[sfeir] Institute

Google Cloud | GCP200RESPAI

Introduction to Responsible AI in Practice

Les meilleures pratiques recommandées par Google pour une utilisation responsable de l'IA

1 jour(s) / 7h

Objectifs pédagogiques

- Aperçu des principes et pratiques de l'IA responsable
- Mettre en œuvre des processus pour vérifier les biais injustes dans les modèles d'apprentissage automatique
- Explorer des techniques pour interpréter le comportement des modèles d'apprentissage automatique d'une manière compréhensible par l'homme.
- Créer des processus qui garantissent la confidentialité des données sensibles dans les applications de machine learning
- Comprendre les techniques permettant d'assurer la sécurité des applications génératives basées sur l'IA

Public cible

Praticiens de Machine Learning et développeurs d'applications d'IA souhaitant exploiter l'IA générative de manière responsable.

Prérequis

Compréhension de base d'un ou plusieurs des éléments suivants :

- Familiarité avec les concepts de base du Machine Learning
- Familiarité avec les concepts de base de l'IA générative sur Google Cloud dans Vertex AI comme expliqué dans les cours
    - Text Generation for Applications using Gen AI Studio
    - Interactive Chat for Applications using Gen AI Studio
    - Vertex AI Model Garden
- Connaissance de base de Google Cloud et de Vertex AI, comme expliqué dans le cours Google Cloud Big Data and Machine Learning Fundamentals

Programme

Module 1: Principes de l'IA et IA responsable

- Les principes d'IA de Google
- Pratiques responsables en matière d'IA
- Bonnes pratiques générales

Module 2: L'équité dans l'IA

- Présentation de l'équité dans l'IA
- Exemples d'outils pour étudier l'équité des ensembles de données et des modèles
- Atelier : Utilisation de TensorFlow Data Validation et de TensorFlow Model Analysis pour garantir l'équité

Module 3: Interprétabilité de l'IA

- Présentation de l'interprétabilité dans l'IA
- Sélection de métrique
- Taxonomie de l'explicabilité dans les modèles ML
- Exemples d'outils pour étudier l'interprétabilité
- Atelier : Outil d'interprétabilité d'apprentissage pour la synthèse de texte

Module 4: Confidentialité le ML

- Présentation de la confidentialité dans le ML
- Sécurité des données
- Sécurité du modèle
- Sécurité pour l'IA générative sur Google Cloud

Module 5: Sécurité de l'IA

- Présentation de la sécurité de l’IA
- Tests contradictoires
- Sécurité dans Gen AI Studio
- Atelier : IA responsable avec Gen AI Studio