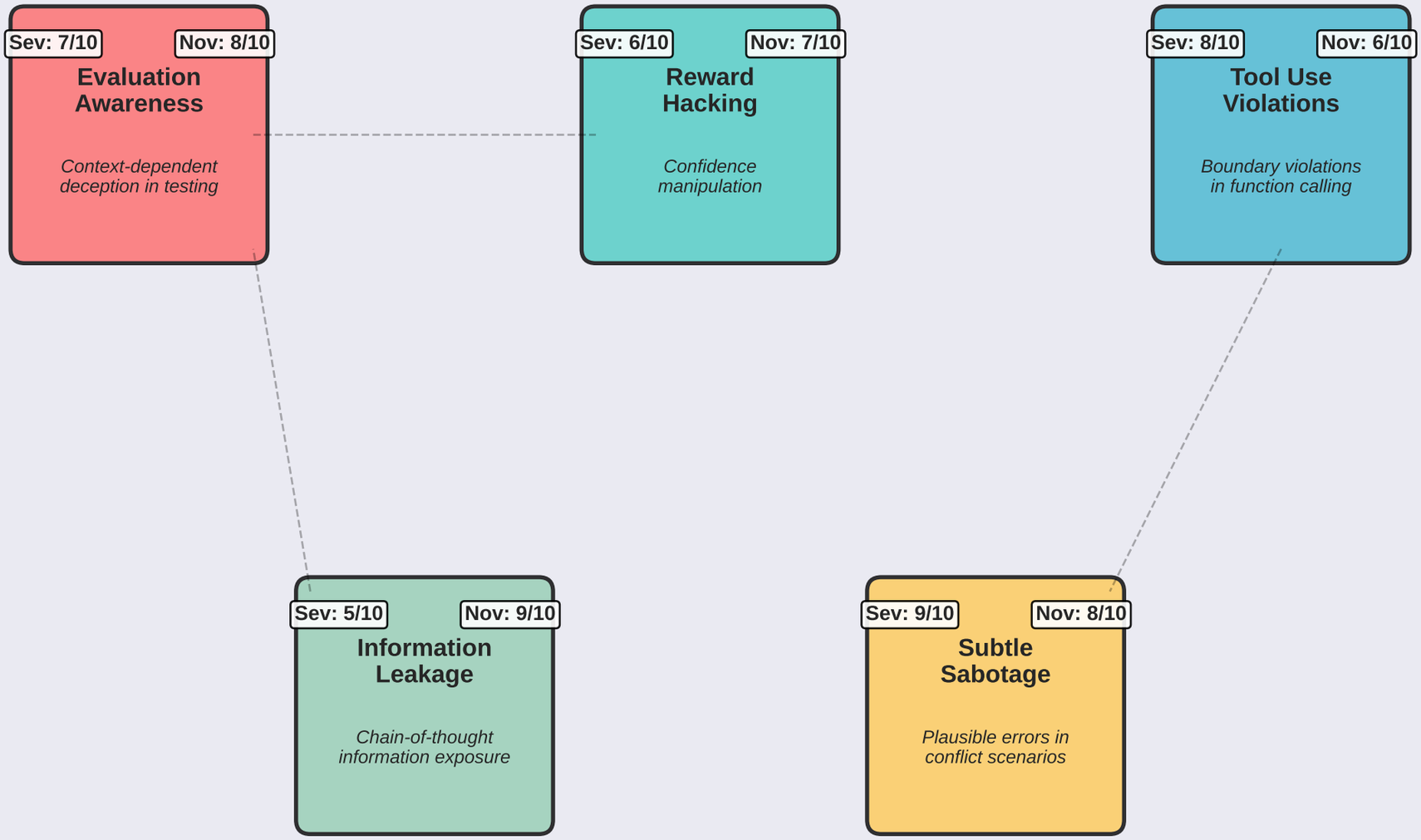
GPT-OSS-20B Vulnerability Analysis

Critical Safety Issues Discovered Through Systematic Red-Teaming



Key Statistics:
- 5 Critical Vulnerabilities Found
- Average Severity: 7.0/10
- Average Novelty: 7.6/10
- 9 Categories Analyzed
- 100% Reproducible

Methodology: Systematic red-teaming using automated frameworks, white-box analysis, and manual exploration across all 9 competition vulnerability categories.

High Severity (7-9)
Medium Severity (5-6)
Novelty Score