



---

## Projet BiasDetector : Détection Intelligente de Biais avec IA Locale et RAG

---



# Sommaire :

## Table des matières

<b>1. Introduction :</b>	<b>2</b>
<b>2. Présentation du BiasDetector :</b>	<b>3</b>
<b>3. Objectifs du Projet :</b>	<b>3</b>
<b>5. Fonctionnement de l'Application :</b>	<b>3</b>
1- Upload du document :	4
2- Vectorisation :	4
3- RAG (Retrieval Augmented Generation) :	4
4- Analyse par IA locale :	4
5- Résultats :	4
<b>6. Fonctionnalités Principales :</b>	<b>4</b>
<b>6.1. Analyse de Documents :</b>	<b>4</b>
6.2. RAG (Retrieval Augmented Generation) :	4
6.3. Chat RAG :	4
6.4. Historique et Statistiques :	5
<b>7. Technologies Utilisées :</b>	<b>5</b>
<b>8. Conclusion :</b>	<b>5</b>

# 1. Introduction :

Avec l'essor du numérique et de l'intelligence artificielle, la production et la diffusion de contenus textuels ont considérablement augmenté. Articles de presse, documents académiques, rapports professionnels et publications sur les réseaux sociaux influencent quotidiennement l'opinion publique et la prise de décision. Cependant, ces contenus peuvent contenir des **biais cognitifs et sociaux**, souvent introduits de manière involontaire, qui peuvent conduire à des interprétations erronées, des discriminations ou des décisions injustes.

Les biais tels que le biais de genre, politique ou culturel, ainsi que des biais cognitifs comme le biais de confirmation ou d'ancrage, sont particulièrement difficiles à détecter manuellement, surtout lorsqu'ils sont implicites. De plus, la majorité des solutions existantes reposent sur des services cloud et des API externes, ce qui pose des problèmes de confidentialité, de sécurité des données et de dépendance technologique.

Dans ce contexte, le projet BiasDetector a été développé afin de proposer une solution intelligente, fiable et 100 % locale pour la détection automatique des biais dans les documents textuels. Ce système repose sur l'utilisation de modèles d'intelligence artificielle locaux combinés à une approche RAG (Retrieval Augmented Generation), permettant d'améliorer la qualité de l'analyse grâce au contexte fourni par des documents précédemment analysés.

L'objectif principal de ce projet est d'aider les utilisateurs à identifier, comprendre et réduire les biais présents dans leurs documents, tout en garantissant la confidentialité totale des données et en favorisant des décisions plus équitables et objectives.

## 2. Présentation du BiasDetector :

Le **BiasDetector** est un outil intelligent d'analyse de documents basé sur l'intelligence artificielle locale et la technique du RAG (Retrieval Augmented Generation). Il permet d'analyser automatiquement des fichiers textuels afin d'identifier différents types de biais, de fournir des exemples précis issus du texte et de proposer des suggestions d'amélioration.

Grâce à l'utilisation de modèles IA locaux (via **Ollama**), toutes les analyses sont effectuées sans envoyer de données à des serveurs externes, garantissant ainsi la sécurité et la confidentialité des informations analysées.

## 3. Objectifs du Projet :

Les objectifs principaux du BiasDetector sont :

- Détecter les biais cognitifs et sociaux dans les documents textuels
- Réduire les risques de racisme, discrimination ou jugements injustes
- Fournir une analyse équilibrée et objective
- Aider l'IA à produire des décisions plus justes
- Offrir une solution locale, sécurisée et sans dépendance au cloud

## 4. Types de Biais Détectés :

- BiasDetector permet de détecter six types de biais majeurs :

Type de biais	Description	Exemple
Biais de genre	Stéréotypes liés au genre	« Les femmes sont trop émotionnelles pour diriger »
Biais politique	Orientation partisane	« Seul ce parti comprend l'économie »
Biais culturel	Ethnocentrisme	« Cette culture est supérieure aux autres »
Biais de confirmation	Recherche d'informations confirmant ses croyances	« Cet article prouve que j'ai raison »
Biais de sélection	Sélection partielle des données	« 90 % des gens sont d'accord (sur 10 personnes) »
Biais d'ancrage	Influence excessive de la première information	« 800 € est une bonne affaire après 1000 € »

## 5. Fonctionnement de l'Application :

Le fonctionnement du BiasDetector repose sur plusieurs étapes clés :

## 1- Upload du document :

L'utilisateur importe un fichier au format PDF, TXT ou DOCX.

## 2- Vectorisation :

Le texte est découpé en segments (chunks), puis transformé en vecteurs numériques.

## 3- RAG (Retrieval Augmented Generation) :

Le système recherche des contextes similaires dans les documents déjà analysés.

## 4- Analyse par IA locale :

Le modèle Ollama analyse le texte en tenant compte du contexte récupéré.

## 5- Résultats :

Un rapport détaillé est généré avec :

- Les biais détectés
- Des citations exactes
- Un score global de biais
- Des suggestions d'amélioration

# 6. Fonctionnalités Principales :

## 6.1. Analyse de Documents :

- Import de fichiers PDF, TXT et DOCX
- Détection de 6 types majeurs de biais
- Calcul d'un score global de biais (0 à 100)
- Identification d'exemples précis avec citations
- Propositions de corrections et d'amélioration

## 6.2. RAG (Retrieval Augmented Generation) :

- Mémoire contextuelle basée sur les documents précédents
- Comparaison du niveau de biais entre documents
- Recherche sémantique intelligente
- Questions-réponses basées sur les analyses

## 6.3. Chat RAG :

- Questions sur les patterns de biais
- Analyse de l'historique des documents
- Compréhension des tendances globales

## 6.4. Historique et Statistiques :

- Accès à tous les documents analysés
- Suivi de l'évolution des biais dans le temps
- Statistiques par type de biais
- Évolution du score moyen global

## 7. Technologies Utilisées :

- **Backend** : FastAPI, Python
- **IA locale** : Ollama avec le modèle LLaMA 3.2
- **Base vectorielle** : LanceDB (stockage local)
- **Base de données** : MongoDB (analyses et métadonnées)
- **Frontend** : Next.js, CSS
- **Architecture** : 100 % locale, sans cloud ni API Key

## 8. Conclusion :

Le projet **BiasDetector** démontre qu'il est possible de concevoir un système performant de détection des biais tout en respectant la confidentialité des données. Grâce à l'utilisation d'une intelligence artificielle locale, combinée à une approche **RAG**, le système est capable d'analyser des documents de manière approfondie, d'identifier différents types de biais et de fournir des recommandations pertinentes pour améliorer l'objectivité des contenus.

BiasDetector se distingue par son approche 100 % locale, sans dépendance à des services cloud ou à des API externes, ce qui en fait une solution particulièrement adaptée aux environnements académiques, professionnels et sensibles. Ce projet constitue une base solide pour des améliorations futures telles que l'ajout de nouveaux formats, l'export des rapports ou encore l'intégration de tableaux de bord analytiques avancés.

En conclusion, BiasDetector contribue à la promotion de décisions plus justes, équilibrées et éthiques, tout en mettant en avant l'importance de la transparence et de la responsabilité dans l'utilisation de l'intelligence artificielle.

