

Responsible AI for Developers: Privacy & Safety

Find links to the resources and further reading mentioned in the training.

[Privacy Considerations in Large Language Models – Google Research Blog](#)

[Generative AI and Data Governance | Vertex AI | Google Cloud](#)

[Identifying Machine Learning Bias With Updated Data Sets | by Jigsaw](#)

[Incorporating Priors with Feature Attribution on Text Classification](#)

[Deep Learning with Differential Privacy](#)

[Extracting Training Data from Large Language Models](#)

[Scaling Instruction-Finetuned Language Models](#)

[Challenges in Detoxifying Language Models](#)