

In this course, you learned ...



You've completed the course: Responsible AI for Developers: Interpretability and Transparency.
Let's recap what you have learned.

In this course, you learned ...



Interpretability and transparency are key tools to help various stakeholders understand AI-based systems.

In this course, we introduced interpretability and transparency in AI. Interpretability and transparency are key to mitigating unfair biases in AI, and are related to the fourth of Google's AI Principles: "Be accountable to people".

In this course, you learned ...



In this course, you learned ...



✓ **Interpretability technique categories:**
Feature-based, Concept-based, and
Example-based methods.

You also learned about interpretability techniques and how they are categorized into Feature-based, Concept-based, and Example-based methods.

In this course, you learned ...



Interpretability technique categories:

Feature-based, Concept-based, and Example-based methods.

Feature-based explanations:

Permutation feature importance, LIME, Integrated Gradients

You learned about feature-based explanations where there are Global techniques such as permutation feature importance and partial dependence plots, and local methods such as LIME, Shapley Values, Integrated gradients, and XRAI.

In this course, you learned ...



✓ Interpretability technique categories:

Feature-based, Concept-based, and Example-based methods.

✓ Feature-based explanations:

Permutation feature importance, LIME, Integrated Gradients

✓ Concept-based explanations:

TCAV and ACE

You also gained knowledge around concept-based explanations, such as TCAV, which aims to provide explanations for arbitrary concepts

In this course, you learned ...



✓ Interpretability technique categories:

Feature-based, Concept-based, and Example-based methods.

✓ Feature-based explanations:

Permutation feature importance, LIME, Integrated Gradients

✓ Concept-based explanations:

TCAV and ACE

✓ Example-based explanations:

Approximate nearest neighbor-based explanations

And you learned about example-based explanations which provide approximate nearest neighbor-based explanations.

In this course, you learned ...



In this course, you learned ...



Tools for **Interpretability** include SHAP library, LIT, and Vertex Explainable AI.

Lastly, you explored a few interpretability tools, such as open source library SHAP, learning interpretability tool, and Vertex explainable AI.

In this course, you learned ...



Tools for **Interpretability** include SHAP library, LIT, and Vertex Explainable AI.



Tools for **Transparency** include Data Card and Model Card.

As well as a few transparency tools, such as data cards for data transparency, and model cards for model transparency.

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. It'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.