IA et éthique

ref

- livre blanc "Les grands défis de l'IA générative » (https://dataforgood.fr/iagenerative/)

- Cours et TP Fidle https://fidles.cnrs.fr
  - (cours plus approfondis et exemples)

- Biais algorithmiques
- IA Générative et sécurité (privacy)
- IA générative et droits d’auteurs
- législation sur l’IA [règlement(UE) 2024/1689
- IA générative et articles scientifiques

IA et éthique

L'utilisation de logiciels d'IA générative pose des questions juridiques mais aussi éthiques.

Le règlement européen sur l'IA (*AI Act*) adopté en 2024 veut imposer des règles de transparence.

- nécessité d'informer clairement l'utilisateur qu'il communique avec une machine,
- entreprises développant des systèmes d'IA générative seraient tenues de préciser si les données utilisées pour développer leurs systèmes sont protégées par des droits d'auteur (textes scientifiques, musiques, photos, etc.).
- Europe a déjà émis le règlement général sur la protection des données (RGPD) qui assure la protection de toutes les données à caractère personnel.

Biais algorithmiques

L’utilisation d’IA génératives pose la question de réponses éventuellement biaisées.
Ces biais peuvent subvenir de manière non-intentionnelle
L’impact de ces biais, leurs reconnaissances, dépendent aussi de notre vision de la société, de notre éducation, etc.

- Voir le livre blanc sur l’IA ou il y a beaucoup d’exemples sur le sexe, l'âge, l’apparence physique

Biais algorithmiques

- loi française définit une vingtaine de critères pour caractériser une discrimination :
- *l'origine, le sexe, la situation de famille, la grossesse, l'apparence physique, la précarité, le patronyme, le lieu de résidence, l'état de santé, la perte d'autonomie, le handicap, les caractéristiques génétiques, les mœurs, l'orientation sexuelle, l'identité de genre, l'âge, les opinions politiques, les activités syndicales, la capacité à s'exprimer dans une langue autre que le français, l'appartenance – ou non – à une ethnie, une nation, une prétendue race ou une religion déterminée*

- *https://www.legifrance.gouv.fr/loda/id/JORFTEXT000018877783/*

Biais algorithmiques

Origine des biais :
Les données (si données sont biaisées, l'IA va exacerber le problème)

- Les données issues des sociétés sont biaisées initialement ! (ex: langage majoritairement utilisé est l'anglais)
- Ex: les biais peuvent provenir d'une surreprésentation de certaines séquences dans le corpus d'entraînement ou d'une mauvaise labellisation des données
- Besoin de représenter la pluralité
- Besoin d'effectuer une sélection de données de qualité
- Caractérisation de bais dans les données est un processus long et non trivial (il existe quelques métriques spécifiques à un contexte)

Biais algorithmiques

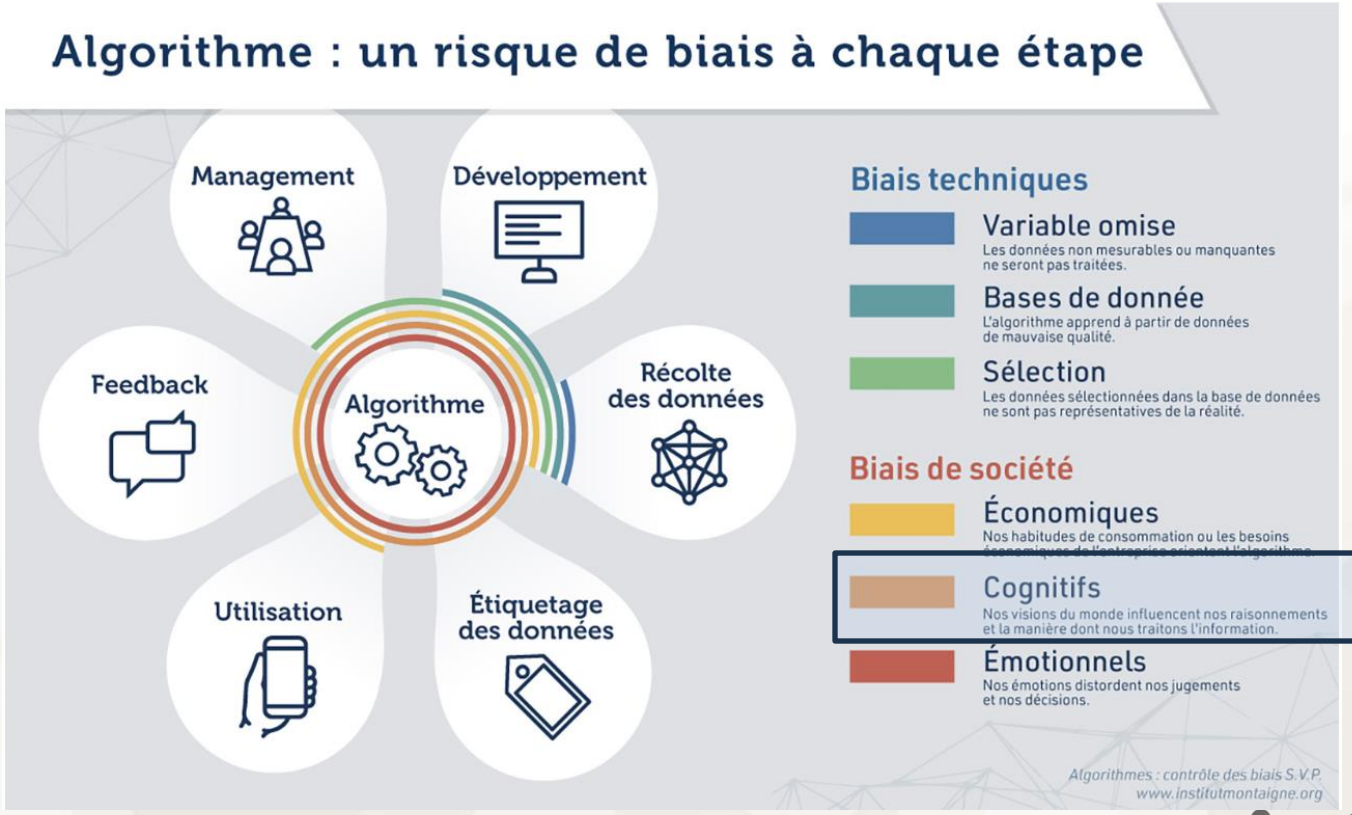
Origine des biais :
Pour éviter des biais il faudrait au moins:

- représenter la pluralité
- une sélection de données de qualité
- Caractériser les biais dans les données. Actuellement, c'est un processus long et non trivial (il existe quelques métriques spécifiques à un contexte)

Pour un grand nombre de LLM, l'accès aux données d'entrainement n'est pas accordé

Biais algorithmiques

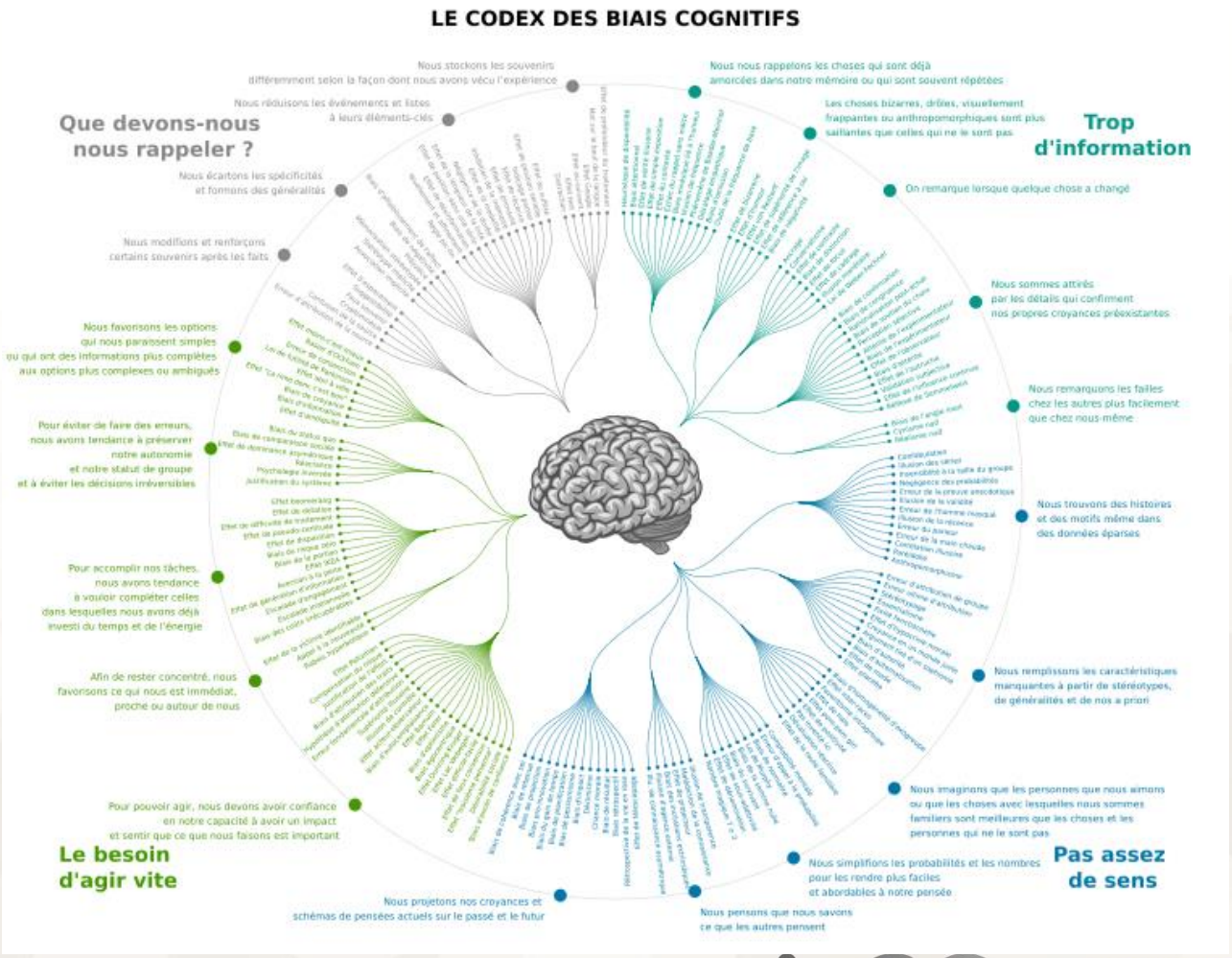
Origine des biais : the big picture



https://www.institutmontaigne.org/publications/algorithmes-controle-des-biais-svp

Biais algorithmiques

Biais cognitifs ?
=> déviation dans le traitement cognitif d'une information.



https://fr.wikipedia.org/wiki/Biais_cognitif

IA générative et droits d'auteurs

Génération de contenu avec IA

- Qui est l'auteur ? (auteurs du contenu d'entrainement, l'IA, ou l'utilisateur ?)
- Et qui détient les droits d'auteur ?

Questions difficiles et toujours en cours
En Europe, un auteur peut demander le re trait de créations sur un jeu de données, et peut attaquer en justice s'il reconnaît son œuvre dans une œuvre générée
Mais est si simple ?
Le temps de la législation est souvent très long

https://www.vie-publique.fr/rapport/277886-transposition-des-exceptions-de-fouille-de-textes-et-de-donnees

IA générative et droits d'auteurs

En Europe, un auteur peut demander le retrait de créations sur un jeu de données, et peut attaquer en justice s'il reconnaît son œuvre dans une œuvre générée

Mais est si simple ?
Le temps de la législation est souvent très long
De nombreuses entreprises (openai, midjourney, etc.) ne donnent aps accès aux jeux d'entrainement

IA Générative et sécurité (privacy)

IA Générative et sécurité (privacy)

En Europe, RGPD
Le RGPD s'applique à tous les traitements de données personnelles effectués dans le cadre des activités d'une société sur le territoire de l'Union Européenne

article 16 du RGPD, les personnes concernées disposent d'un
droit de rectification des données personnelles les concernant. Dès lors que des
données personnelles inexactes sont traitées,

IA Générative et sécurité (privacy)

- Toute utilisation de services « gratuits » implique souvent une collecte et vente de données personnelle
- Souvent nous y consentons (et donnons droit au-delà des lois de protection telles le RGPD)

- Ex: chatgpt : utilisation des réponses et des avis pour effectuer du finetuning (et ?)

- Attention aux données transmises (réponses, contexte de vie privée, fichiers)

législation sur l'IA [règlement(UE) 2024/1689

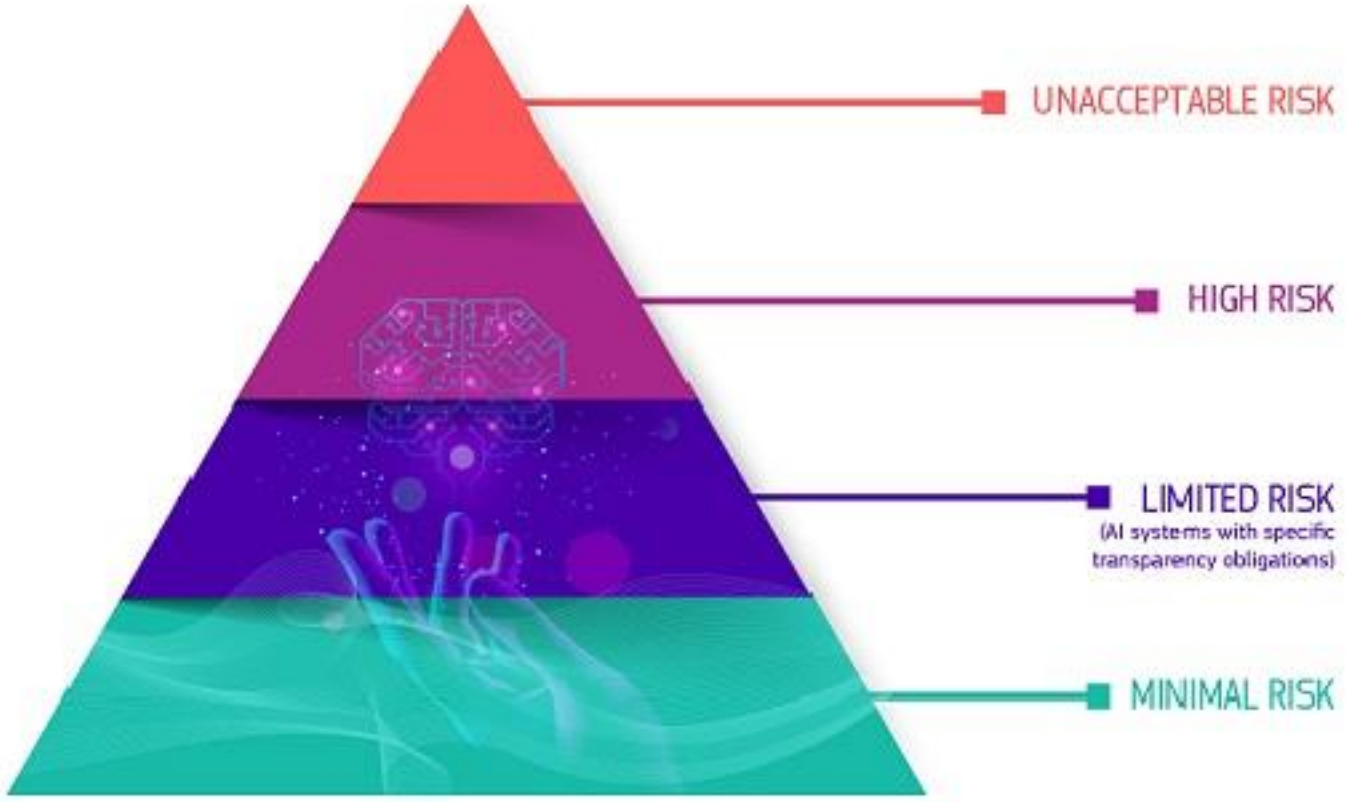
- mesures visant à soutenir le développement d'une IA digne de confiance

- traiter les risques spécifiquement créés par les applications d'IA
- interdire les pratiques d'IA qui présentent des risques inacceptables
- établir une liste des demandes à haut risque; fixer des exigences claires pour les systèmes d'IA destinés aux applications à haut risque
- définir des obligations spécifiques pour les déployeurs et les fournisseurs d'applications d'IA à haut risque
- exiger une évaluation de la conformité avant la mise en service ou la mise sur le marché d'un système d'IA donné
- mettre en place des mesures d'exécution après la mise sur le marché d'un système d'IA donné
- mettre en place une structure de gouvernance aux niveaux européen et national

- https://digital-strategy.ec.europa.eu/fr/policies/regulatory-framework-ai

législation sur l'IA [règlement(UE) 2024/1689

- **Une approche fondé sur les risques**



législation sur l'IA [règlement(UE) 2024/1689

- **Exemple de risque élevé :** la formation scolaire ou professionnelle, qui peut déterminer l'accès à l'éducation et au parcours professionnel d'une personne (par exemple, la notation des examens);

- Obligations :
- des systèmes adéquats **d'évaluation et d'atténuation des risques**;
- **qualité élevée des ensembles de données** alimentant le système afin de réduire au minimum les **risques et les résultats discriminatoires**
- journalisation de l'activité pour assurer la **traçabilité** des résultats
- une **documentation détaillée** fournissant toutes les informations nécessaires sur le système et sa finalité pour permettre aux autorités d'évaluer sa conformité;
- information claire et adéquate du déployeur
- des mesures de **surveillance humaine** appropriées pour réduire au minimum les risques;
- **haut niveau de robustesse, de sécurité et de précision**