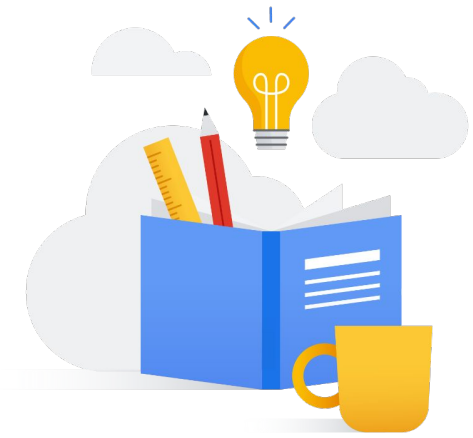
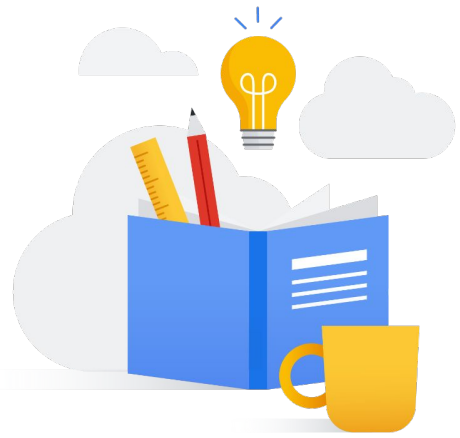
In this course, you learned ...



You've completed this course: Responsible AI for Developers: privacy and safety.

Let’s recap what you have learned.

In this course, you learned ...



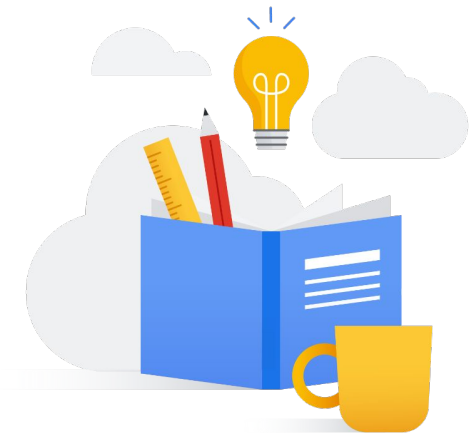
- Privacy in training data approaches:

In this course, we introduced privacy in machine learning and some best practices on privacy.

Privacy relates to the fifth of Google's AI Principles: "Incorporate privacy design principles".

We described two approaches to help developers achieve privacy in training data,

In this course, you learned ...

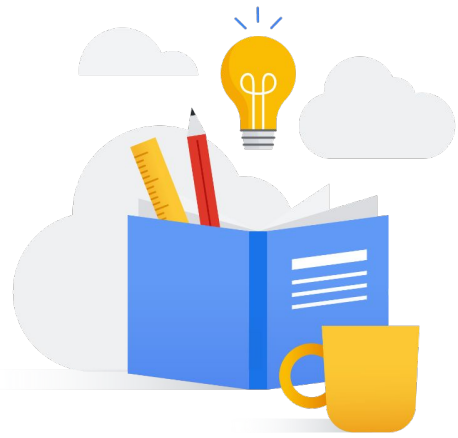


Privacy in training data approaches:

- De-identify your data.

de-identify your data

In this course, you learned ...

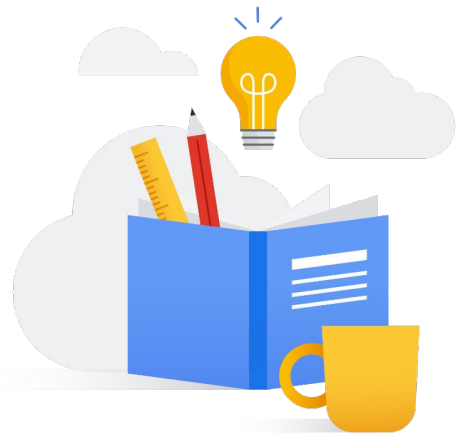


Privacy in training data approaches:

- De-identify your data.
- Randomize your data.

and randomize your data.

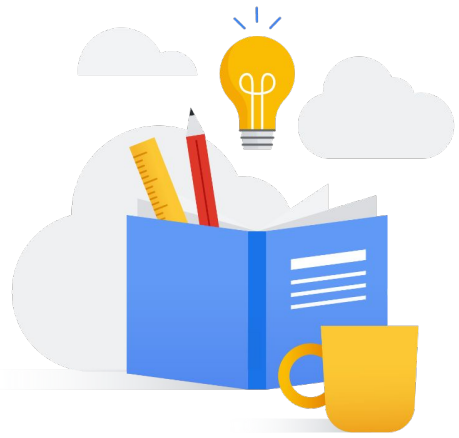
In this course, you learned ...



- Privacy in training data approaches:
  - De-identify your data.
  - Randomize your data.
- De-identify data through redaction, replacement, masking, tokenization, bucketing, or shifting.

To de-identify your data, you can apply redaction, replacement, masking, tokenization, bucketing or shifting.

In this course, you learned ...

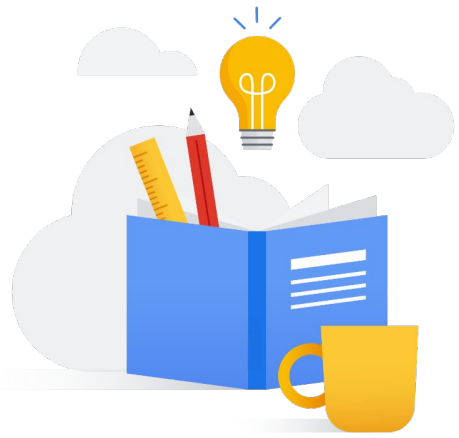


- Privacy in training data approaches:
  - De-identify your data.
  - Randomize your data.
- De-identify data through redaction, replacement, masking, tokenization, bucketing, or shifting.
- Randomize data by applying data perturbation or differential privacy.

To randomize your data, you can apply Data Perturbation or Differential Privacy.

It is recommended to apply the most suitable technique based on your business requirements for data privacy.

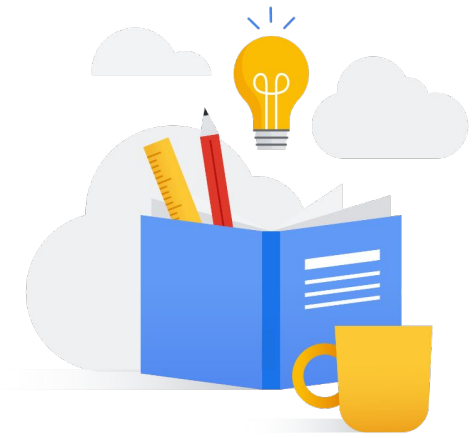
In this course, you learned ...



- Privacy in machine learning training achieved through:

We introduced two approaches to help developers achieve privacy in machine learning model training,

In this course, you learned ...

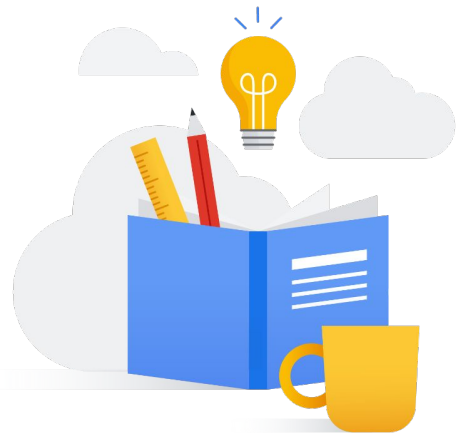


Privacy in machine learning training achieved through:

- DP-SGD

Differentially Private Stochastic Gradient Descent, also known as DP-SGD,

In this course, you learned ...

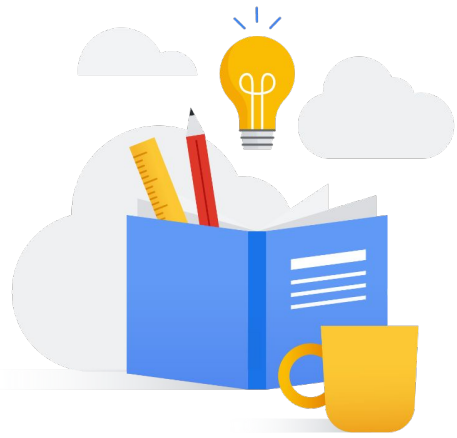


Privacy in machine learning training achieved through:

- DP-SGD
- Federated learning

and Federated learning.

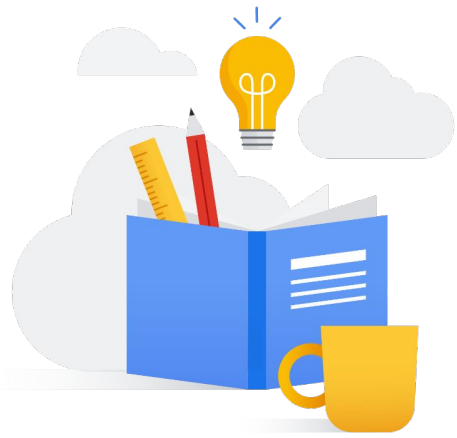
In this course, you learned ...



- Privacy in machine learning training achieved through:
  - DP-SGD
  - Federated learning
- System security best practices in Google Cloud for sensitive data protection:

You also learned about system security best practices in Google Cloud.

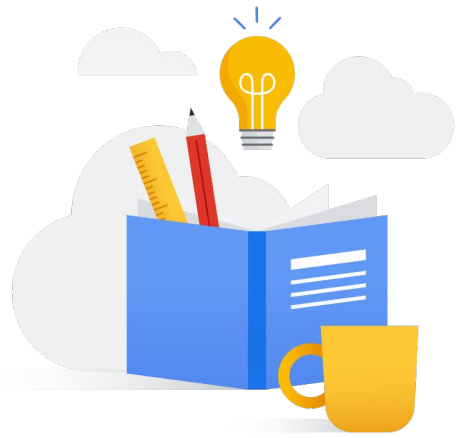
In this course, you learned ...



- Privacy in machine learning training achieved through:
  - DP-SGD
  - Federated learning
- System security best practices in Google Cloud for sensitive data protection:
  - Cloud DLP

Focus on sensitive data protection with the Cloud Data Loss Prevention (DLP) API,

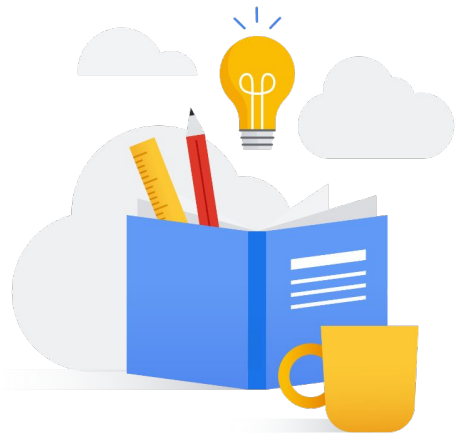
In this course, you learned ...



- Privacy in machine learning training achieved through:
  - DP-SGD
  - Federated learning
- System security best practices in Google Cloud for sensitive data protection:
  - Cloud DLP
  - Encryption KMS

Encryption with Key Management Service (KMS),

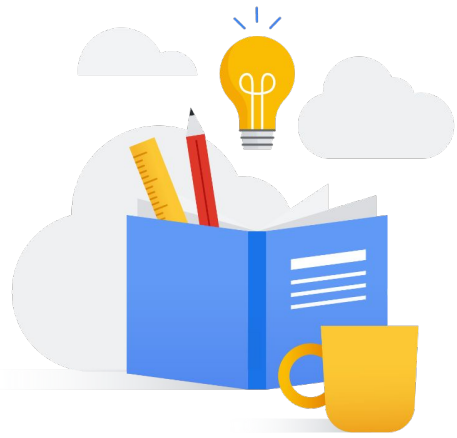
In this course, you learned ...



- Privacy in machine learning training achieved through:
  - DP-SGD
  - Federated learning
- System security best practices in Google Cloud for sensitive data protection:
  - Cloud DLP
  - Encryption KMS
  - Access control

Access control with Identity & Access Management (IAM),

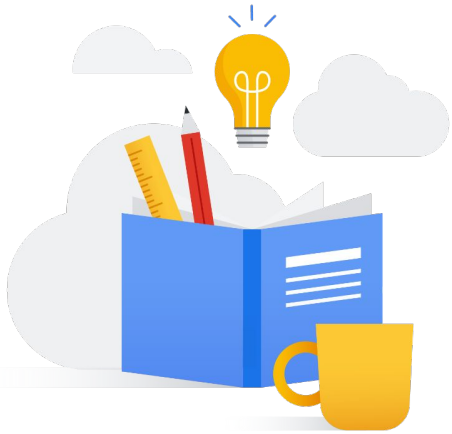
In this course, you learned ...



- Privacy in machine learning training achieved through:
  - DP-SGD
  - Federated learning
- System security best practices in Google Cloud for sensitive data protection:
  - Cloud DLP
  - Encryption KMS
  - Access control
  - Monitoring

and Monitoring.

In this course, you learned ...

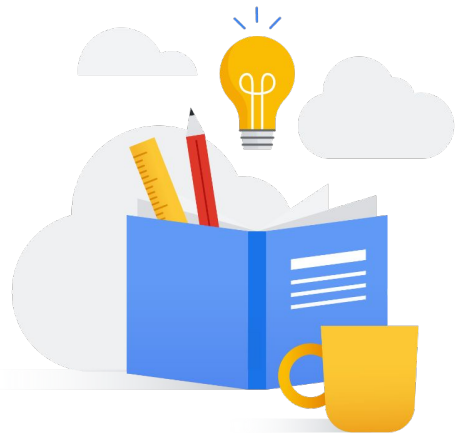


Safety is related to Google's third AI principle:

- Be built and tested for safety

We also introduced AI safety. Safety directly relates to the third of Google's AI Principle: "Be built and tested for safety".

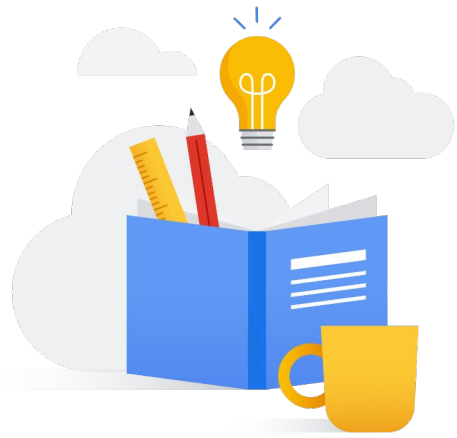
In this course, you learned ...



- Major failure modes and adversarial testing techniques for safety evaluation.

You learned about several major failure modes and adversarial testing techniques for safety evaluation.

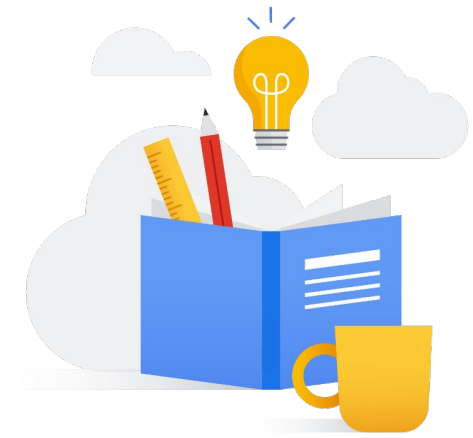
In this course, you learned ...



- Major failure modes and adversarial testing techniques for safety evaluation.
- Safety classifiers, and input and output safeguards that help you prevent harm.

You identified what are safety classifiers and what are input, output safeguards that help you prevent harm.

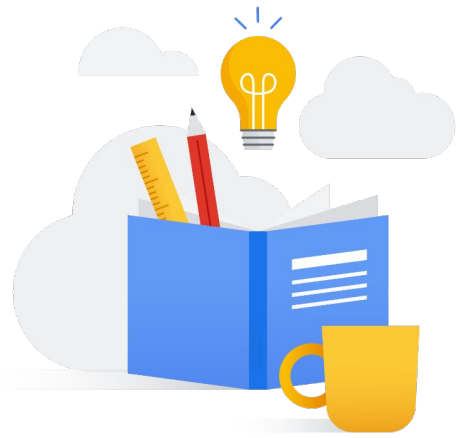
In this course, you learned ...



- Two approaches to teach safety to an AI model:

You learned about two approaches,

In this course, you learned ...

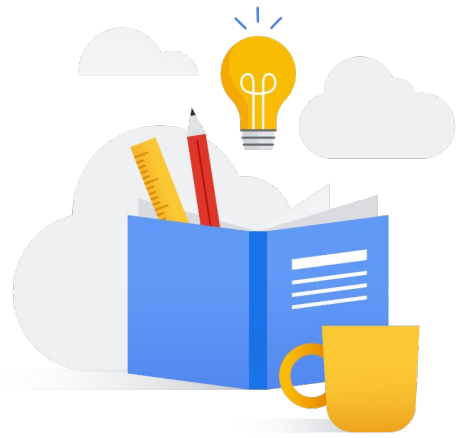


Two approaches to teach safety to an AI model:

- Instruction-tuning

Instruction-tuning

In this course, you learned ...

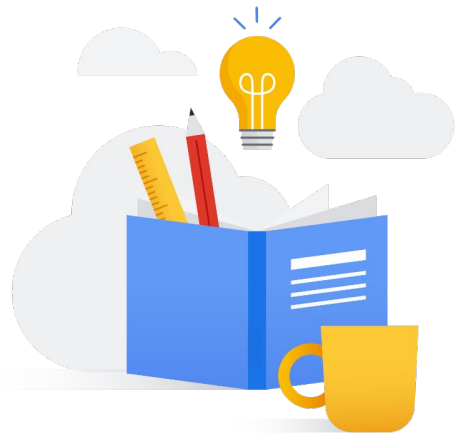


Two approaches to teach safety to an AI model:

- Instruction-tuning
- Reinforcement learning from human feedback

and Reinforcement Learning from Human Feedback, to teach safety to AI model.

In this course, you learned ...



- Two approaches to teach safety to an AI model:
  - Instruction-tuning
  - Reinforcement learning from human feedback
- Safety settings in Gemini and Natural Language API

You also learned the safety settings in Gemini and Natural Language API.

Stay tuned!

As artificial intelligence continues its rapid ascent, the conversation around responsible AI becomes ever more vital.

New technological developments constantly present fresh challenges and opportunities in this domain.

t’s even more important now to ensure that when you develop for AI, you are equipped with the latest insights and best practices for responsible AI implementation.