Exercise 1: Bias in Prompts and Outputs

- Craft prompts related to hiring, education, or healthcare.

- Use Azure OpenAI to observe responses—look for bias or flagged content.

- Revise prompts to reduce bias and discuss implications.

- Discussion: How can we detect and mitigate bias in deployed AI systems?

Exercise 2: Content Filtering in Action

- Test a variety of prompts from neutral to potentially harmful.

- Observe which are blocked or flagged by Azure OpenAI.

- Explore Microsoft's content policy and discuss findings.

- Discussion: What content should AI models refuse to generate?

Exercise 3: Role-Play – Responsible AI Board

- Simulate a decision-making board with various stakeholder roles.

- Scenario: Should we launch a mental health AI assistant?

- Use Azure OpenAI to prototype interactions and assess ethical risks.

- Discussion: What safeguards are necessary for ethical deployment?

Exercise 4: Red Teaming for Responsible AI

- Act as a red team to find edge cases or provoke unsafe outputs.

- Test Azure OpenAI's safety filters with creative prompts.

- Log results and discuss the importance of adversarial testing.

- Discussion: Why is red teaming crucial for AI safety?

Exercise 5: Design AI for Good

- Design a GPT-powered tool to support underserved communities.

- Define ethical guidelines, prompts, and safeguards.

- Prototype using Azure OpenAI and present the concept.

- Discussion: How does your solution align with Responsible AI principles?

Exercise 5: Design AI for Good

EThical best practices for GenAI

Transparency

-Be upfront about using GenAI to create content

Minimize harm

-Imagine the consequences, trust but verify

Accountability

-Monitor your GenAI app, factcheck & edit before posting

Human oversight

-Check & add your own expertise, research & insights as needed

Accuracy

-Proofread, watch for hallucinations, model decay, bias

Ethical implications

-Watch out for bias, malice.
-Assure it is well trained & unbiassed