Questionnaire d’auto-évaluation: Identifier et prévenir les biais dans vos projets

# NB: Ce questionnaire expérimental est largement inspiré de la liste d’évaluation issue des [lignes directrices en matière d’éthique pour une IA de confiance](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60427) rédigées par des experts indépendants de la CE. Ce document est à la base du [livre blanc sur l’IA](https://op.europa.eu/fr/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1/language-fr) publié en février dernier. Livre blanc qui annonce la très probable publication d’une nouvelle règlementation sur les systèmes d’intelligence artificielle, notamment ceux d’apprentisage automatique.

Pouvoir répondre à ce questionnaire est une façon d’anticiper une réglementation qui ne manquera pas de se mettre en place et de toute façon, même s’ils sont difficiles à détecter, les risques de discrimination tombent déjà sous le coup de la loi !

# Questions préliminaires

- Rappelez l’objectif de votre système d’Intelligence Artificiel (SIA) et décrivez (population cible, échantillonnage, variables, présence ou non de données personnelles) le corpus de données - recueillies ou à recueillir - nécessaires à sa réalisation.

- Avez-vous évalué la balance bénéfices/risques de la mise en place de votre solution d’Intelligence artificielle?

- Quels sont les impacts sur votre organisation de la mise en place de la solution d’Intelligence?

# Questions portant sur les biais dans les données

- Avez-vous évalué et reconnu les éventuelles limites provenant de la composition des ensembles de données utilisés?

- 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? Ex: outils de visualisation des données, enquêtes qualitatives de terrain, etc.

- Les données sont-elles susceptibles de refléter des erreurs de mesures ? D’être affectées d’absences (données manquantes)? De nécessiter des transformations et/ou ré-équilibrages (groupes rares)?

- Si des personnes sont susceptibles d’être impactées, avez-vous vérifié la présence possible de biais sociétaux: dépendance entre la variable cible du SIA et des groupes (sexe, origine ethnique…) protégés par la loi?

- Quelle procédure est prévue pour détecter des données éventuellement atypiques, aberrantes ou erronées pour constituer la base d’apprentissage puis en cours d’exploitation?

# Définition de l'équité

- Avez-vous prévu une **définition** appropriée de l’«équité» que vous appliquez dans la conception des systèmes d’IA? Quel est l’objectif souhaitable de l’algorithme ?

* 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?
* Prévoyez-vous des mécanismes visant à garantir l’équité dans vos systèmes d’IA? Avez-vous envisagé d’autres mécanismes potentiels?

# Questions sur les algorithmes développés et leurs impacts

## Questions préliminaires sur le fonctionnement de l’algorithme

- Caractériser le type d’algorithme mis en place: est-il non-supervisé ou supervisé, par renforcement?

- L’algorithme est-il entraîné sur vos propres données? Votre solution a-t-elle recours à un algorithme pré-entraîné?

- Comment ont été sélectionné l'échantillon d'apprentissage et l’échantillon test ? Quels effectifs? Préciser la procédure de validation pour l'optimisation des hypers paramètres.

- Quelles sont les précisions obtenues? Sont elles acceptables au regard de l’objectif poursuivi? Les décisions sont-elles assorties d’un niveau de confiance?

- Quelle procédure est mise en place pour détecter une éventuelle dégradation du SIA en cours d’exploitation?

## Variabilité de la décision algorithmique

- Avez-vous évalué si, dans des conditions identiques, une éventuelle variabilité des décisions est possible?

* Le cas échéant, avez-vous réfléchi aux causes probables?
* Distribution des erreurs en fonction des groupes ?

# Responsabilité et redevabilité

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?

* 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?

- Décrire les interventions humaines et identifier les responsabilités dans les différents processus de recueil des données, conception, déploiement, surveillance d’un SIA.

- 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?

- 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?