EduGroupe
Accompagner pour réussir

Déjouer les biais de l'Intelligence Artificielle (IA) par la pensée critique

Référence : **IA008** Durée : **1 jour (7 heures)** Certification : **Aucune**

Connaissances préalables

- Aucun prérequis

Profil des stagiaires

- Toute personne souhaitant renforcer son esprit critique face à l'IA

Objectifs

- A l'issue de la formation, le stagiaire sera capable de :
- Identifier les fondements de l'esprit critique
- Comprendre les biais cognitifs et algorithmiques
- Adopter une posture critique, éthique et stratégique face à l'Intelligence Artificielle Générative (IAG)

Certification préparée

- Aucune

Méthodes pédagogiques

- Apports théoriques interactifs
- Démonstrations en direct
- Ateliers collaboratifs
- Travaux dirigés
- Études de cas réels
- Exercices de synthèse
- Supports pédagogiques numériques
- Signature d'une feuille d'émargement pour attester de la présence à chaque demi-journée de formation

Formateur

- Consultant-formateur expert en Intelligence Artificielle (IA)

Méthodes d'évaluation des acquis

- Participation et réalisation d'exercices tout au long de la formation
- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

Contenu du cours

1. Matin (3h30)

2. Introduction aux biais en IA – 30 min

- Définitions clés
- Exemples récents de biais médiatisés
- 💡 *Démonstration en direct d'une réponse IA biaisée - Atelier : Identifier un biais simple dans une réponse ChatGPT.*

3. Typologies de biais – 45 min

- Biais liés aux données (échantillonnage, représentativité)
- Biais algorithmiques (optimisation, fonctionnalités latentes)
- Biais d'interaction utilisateur (confirmation, suggestions)
- Biais systémiques (culture, réglementation, marché)
- 💡 *Atelier : Cartographier les biais possibles pour un cas métier*

4. Pensée critique appliquée – 1h

- Méthodes QQOQCP et "lateral reading"
- Triangulation et vérification de sources
- Grilles d'analyse rapide
- 💡 *Atelier : Évaluer la fiabilité d'une réponse à forte incertitude*

5. Détection de biais dans les LLM – 1h15

- Tests A/B et "prompts sentinelles"
- Indicateurs : toxicity, demographic parity
- Outils open source : Bias Bench, Aequitas
- 💡 *Atelier : Appliquer un outil de scoring sur des réponses générées*

6. Après-midi (3h30)

7. Stratégies de mitigation – 1h15

- Curations et enrichissement des données
- Prompt engineering anti-biais
- Post-traitement et boucles de rétroaction humaine
- 💡 *Atelier : Réécrire un prompt et post-traiter la sortie pour réduire un biais*

8. Cadre juridique, éthique et responsabilité – 45 min

- RGPD, AI Act, normes ISO/IEC 42001
- Responsabilité éditoriale et transparence
- Principes de sécurité et de gouvernance
- Atelier : Élaborer une check-list éthique pour son organisation

9. Atelier de synthèse : audit complet d’un cas réel – 1h30

- Analyse complète d’une interaction IA : détection, mesure et mitigation
- 💡 *Activités : Travail en équipe sur un cas réel - Restitution rapide et élaboration d’un plan d’actions.*

Notre référent handicap se tient à votre disposition au 01.71.19.70.30 ou par mail à referent.handicap@edugroupe.com pour recueillir vos éventuels besoins d’aménagements, afin de vous offrir la meilleure expérience possible.