

Fiche 4 — Éthique et algorithmes

Introduction

Les algorithmes sont partout : ils recommandent, filtrent, calculent, classent, décident. Ils influencent ce que nous voyons, ce que nous achetons, et parfois même ce que nous pensons.

Mais derrière chaque algorithme, il y a un **développeur** — une personne qui fait des choix : quelles données utiliser, quels critères privilégier, quelle transparence offrir.

L'éthique des algorithmes, c'est la conscience que **le code n'est jamais neutre**.

Un algorithme n'a pas d'intention, mais il a des effets.

Et ces effets doivent être anticipés, compris et assumés.

1. Comprendre ce qu'est un algorithme

1.1 Définition simple

Un **algorithme** est une suite d'instructions permettant de résoudre un problème ou d'automatiser une tâche.

Dans le numérique, il s'exprime sous forme de code : tri, recommandation, classification, prédiction...

1.2 Algorithme ≠ Intelligence

L'algorithme ne pense pas ; il **exécute**.

L'intelligence artificielle, elle, apprend à partir de données : plus elle en absorbe, plus elle reproduit leurs régularités — et donc leurs biais.

1.3 Pourquoi parler d'éthique ?

Parce qu'un algorithme mal conçu peut :

- exclure (discrimination automatisée),
- manipuler (désinformation, addiction),
- surveiller (collecte abusive),
- fragiliser (biais invisibles).

L'éthique vise à remettre **de la responsabilité** dans le développement.

2. Les biais algorithmiques

2.1 Origine des biais

Source	Exemple
Données	Base d'entraînement non représentative (ex : IA de recrutement favorisant les hommes)
Conception	Poids mal équilibré entre variables (ex : survalorisation d'un critère social ou économique)
Utilisation	Mauvaise interprétation des résultats ou absence de vérification humaine

2.2 Conséquences

- Discrimination indirecte (genre, origine, âge, lieu).
- Décisions automatisées injustifiées (crédit, embauche, notation).
- Perte de confiance dans le numérique.

2.3 Exemples emblématiques

- **COMPAS** : outil de prédiction de récidive pénale biaisé contre les personnes noires.
- **Amazon Recruiting Tool** : IA RH qui défavorisait les candidatures féminines.
- **YouTube / TikTok** : algorithmes de recommandation favorisant la polarisation.

3. Transparence et explicabilité

3.1 La "boîte noire" algorithmique

Les modèles complexes (IA, deep learning) produisent souvent des résultats sans explication claire.

Mais pour être légitime, un système doit être **compris** :

- par les utilisateurs,
- par les développeurs eux-mêmes,
- par les régulateurs (si nécessaire).

3.2 Principes d'explicabilité

- Décrire **les données utilisées**.
- Expliquer **les critères de décision**.
- Justifier **les limites** de l'algorithme.
- Documenter **le processus d'entraînement et de validation**.

Outils utiles : LIME, SHAP, ELI5, TensorBoard — pour interpréter les modèles d'IA.

3.3 Transparence envers les utilisateurs

Informers clairement :

- qu'un algorithme est utilisé,
- dans quel but,
- avec quelles données.

Un utilisateur averti n'est pas un obstacle : c'est un partenaire informé.

4. Responsabilité du développeur

4.1 Éthique professionnelle

- Se poser les bonnes questions avant de coder :
 - “Cet algorithme peut-il causer du tort ?”,
 - “À qui profite-t-il ?”,
 - “Qui risque d'en être exclu ?”
- Alerter son équipe ou sa hiérarchie en cas de dérive.
- Refuser de coder un système contraire à la loi ou aux valeurs fondamentales.

4.2 Documentation et traçabilité

- Décrire les hypothèses, choix de conception, limites identifiées.
- Conserver les jeux de données d’entraînement et de validation.
- Produire un **rapport éthique** (inspiré des “datasheets for datasets” de Gebru & Mitchell).

4.3 Le rôle du collectif

L’éthique ne repose pas sur un individu isolé.

Elle s’organise : **comités d’éthique, revues croisées, codes de conduite.**

Le dialogue entre développeurs, chercheurs, juristes et citoyens est essentiel.

5. Cadres et référentiels existants

Référence	Origine	Principes clés
Charte de Montréal pour une IA responsable	2018 (Canada)	Bienveillance, autonomie, justice, transparence
Lignes directrices de la Commission européenne (2019)	UE	Éthique, explicabilité, supervision humaine
INR – Numérique responsable	France	Transparence, sobriété, équité, respect des droits fondamentaux
IEEE – Ethically Aligned Design	International	Éthique dès la conception, diversité des équipes

6. Activités pratiques

Atelier 1 — Analyse critique d’un algorithme

- Choisissez un service connu utilisant des algorithmes (TikTok, Waze, Spotify, LinkedIn...).
- Décrivez :
 - ses objectifs et son fonctionnement supposé,

- les données collectées,
- les biais ou risques potentiels,
- les effets sur les utilisateurs.
- Proposez des pistes d'amélioration éthique ou technique.

Atelier 2 — Rédaction d'une mini-chartre éthique

- En groupe, rédigez une charte de 5 principes à respecter dans vos futurs projets.
- Exemples : transparence, non-discrimination, sobriété, supervision humaine, documentation.
- Publiez-la dans un fichier `ETHICS.md` ou affichez-la dans votre dépôt GitHub.

Livrables attendus

- Fiche d'analyse d'un algorithme ou service existant (1 page).
- Mini-chartre éthique de projet (`ETHICS.md`).
- Présentation orale optionnelle en petits groupes.

Évaluation

Critère	Description	Pondération
Compréhension des enjeux	Clarté de l'analyse et pertinence des risques identifiés	40 %
Capacité réflexive	Esprit critique, cohérence éthique, sens de la nuance	40 %
Qualité de la restitution	Structure, précision et lisibilité des livrables	20 %

Pour aller plus loin

- Lire : *Weapons of Math Destruction* — Cathy O'Neil (2016).
- Explorer : *AI Now Institute* et *Algorithmic Justice League*.

- Suivre : les travaux du *Conseil de l'IA* (Union européenne, CNIL, INRIA).
- Expérimenter : outils d'explicabilité (LIME, SHAP), documentation de jeux de données.

Message clé

L'éthique n'est pas une option morale, c'est une **compétence technique et citoyenne**.
Coder, c'est décider : chaque ligne trace une frontière entre ce qui est juste, et ce qui ne l'est pas.

Le développeur responsable sait **où il place cette frontière**.