



## Rapport de l'association pour la Dignité humAine

### Fiabilité des IA génératives et implications politiques et sociales

---

#### Résumé exécutif

Ce rapport vise à évaluer la fiabilité des intelligences artificielles génératives, en mettant en lumière leurs capacités, leurs limites et les enjeux qui découlent de leurs usages. À mesure que ces technologies sont déployées à large échelle dans divers secteurs de la société, il devient essentiel de comprendre comment elles fonctionnent, ainsi que les implications politiques et sociales qui en résultent. L'utilisation croissante des IA génératives dans des contextes tels que la communication, le traitement de l'information et la prise de décision soulève des questions critiques concernant leur fiabilité et la manière dont elles peuvent influencer la société. Ce document aborde ces préoccupations afin d'encadrer le débat sur l'utilisation responsable de ces technologies.

#### Principaux écueils des IA génératives détaillés dans le rapport

- **Biais dans les données** : les modèles génératifs sont souvent entraînés sur des ensembles de données contenant des préjugés culturels, sociaux ou d'autres biais qui se reflètent ensuite dans leurs productions.
- **Hallucinations** : capacité des modèles à générer des informations inexactes ou totalement fictives, souvent présentées comme des faits.
- **Raisonnements fallacieux** : tendance à produire des réponses qui semblent logiques, mais qui reposent parfois sur des fondements erronés ou des généralisations abusives.
- **Tendance à mentir** : les IA ont parfois tendance à répondre de manière positive à toutes les requêtes, même si cela implique de déformer la réalité ou de fournir des informations inexactes.

**Le déploiement des intelligences artificielles génératives à large échelle dans notre société pourrait induire de nombreuses implications politiques et sociales particulièrement critiques :**

- **Manipulation de l'information** : ces outils induisent un risque accru de désinformation et d'utilisation malveillante des contenus générés et ce avec un pouvoir d'amplification encore jamais vu.
- **Impact sur la démocratie** : le pouvoir de désinformation induit une influence potentielle sur les opinions publiques et le processus démocratique, notamment en altérant la perception de la réalité (particulièrement en lien avec le ciblage d'information sur les réseaux sociaux avec l'effet bulle de filtre).
- **Inégalités exacerbées** : le renforcement d'un fossé numérique entre celles et ceux maîtrisant l'usage de ces outils et potentiellement une aggravation des biais existants qui pourraient affecter des groupes déjà marginalisés.
- **Déstabilisation du marché du travail** : l'automatisation de tâches liés aux métiers intellectuels de type service pourrait engendrer un bouleversement en terme d'emploi dans de nombreux secteurs et impacter une part significative de la population active.
- **Érosion de la confiance** : il est probable qu'une méfiance accrue envers les informations et les contenus produits se développe, réduisant aussi la crédibilité des médias et des institutions.
- **Questions éthiques** : il est crucial de débattre sur la responsabilité des concepteurs et des utilisateurs de ces IA, notamment en matière de propriété intellectuelle et de droits d'auteur, mais aussi en terme d'impact sur les futurs citoyens et citoyennes exposés à ces technologies.

### **Points d'attention et dangers potentiels**

Les intelligences artificielles génératives soulèvent des préoccupations spécifiques, notamment en matière de prise de décision politique et d'utilisation pour des raisons sécuritaires. Leur capacité à façonner l'opinion publique et à générer de la désinformation pourrait avoir des conséquences désastreuses sur la démocratie, notamment en période électorale. De plus, l'utilisation de ces technologies dans le cadre des politiques de sécurité pourrait mener à des violations des droits civils et à une surveillance accrue des citoyens, compromettant ainsi les libertés individuelles. L'absence de régulations adéquates pourrait aggraver ces risques, appelant à une vigilance accrue des décideurs politiques et des parties prenantes pour s'assurer que ces outils soient utilisés de manière éthique et responsable dans l'intérêt public.