

Introduction to Responsible AI in Practice

Les recommandations de Google pour une utilisation responsable de l'IA

Cours Pratique de 1 jour - 7h
Réf : RAI - Prix 2024 : 980€ HT

Avec cette formation, vous effectuerez une exploration de haut niveau des meilleures pratiques recommandées par Google pour une utilisation responsable de l'IA dans différents domaines d'intérêt : équité, interprétabilité, confidentialité et sécurité. Vous apprendrez comment exploiter différents outils open source et outils sur Vertex AI pour explorer ces concepts et prendre le temps de réfléchir aux différents défis qui se posent avec l'IA générative.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Découvrir les principes et pratiques de l'IA responsable

Mettre en œuvre des processus pour vérifier les biais injustes dans les modèles d'apprentissage automatique

Savoir interpréter le comportement des modèles d'apprentissage automatique de 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

MÉTHODES PÉDAGOGIQUES

Animation de la formation en français. Support de cours officiel en anglais et au format numérique. Bonne compréhension de l'anglais à l'écrit.

CERTIFICATION

Cours officiel sans certification.

LE PROGRAMME

dernière mise à jour : 10/2023

1) Principes de l'IA et IA responsable

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

2) L'équité dans l'IA

- L'équité dans l'IA.
- Exemples d'outils pour étudier l'équité des ensembles de données et des modèles.

Travaux pratiques : Atelier : Utilisation de TensorFlow Data Validation et de TensorFlow Model Analysis pour garantir l'équité.

3) Interprétabilité de l'IA

- L'interprétabilité dans l'IA.
- Sélection de métrique.
- Taxonomie de l'explicabilité dans les modèles ML.

PARTICIPANTS

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

PRÉREQUIS

Connaissance de base de Google Cloud et de Vertex AI, comme expliqué dans le cours Réf. GCD. Compréhension des concepts de base de l'IA générative sur Google Cloud dans Vertex AI.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils sont agrées par l'éditeur et sont certifiés sur le cours. Ils ont aussi été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum trois à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Évaluation des compétences visées en amont de la formation.

Évaluation par le participant, à l'issue de la formation, des compétences acquises durant la formation.

Validation par le formateur des acquis du participant en précisant les outils utilisés : QCM, mises en situation...

À l'issue de chaque stage, ITTCERT fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Les participants réalisent aussi une évaluation officielle de l'éditeur. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les ressources pédagogiques utilisées sont les supports et les travaux pratiques officiels de l'éditeur.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

- Exemples d'outils pour étudier l'interprétabilité.

Travaux pratiques : Atelier : Outil d'interprétabilité d'apprentissage pour la synthèse de texte.

4) Confidentialité dans le Machine Learning

- Confidentialité dans le ML.

- Sécurité des données.

- Sécurité du modèle.

- Sécurité pour l'IA générative sur Google Cloud.

5) Sécurité de l'IA

- Sécurité de l'IA.

- Tests contradictoires.

- Sécurité dans Gen AI Studio.

Travaux pratiques : Atelier : IA responsable avec Gen AI Studio.

LES DATES

CLASSE À DISTANCE

2025 : 21 mars, 13 oct.

PARIS

2025 : 21 mars, 13 oct.