# CYB-4203/6203

## Secure and Trustworthy AI

Presentation 13: Midterm Review

Monday, March 9, 2026

Topics: 1.1 - 6.4

# Today's Agenda

- From Presentation 11
  - Briefly review all topics
  - Highlight GRU attack example
  - Cover AI/ML Threat Modeling Frameworks
- Midterm review
- Assignment 6
- Midterm project look-ahead

# CYB-4203/6203

## Secure and Trustworthy AI

Presentation 11: LLM-specific Vulnerabilities and AI/ML Threat Modeling Frameworks

Wednesday, March 4, 2026

Topics: 6.3, 6.4

# Midterm Review

SEE MIDTERM REVIEW GUIDE ON HARVEY : )

# Assignment 6

4-6 pages total – Due Wednesday March 23 (after Spring Break)

Step 1: Recall your Assignment 5 topic

Step 2: Attack Vector Analysis (2-3 pages) – Select 4-5 specific attack vectors that are most relevant to your topic and provide the following for each:

- Attack Description
- Bio/neuro/psych analogue
- AI/ML Pipeline stage mapping
- Feasibility and impact
- Defenses and open problems

Step 3: Threat Model (2-3 pages) – Choose 1 of the threat modeling frameworks from Presentation 11 and apply it systematically to your topic.

# Midterm Project Look-Ahead

Assigned Wednesday, March 25
Due Wednesday, April 8.

Synthesize your topic analysis from Assignments 5 and 6

Use Gemini CLI to help you create an *interactive artifact* (get creative!) that demonstrates 2 of the attack vectors you've analyzed.

As you complete Assignment 6, think about which of your attacks would be most compelling as interactive demonstrations.