



# Éthique et données : chartes et guides de bonnes pratiques

Karën Fort

karen.fort@loria.fr / <https://members.loria.fr/KFort>

14 mars 2023

## Sources d'inspiration

- ▶ AI HLEG : recommandation du groupe d'experts européen sur l'IA (EN)
- ▶ Une grille d'analyse conséquentialiste [Lefeuvre et al., 2015] (FR)
- ▶ CERNA : éthique de la recherche en apprentissage machine (FR)
- ▶ CCNE : enjeux éthiques des agents conversationnels (FR)
- ▶ Data Ethics Decision Aid (DEDA)

Le piège des bonnes pratiques

Éthique de l'apprentissage machine

Recommandations du High-Level Experts Group (HLEG) on IA

Pour finir



## Les bonnes pratiques ne résoudre pas le problème

*"Currently, AI ethics is failing in many cases. Ethics lacks a reinforcement mechanism. Deviations from the various codes of ethics have no consequences. And in cases where ethics is integrated into institutions, it mainly serves as a marketing strategy. Furthermore, empirical experiments show that reading ethics guidelines has no significant influence on the decision-making of software developers."*  
[Hagendorff, 2020]

## Rappel : au-delà des bonnes pratiques

Les checklists et les grilles sont attirantes :

- ▶ elles sont simples
- ▶ elles donnent une illusion d'exhaustivité

Mais elles sont loin d'être suffisantes :

*" Neither the risk analysis informed by engineering practice, nor the socially informed engineering practice can be replaced by the other." [Gurses et al., 2011]*

## Profiter des grilles sans oublier leurs limites

1. commencer à réfléchir/discuter **sans** elles
2. les utiliser en complément
3. ne pas limiter sa réflexion parce qu'on a tout vérifié sur la liste/grille

## Des lectures que je recommande (voir Sources)

- ▶ AI HLEG : recommandation du groupe d'experts européen sur l'IA (EN)
- ▶ Une grille d'analyse conséquentialiste [Lefeuvre et al., 2015] (FR)
- ▶ CERNA : éthique de la recherche en apprentissage machine (FR)
- ▶ CCNE : enjeux éthiques des agents conversationnels (FR)



Le piège des bonnes pratiques

**Éthique de l'apprentissage machine**

Recommandations du High-Level Experts Group (HLEG) on IA

Pour finir

# CERNA

= Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistene

- ▶ créée en 2012, n'existe plus aujourd'hui (CCNE)
- ▶ plusieurs acteurs de la recherche publique : Inria, CEA, CNRS, etc.
- ▶ a organisé plusieurs journées d'études (conférences) sur différents sujets :
  - ▶ les robots
  - ▶ l'apprentissage machine (*machine learning*)
  - ▶ etc.
- ▶ groupes de travail, produisant des rapports

# Éthique de l'apprentissage machine : six thèmes

1. les données
2. l'autonomie
3. l'explicabilité
4. la prise de décision
5. le consentement
6. la responsabilité

Le piège des bonnes pratiques

Éthique de l'apprentissage machine

Recommandations du High-Level Experts Group (HLEG) on IA

Pour finir

# Guide de bonnes pratiques éthiques pour une IA de confiance

<https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf>

4 principes éthiques :

1. respect de l'autonomie humaine
2. prévention de toute atteinte
3. équité
4. explicabilité

+ tensions entre eux : les décisions prises doivent être documentées et argumentées

## Respect de l'autonomie humaine

<https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf>

*"En l'absence de justification, les systèmes d'IA ne devraient pas subordonner, contraindre, tromper, manipuler, conditionner ni régenter des êtres humains. Au contraire, les systèmes d'IA devraient être conçus afin d'augmenter, de compléter et de favoriser les compétences cognitives, sociales et culturelles. La répartition des tâches entre êtres humains et systèmes d'IA devrait suivre des principes de conception centrés sur l'humain et donner à l'être humain une possibilité réelle de poser des choix. En d'autres termes, **il convient de veiller à la supervision et au contrôle humains sur les processus de travail des systèmes d'IA.**"*

*"moins un être humain peut exercer de contrôle sur un système d'IA, plus il faut approfondir les essais et renforcer la gouvernance."*

## Prévention de toute atteinte

<https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf>

*"Les systèmes d'IA ne devraient ni porter atteinte, ni aggraver toute atteinte portée, ni nuire aux êtres humains d'une quelconque autre manière. Cela englobe la protection de la dignité humaine ainsi que de l'intégrité mentale et physique. Les systèmes d'IA et les environnements dans lesquels ils évoluent doivent être sûrs et sécurisés. [...] Il convient également d'accorder une attention particulière aux situations dans lesquelles les systèmes d'IA peuvent entraîner ou aggraver des incidences négatives du fait d'asymétries de pouvoir ou d'information, par exemple entre les employeurs et les travailleurs, entre les entreprises et les consommateurs ou entre les pouvoirs publics et les citoyens."*

# Équité

<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

*"La mise au point, le déploiement et l'utilisation de systèmes d'IA doivent être équitables. [...] En outre, l'utilisation de systèmes d'IA ne devrait jamais avoir pour conséquence de tromper les utilisateurs (finaux) ou de limiter leur liberté de choix. L'équité implique en outre que les professionnels de l'IA devraient respecter le principe de proportionnalité entre la fin et les moyens, et examiner de manière attentive la manière de trouver un équilibre entre des intérêts et des objectifs en concurrence. Le volet procédural de l'équité suppose la capacité de contester les décisions prises par des systèmes d'IA et par les êtres humains qui les utilisent, ainsi que celle d'introduire un recours efficace à l'encontre de ces décisions. Pour ce faire, l'entité responsable de la décision doit pouvoir être identifiée, et le processus de prise de décisions devrait pouvoir être expliqué."*



# Explicabilité

<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

*"Cela signifie que les **processus doivent être transparents**, que les capacités et la finalité des systèmes d'IA doivent être communiquées ouvertement, et que les décisions – dans la mesure du possible – doivent pouvoir être expliquées aux personnes directement et indirectement concernées. [...] La mesure dans laquelle l'explicabilité est nécessaire dépend fortement du contexte et de la gravité des conséquences si ce résultat est erroné ou imprécis d'une autre manière."*

Le piège des bonnes pratiques

Éthique de l'apprentissage machine

Recommandations du High-Level Experts Group (HLEG) on IA

**Pour finir**

CQFR : Ce Qu'il Faut Retenir (de tout le cours)



- ▶ l'éthique est vendeuse
- ▶ les discussions éthiques permettent de développer des fonctionnalités intéressantes
- ▶ il faut prendre le temps de discuter un projet d'un point de vue éthique. . .
- ▶ . . . mais ce temps n'est pas nécessairement long



Gurses, S., Troncoso, C., and Diaz, C. (2011).

Engineering privacy by design.

In Computers, Privacy & Data Protection.



Hagendorff, T. (2020).

The ethics of ai ethics : An evaluation of guidelines.

Minds & Machines, 30 :99–120.



Lefevre, A., Antoine, J.-Y., and Allegre, W. (2015).

Ethique conséquentialiste et traitement automatique des langues : une typologie de facteurs de risques adaptée aux technologies langagières.

In

Atelier Ethique et TRaitement Automatique des Langues (ETeRNAL'

Actes de la 1e Ethique et TRaitement Automatique des Langues (ETeRNAL'2015), Caen (France), pages 53–66, Caen, France.