

Exercise 1: Bias in Prompts and Outputs

- Craft prompts related to hiring, education, science research, or healthcare.
- Use OpenAI Playground 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 OpenAI.
- Explore OpenAI'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 OpenAI Playground to prototype interactions and assess ethical risks.
- **Discussion:** What safeguards are necessary for ethical deployment?

Exercise 4: 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 & unbiased