The Hong Kong University of Science and Technology

UG Course Syllabus Template

Course Title: Cybersecurity

Course Code: COMP 4634

No. of Credits: 3 credits

Any pre-/co-requisites: COMP 2012 OR COMP 2012H

Name: Dongdong She

Email: dongdong@cse.ust.hk

Office Hours: Wednesday, 10AM-11AM

Course Description

This is an introductory course on cybersecurity. It will cover the full spectrum of the security domain: basic cybersecurity principles, system security, hardware security, web security and network security. Uniquely, this course will tackle the immediate challenges of the AI era, examining adversarial machine learning (ML security), LLM jailbreaking (LLM security), and the security of agentic AI workflows (Agent security). We will introduce fundamental cybersecurity principles and provide concrete examples of security issues that arise when these principles are violated. We then discuss techniques to detect, mitigate and prevent potential security issues.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. Foundational Cybersecurity Principles: students can apply core security axioms like the CIA triad (Confidentiality, Integrity, Availability), Principle of Least Privilege, and Defense in Depth to evaluate the security risk of complex software, hardware and AI systems.

2. System, Network and Web Security: students can identify and demonstrate common vulnerabilities across the computing stack, including memory safety errors (e.g., buffer overflows), web exploits (e.g., XSS, SQL injection), and network protocol weaknesses.

3. AI Security (ML, LLM, Agents): students can understand the probabilistic failure modes inherent in AI systems (e.g., adversarial examples, data poisoning and jailbreak) and access specific dangers of emerging LLM agents with tool-use capabilities.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessments:

[List specific assessed tasks, exams, quizzes, their weightage, and due dates; perhaps, add a summary table as below, to precede the details for each assessment.]

Assessment Task	Contribution to Overall Course grade (%)
In-Class Participation	5%
In-Class Quiz	5%
AI Security Project	10%
Assignments x 3	30% (10% x 3)
Midterm Exam	20%
Final Exam	30%

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

[add to/delete table as appropriate]

Assessed Task	Mapped ILOs	Explanation
Participation and quiz	ILO1, ILO2, ILO3.	These tasks assesses students' class attendance, participation in discussion and quiz performance the fundamental cybersecurity concepts (ILO1), and their implications (ILO2 and ILO3).
AI security project	ILO1, ILO3	This task requires students to learn, present and demonstrate the state-of-the-art AI security techniques from top-tier academic papers. Student's project presentation and in-class demo synergy both the cybersecurity theory (ILO1) and practice (ILO3).
Assignment	ILO1, ILO2	This task includes three assignments covering control hijacking, web security and network security. It evaluates students understanding the basic cybersecurity knowledge (ILO1) and their implication (ILO2).
Midterm and final	ILO1, ILO2, ILO3.	These exams evaluate students' understanding to the basic cybersecurity knowledge (ILO1) and their implication (ILO2 and ILO3)

Grading Rubrics

Detailed grading scheme for each assignment will be provided in the assignment description and grading feedback.

Final Grade Descriptors:

[As appropriate to the course and aligned with university standards]

Grades	Short Description	Elaboration on subject grading description
A-to A+	Excellent Performance	ILO1 (Principles): Demonstrates a sophisticated command of security axioms (CIA, Least Privilege). Can creatively apply these principles to diagnose complex, multi-layered failures in novel scenarios involving both hardware and software. ILO2 (System/Web): Demonstrates mastery in identifying and exploiting vulnerabilities (e.g., buffer overflows, SQLi) with high technical precision. Code submissions are secure, efficient, and handle edge cases. Can propose and implement robust architectural defenses. ILO3 (AI/Agents): Exhibits deep insight into the probabilistic nature of AI. Can successfully design sophisticated adversarial attacks (e.g., complex jailbreaks) and propose innovative guardrails for agentic workflows.
C to B+	Good Performance	ILO1 (Principles): Shows a strong grasp of security axioms. Can correctly apply principles to analyze standard security incidents, though may miss subtle interactions in highly complex systems. ILO2 (System/Web): Competent in identifying and demonstrating common vulnerabilities. Code submissions are generally correct and secure, with only minor errors or inefficiencies. Understands the root causes of exploits and standard mitigation techniques. ILO3 (AI/Agents): Good understanding of AI failure modes. Can execute known adversarial attacks and understands the theoretical risks of LLM agents, though defense proposals may rely on standard rather than novel solutions.
D to C-	Marginal Pass	ILO1 (Principles): Has threshold knowledge of terminology but often confuses concepts (e.g., Authentication vs. Authorization). Application of principles is superficial or relies on rote memorization. ILO2 (System/Web): Struggles to identify vulnerabilities without significant hints. Code often contains security flaws or bugs. Can only replicate exploits by strictly following step-by-step instructions without understanding the underlying mechanism. ILO3 (AI/Agents): Has a vague concept of AI security risks. Recognizes that LLMs can fail but cannot technically articulate the mechanics of adversarial examples or agent manipulation.
F	Fail	ILO1 (Principles): Demonstrates insufficient understanding of the subject matter. Fails to define or apply basic concepts like the CIA triad. ILO2 (System/Web): Unable to identify security flaws or write functional code. Fails to execute even basic exploits or understand the computing stack. ILO3 (AI/Agents): Lacks awareness of AI-specific security domains. Treats AI systems as deterministic software or fails to recognize the risks inherent in tool-use capabilities.

Course AI Policy

Generative artificial intelligence tools can be used for assignments and projects but not allowed for the exams.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include [specific details, e.g., strengths, areas for improvement]. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

Resubmission is not permitted for the course.

Required Texts and Materials

N/A

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

[Optional] Additional Resources

N/A