Training: Google Cloud
Introduction to Responsible AI in Practice

TRANING TERMS

2025-07-25 | 1 day | Kraków / Virtual Classroom
2025-08-29 | 1 day | Warszawa / Virtual Classroom

TRAINING GOALS:

The development of AI has created new opportunities to improve the lives of people around the world, from business to healthcare to education. It has also raised new questions about the best way to build fairness, interpretability, privacy, and safety into these systems. In this course, you will do a high-level exploration of Google's recommended best practices for responsible AI usage across different areas of focus: Fairness, Interpretability, Privacy and Safety. Along the way, you will learn how you can leverage different open-source tools and tools on Vertex AI to explore these concepts and spend time considering the different challenges that arise with generative AI.

Course objectives

- Overview of Responsible AI principles and practices
- Implement processes to check for unfair biases within machine learning models
- Explore techniques to interpret the behavior of machine learning models in a human-understandable manner
- Create processes that enforce the privacy of sensitive data in machine learning applications
- Understand techniques to ensure safety for generative AI-powered applications

Audience

Machine learning practitioners and AI application developers wanting to leverage generative AI in a responsible manner

CONSPECT:

- Module 1 AI Principles and Responsible AI
  - Google's AI Principles
  - Responsible AI practices
  - General best practices
- Module 2 Fairness in AI

- Overview of Fairness in AI
- Examples of tools to study fairness of datasets and models
- Lab: Using TensorFlow Data Validation and TensorFlow Model
- Analysis to Ensure Fairness

- Module 3 Interpretability of AI
  - Overview of Interpretability in AI
  - Metric selection
  - Taxonomy of explainability in ML Models
  - Examples of tools to study interpretability
  - Lab: Learning Interpretability Tool for Text Summarization

- Module 4 Privacy in ML
  - Overview of Privacy in ML
  - Data security
  - Model security
  - Security for Generative AI on Google Cloud

- Module 5 AI Safety
  - Overview of AI Safety
  - Adversarial testing
  - Safety in Gen AI Studio
  - Lab: Responsible AI with Gen AI Studio

REQUIREMENTS:

To get the most out of this course, participants should have:

- Familiarity with basic concepts of machine learning
- Familiarity with basic concepts of generative AI on Google Cloud in Vertex AI

Difficulty level

CERTIFICATE:

The participants will obtain certificates signed by Google Cloud.

TRAINER:

Authorized Google Cloud Trainer

Compendium Education Center Ltd.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Bielska 17, 02-394 Warszawa, tel.: (22) 417 41 70

BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007