The Hong Kong University of Science and Technology (Guangzhou)

UG Course Syllabus Template

Ethics, Privacy and Security in Artificial Intelligence

AIAA 2290

[No. of Credits]

[Any pre-/co-requisites]

Name: Xuming HU

Email: xuminghu@hkust-gz.edu.cn

Office Hours:

Every Friday morning 10:00-12:00, E3-303

Every Monday evening 18:00-19:00, E3-303

Course Description

This introductory course surveys the explosive area of AI ethics and illuminates relevant AI concepts with no prior background needed. Key topics include Fake News Bots; AI Driven Social Media Displacing Traditional Journalism; drone Warfare; Elimination of Traditional Jobs; Privacy-violating Advertising; Monopolistic Network Effects; Biased AI Decision/Recognition Algorithms; Deepfakes; Autonomous Vehicles; Automated Hedge Fund Trading, etc. Through the course, students will be able to understand how human civilization will survive amid the rise of AI, what are the new rules in the new era, how to preserve ethics when facing the threats of extinction and what are engineers' and entrepreneurs' ethical responsibilities.

Intended Learning Outcomes (ILOs)

1. Demonstrate a comprehension of ethics, security and privacy risks in AI technology.
2. Demonstrate a comprehension of ethically compliant AI technologies.
3. Demonstrate a comprehension of the universal defence approaches to AI privacy and security.
4. Demonstrate a comprehension of ethical, privacy and security risks in specific areas (e.g. NLP, CV, and Embodied AI) and their existing solutions.
5. Recognize the limitations of current methods and make improvements.

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:

Assessment Task	Contribution to Overall Course grade (%)	Due date
In-class Quiz	20%	28/04/2025
Student Presentation	25%	07/05/2025
Final Report	55%	26/05/2025

1. **In-class Quiz:**
    a. Description: Each week, there will be a short quiz consisting of 2 to 5 multiple-choice or multiple-answer questions, covering the key concepts discussed in class. These quizzes are designed to assess students' understanding of the material on a weekly basis.
    b. Format:
        i. Number of Questions: 2 to 5 multiple-choice or multiple-answer questions.
        ii. Duration: 10-15 minutes.
    c. Content: The questions will focus on key topics from the week's lectures, including ethical, privacy, or security issues in AI, and any other relevant concepts discussed.
2. **Student Presentation:**
    a. Description: Each student is required to deliver a presentation on a relevant topic within the course scope.
    b. Format:
        i. Duration:
        ii. 9 minutes for the presentation.
        iii. 2-3 minutes for a Q&A session with peers and the instructor.
    c. Content: Presentations should demonstrate a clear understanding of ethical, privacy, or security issues in AI, supported by relevant case studies or research.
3. **Final Report:**
    a. Description: A comprehensive written report that explores a specific topic related to Ethics, Privacy, or Security in AI.
    b. Requirements:
        i. Length: limited to 8 pages, double-spaced.
        ii. Format: Follow the given template.
    c. Content:
        i. Introduction and background of the chosen topic.
        ii. Detailed analysis and discussion of relevant issues.
        iii. Case studies or examples to illustrate key points.
        iv. Conclusion and recommendations.
    d. Late Submission Policy: Late submissions will be penalized. We will deduct 3% of the overall score for every 24 hours after the deadline (2025/05/26).

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
In-class Quiz	ILO1, ILO2, ILO3, ILO4	This task assesses students' ability to explain and apply ethical, security, and privacy risks in AI technology (ILO1), demonstrate understanding of ethically compliant AI technologies

		(ILO2), evaluate universal defence approaches to AI privacy and security (ILO3), and recognize risks and solutions in specific AI domains such as NLP, computer vision, and embodied AI (ILO4).
Student Presentation	ILO1, ILO2, ILO3, ILO4, ILO5	Students begin by demonstrating their understanding of ethical, privacy, and security risks in AI technology (ILO1), followed by an explanation of ethically compliant technologies relevant to their topic (ILO2). They are then expected to evaluate general defense approaches to AI privacy and security, such as differential privacy or federated learning (ILO3), and analyze how these concerns and solutions apply to a specific domain like natural language processing, computer vision, or embodied AI (ILO4). In the reflective component, students critically examine the limitations of current approaches and propose thoughtful improvements (ILO5), thereby synthesizing knowledge and demonstrating higher-order thinking.
Final Report	ILO1, ILO2, ILO3, ILO4, ILO5	In the introduction and background sections, students identify and explain the ethical, privacy, and security risks involved in AI systems (ILO1). The literature review showcases their understanding of ethically compliant technologies and frameworks (ILO2). The main body includes an evaluation of universal defense methods (ILO3) and a domain-specific case analysis that explores practical challenges and existing solutions in areas such as NLP, CV, or embodied AI (ILO4). Finally, in the conclusion and recommendation sections, students reflect on the limitations of current methods and articulate possible improvements or future research directions (ILO5), demonstrating a critical and integrative approach to the subject matter.

Grading Rubrics

1. Rubric and Evaluation Form of Student Presentation:

Content

Student	Significance	Originality	Diversity	Workload	Overall

Score Rubrics: **1** (Poor) | **2** (Below Average) | **3** (Borderline) | **4** (Up Average) | **5** (Good) | **6** (Very Good) | **7** (Excellent)

Presentation

Student	Introduction to the Topic	Organization and Logic	Clarity of Presentation	Fluency of Presentation	Overall

Score Rubrics: **1** (Poor) | **2** (Below Average) | **3** (Borderline) | **4** (Up Average) | **5** (Good) | **6** (Very Good) | **7** (Excellent)

Q & A

Student	Clarity of Answers	Response Organization	Accuracy and Relevance	Interaction	Overall

Score Rubrics: **1** (Poor) | **2** (Below Average) | **3** (Borderline) | **4** (Up Average) | **5** (Good) | **6** (Very Good) | **7** (Excellent)

2. Rubric and Evaluation Form of Final Report:

Content Quality (60 points)	Fundamental Concept Understanding (25 points)	Clearly explains key concepts with well-structured definitions and real-world examples. (22-25 points)	Covers key topics but lacks depth or fails to provide concrete examples. (17-21 points)	Mentions concepts but with vague explanations or missing critical details. (10-16 points)	Displays little to no understanding of core AI ethics, security and privacy concepts. (0-9 points)
	Score				

	Review and Analyze Existing Research (20 points)	Provides a comprehensive literature review with relevant and up-to-date papers, critically analyzing methodologies and findings. (18-20 points)	Covers related works but lacks deep comparison or critical evaluation. (14-17 points)	Lists some related works without much discussion or relevance. (8-13 points)	Little to no research cited, or references are outdated/irrelevant. (0-7 points)
	Score				
	Practical Application & Critical Analysis (15 points)	Demonstrates strong analytical skills by applying AI ethics, security and privacy concepts to real-world scenarios with insightful evaluation and innovative solutions. (13-15 points)	Provides relevant applications and some critical analysis but lacks depth, originality, or a strong connection to practical challenges. (10-12 points)	Mentions basic applications but offers superficial analysis, with limited critical thinking or weak real-world relevance. (6-9 points)	Shows little to no application of concepts, with unclear, incorrect, or missing critical analysis.(0-5 points)
	Score				
Expression and Format (40 points)	Language Expression (15 points)	Clear and fluent language expression, easy to understand. (13-15 points)	Generally good expression, but some parts are not fluent. (10-12 points)	Unclear expression, affects understanding.(6-9 points)	Severely unclear expression. (0-5 points)
	Score				
	Format Adherence (15 points)	Strictly follows format requirements, well-structured. (13-15 points)	Basically follows format requirements, with minor deviations. (10-12 points)	Insufficient format adherence, noticeable deviations. (6-9 points)	Fails to follow format requirements. (0-5 points)
	Score				
	References (10 points)	Appropriately selected references, correct citation format. (9-10 points)	Generally appropriate references, minor issues in format or selection. (7-8 points)	Inadequate references or incorrect citation format. (4-6 points)	Severely lacking references or improper format. (0-3 points)
	Score				

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive understanding of ethical theories, privacy implications, and security frameworks in AI. Applies concepts to complex scenarios with clarity and originality. Critically evaluates AI systems for compliance and fairness, and proactively engages in responsible AI discourse with peers. Shows leadership in applying ethical principles in real-world contexts.
B	Good Performance	Shows solid knowledge of key ethical issues, privacy concerns, and AI security practices. Can analyze case studies with logical reasoning and identify major ethical and technical challenges. Demonstrates consistent engagement with course content and contributes meaningfully to discussions on responsible AI development.
C	Satisfactory Performance	Possesses basic understanding of ethics, privacy, and security in AI. Can address familiar issues and apply general principles to straightforward cases. Shows some critical thinking but may

		struggle with complex or nuanced dilemmas. Participates adequately in coursework and group discussions.
D	Marginal Pass	Has minimal grasp of fundamental concepts in AI ethics, privacy, and security. Can recognize obvious concerns but lacks depth in analysis or application. Demonstrates effort and potential for growth, but needs guidance to meet academic and professional standards in the field.
F	Fail	Fails to understand key ethical, privacy, and security concepts related to AI. Inadequately analyzes issues or misapplies theoretical frameworks. Shows little engagement with the material and does not meet basic expectations for professional or academic development.

Course AI Policy

Three Always Principles for Using Generative AI in This Course:

1. Always Acknowledge AI Use: When using AI tools in coursework, clearly acknowledge and cite their contributions. Transparency about the use of AI ensures academic honesty and helps clarify the origin of ideas or content used in assignments.
2. Always Evaluate AI Output Critically: While AI can be a valuable aid, its responses must be critically assessed for accuracy and relevance. Students should verify AI-generated content against reliable sources and apply their own judgment to determine its appropriateness in academic work.
3. Always Uphold Academic Integrity: AI should complement—not replace—your own effort, understanding, and analysis. It must be used as a support tool, with all work ultimately reflecting your individual learning and intellectual engagement.

Communication and Feedback

To support students' learning and address any questions, a weekly office hour is held every Friday from 10:00 AM to 12:00 PM, where students are encouraged to seek clarification and engage in discussion on course content or assignments.

In addition, the course actively collects anonymous student feedback through regular surveys throughout the semester. This feedback is reviewed in a timely manner and used to make responsive adjustments to course delivery and structure, ensuring an open and adaptive learning environment.

Resubmission Policy

Resubmission opportunities are not provided for in-class quiz missed due to absence, except in cases of special circumstances such as medical leave or official leave approved by the university. All assignments, along with their respective content and deadlines, will be communicated at least two weeks in advance. Late submissions will not be accepted. The date for the final report will be provided at least one month in advance. Attendance for both is mandatory and, in principle, cannot be missed.

Required Texts and Materials

Recent research papers recommended by course instructor.

Academic Integrity

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