de respecter
- Démarche casuistique : proportionner l'importance de certaines valeurs aux bénéfices et aux risques réels et potentiels de l'action à court et long terme

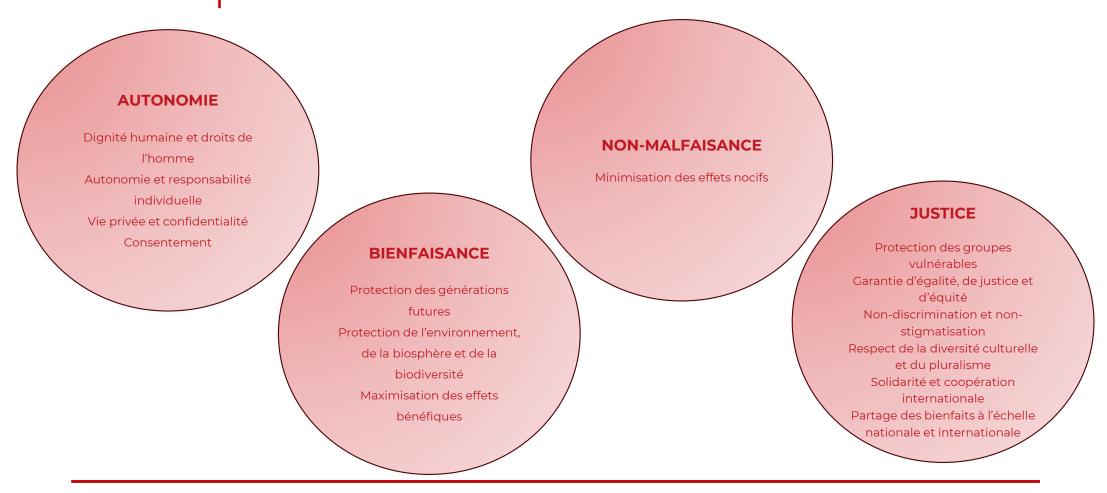


Il existe des « principes universels fondés sur des valeurs éthiques communes afin de guider le développement scientifique et technologique ainsi que les transformations sociales [...] en tenant compte de la responsabilité des générations présentes envers les générations futures. »

Déclaration universelle sur la bioéthique et les droits de l'Homme, Conférence Générale de l'UNESCO, Paris, 19 octobre 2005

Ekitia – 2023 4





Ekitia - 2023

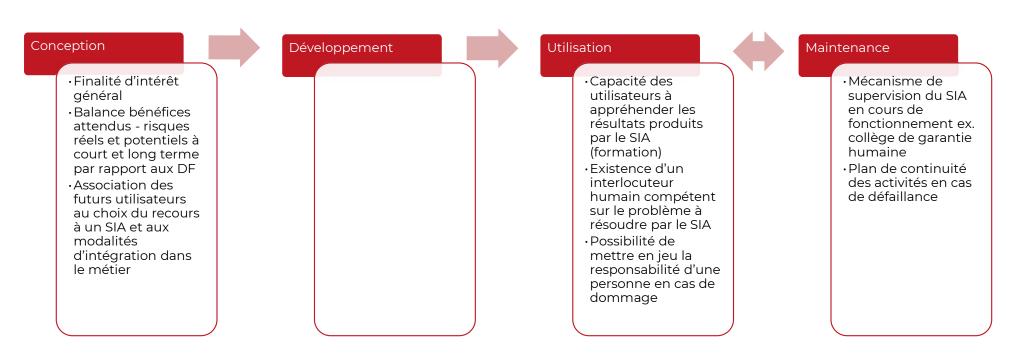
Emergence de principes éthiques spécifiques aux enjeux liés au recours à l'IA dans le cadre de la mise en œuvre de politiques publiques :

- Primauté humaine
- Performance
- Equité et non-discrimination
- Transparence
- Sûreté
- Soutenabilité environnementale
- Autonomie stratégique

Source: Etude à la demande du Premier Ministre « IA et action publique: construire la confiance, servir la performance, Conseil d'Etat, 31 mars 2022, accessible ici: https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance



Primauté humaine : « Un humain doit veiller à ce que les SIA fonctionnent à son bénéfice, se porter garant de leur bon fonctionnement et répondre des conséquences de ses dysfonctionnements. »



Ekitia – 2023 7



Performance: « Le principe de performance inclut la minimisation, la prévention et la résolution rapide des dysfonctionnements du système en conditions normales d'utilisation »

Conception

- Définition d'indicateurs de performance (exactitude. robustesse technique. temps de réponse, satisfaction des utilisateurs)
- Définition du niveau de performance acceptable au regard des conséquences de l'erreur

Développement

- · Evaluation de la performance par rapport aux indicateurs et au niveau de performance attendus
- · Oualité et pertinence des données d'apprentissage
- · Choix d'un algorithme approprié
- Prévention d'incidents résultant d'accès non malveillants ou de cas fortuits

Utilisation

- Condition d'amélioration de la qualité du service rendu au regard d'une prestation humaine ou une prestation numérique antérieure
- · Condition de satisfaction des utilisateurs et des destinataires du SIA

Maintenance

- Neutralisation des anomalies
- · Garantie de disponibilité et de stabilité de l'outil
- · Résilience vis-àvis d'incidents résultant d'accès non malveillants ou de cas fortuits



Equité et non-discrimination : « Un SIA ne saurait aboutir à un résultat inéquitable et, en particulier, à un résultat qui introduirait une discrimination injustifiée. »

Conception

- · Choix de l'acception de l'équité qui guidera le fonctionnement du SIA (équité collective, équité individuelle...)
- · Evaluation des risques de discrimination directe ou indirecte engendrés par le recours au SIA

Développement

- · Paramétrage du SIA en cohérence avec l'acceptation de l'équité choisie
- Analyse critique des données d'apprentissage et du modèle visant à prévenir les biais discriminatoires
- Représentativité sociale des équipes de développement

Utilisation

- · Amélioration de la qualité du service rendu
- Sensibilisation des agents utilisateurs de SIA aux problématiques liées aux biais discriminatoires
- Accessibilité du SIA vis-à-vis des usagers

Maintenance

 Mise en place d'un système de gestion des risques visant à prévenir la survenance de biais discriminatoires



Transparence: « Le principe de transparence impose la formalisation des choix et arbitrages effectués lors de la conception et des constats effectués après la mise en service. »

Conception

 Documentation des objectifs du recours à un SIA. des choix et arbitrages effectués lors de la conception ·Association des parties prenantes

à la conception

Développement

 Documentation des choix de conception du modèle algorithmique, des jeux de données utilisés et de l'ensemble des processus

Utilisation

- ·Capacité de l'utilisateur du SIA à comprendre et restituer les éléments-clés du « raisonnement » (exigence d'explicabilité) ·Droit d'accès du
- public à la documentation du SIA (art. L311-2 et L312-1-3 CRPA) ·Information des
- personnes de l'utilisation d'un SIA à leur égard (exigence de lovauté)

Maintenance

 Documentation des choix et arbitrages effectués après la mise en service



Sûreté : « La conception d'un SIA public de confiance doit impérativement intégrer [...] la prévention des attaques et la résolution de leurs conséquences. »

Conception

- ·Identification de vulnérabilités liées au SI dans lequel s'insère le SIA
- ·Identification des « points faibles » spécifiques au SIA (ex. empoisonnement des données d'apprentissage, leurre du SIA, vol de données...)

Développement

 Mise en place de mesures préventives contre des pratiques malveillantes susceptibles d'affecter l'intégrité, la disponibilité ou la confidentialité des données d'entraînement et du modèle

Utilisation

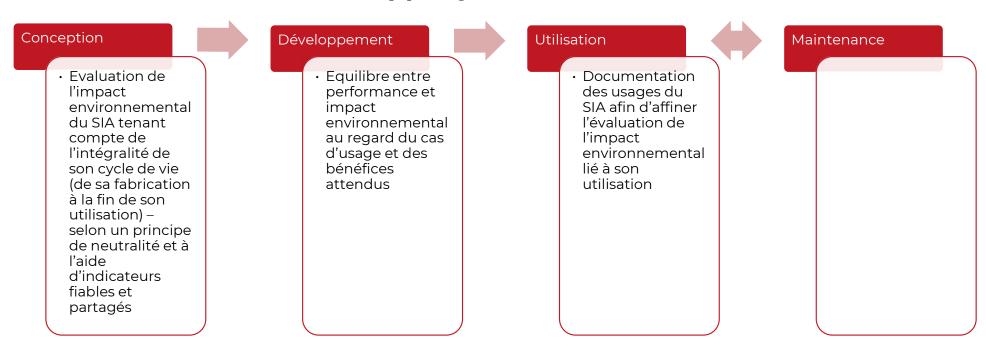
 Notification des brèches de sécurité à l'autorité compétente, voire au public si nécessaire

Maintenance

- Détection et neutralisation des tentatives d'attaque et des attaques
- ·Résolution des conséquences des attaques

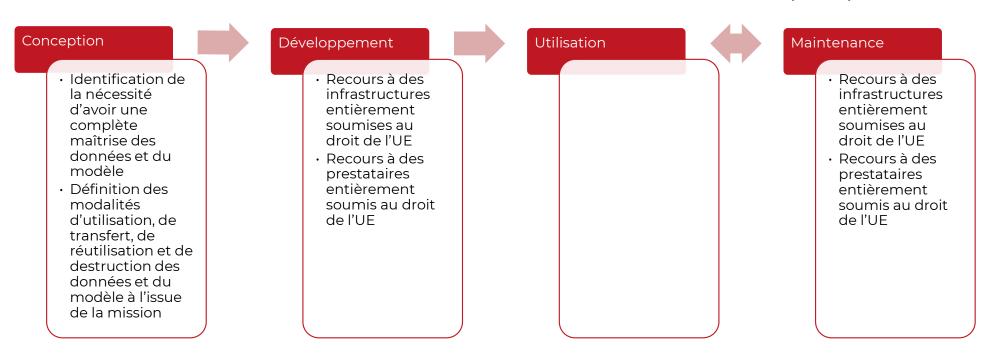


Soutenabilité environnementale : « La décision de recourir à un SIA, l'arbitrage entre l'achat et la conception en régie et, selon le cas, les critères de l'achat public et les choix internes de conception doivent [...] intégrer cette dimension. »





Autonomie stratégique: Lorsqu'il importe d'assurer une complète maîtrise des données, outre le recours à des infrastructures propres, il apparaît légitime que seules les entreprises entièrement soumises au droit de l'UE soient admises à candidater à une commande publique.



Quelle est la portée d'une démarche éthique?

- > Participer à la définition d'une doctrine administrative de l'IA de confiance
- > Construire la confiance avec les utilisateurs et les destinataires de SIA, qu'ils soient agents ou usagers du service public
- Anticiper de futures exigences légales concernant les SIA « à haut-risque » au sens de la proposition de règlement sur l'IA, publiée par la Commission européenne en avril 2021 et dont l'adoption est prévue prochainement

Quelques ressources pour aller plus loin:

- Groupe d'experts de haut niveau sur l'IA (UE), <u>Lignes directrices en matière d'éthique pour une IA digne</u> de confiance, 8 avril 2019
- OCDE, <u>Recommandation du Conseil sur l'IA</u> (OECD/LEGAL/0449), adoptée le 22 mai 2019
- Groupe d'experts de haut niveau sur l'IA (UE), <u>Assessment List For Trustworthy AI for self-assesment</u>, 17
 juillet 2020
- Commission européenne, <u>COM(2021)206 Proposition de Règlement européen établissant des règles</u>
 harmonisées concernant l'IA, 21 avril 2021
- UNESCO, <u>Recommandation sur l'éthique de l'IA</u>, adoptée le 23 nov. 2021
- Conseil de l'Europe (Comité sur l'IA), <u>Projet zéro révisé de convention [cadre] sur l'IA, les droits de l'Homme, la démocratie et l'Etat de droit</u>, 6 janv. 2023



Quelques ressources pour aller plus loin:

- Ethical AI, Guide pratique pour des IA éthiques, sept. 2021
- Hub France IA, L'IA éthique en pratique Opérationnaliser votre SIA avec une démarche éthique, mai
 2023

• ...



MERCI POUR VOTRE ATTENTION!

Pour tout renseignement complémentaire, n'hésitez pas à me contacter : anthea.serafin@ekitia.fr