Pierian Training

Vision and Images

Gemini Python API

- **Gemini Vision**
  - Beyond the text model, there is “gemini-pro-vision” which allows for multimodal input, where we can pass in a list of text along with a corresponding image.
  - We could ask for a description of an image, or ask the model to use the image for context to a query.

Gemini Python API

- **Gemini Vision**
  - It should be noted that the vision model has about half the allowed max output tokens ~4000 as the text model and the vision model is not optimized for multi-turn chat.
  - When using the vision model, try to optimize for use cases for a single input and output generation.

Gemini Python API

- **Gemini Vision**
  - Let’s explore how to use the vision model with the Python API!