Introduction to the Ethics of Artificial Intelligence and Data Science

Prof. Linus Huang (LinusHuang@ust.hk)
Office: Room 2387
Time: Friday, 9:00-11:50
Location: Lecture Theater K
Office Hours:
by appointment

COURSE DESCRIPTION

How do AI and data science shape the world we live in? How should we govern their development responsibly? What kind of future do we want to build with them? This course empowers you to engage critically with these urgent questions. Through real-world case studies, you'll examine both the perils and promises of emerging technologies. You'll gain insight into issues like algorithmic bias, data surveillance, and online manipulation, while also exploring AI's potential for social good. The course highlights diverse global perspectives, contrasting Eastern and Western approaches.

Each week, you'll tackle exciting topics like privacy, inequality, human-AI interaction, and the future of work. You'll develop practical skills for ethical tech design and learn how AI is governed on a societal level. The course culminates in a final group project where you get to apply your new knowledge to a topic of your choice. Join us as we investigate the fascinating intersections of ethics, culture, and technology. This course will equip you to navigate our AI-driven world thoughtfully and help shape it for the better. No technical background required - just an open and curious mind!

LEARNING OUTCOMES

Throughout the course, students will: (1) become familiar with contemporary texts related to the ethical considerations and social implications of AI and data science; (2) develop their own perspectives on the foundational ethical questions surrounding emerging technologies; (3) analyze impacts of AI using diverse and comparative frameworks, and apply ethical reasoning to real-world issues; (4) improve written and oral communication skills to effectively discuss ideas related to AI ethics; and (5) develop independent research and teamwork skills to collaboratively investigate AI's societal implications and propose solutions.

COURSE READINGS

- ✓ All readings and videos will be made available electronically.
- ✓ The instructor reserves the right to make changes to the syllabus, including paper due dates.

READING SCHEDULE

Date	Topic	Assessment
Week 1 (Sep 1)	**Class Cancelled**	
Week 2 (Sep 8)	**Digital Age Dive:** Introduction to AI and data science.	Quiz 1
Week 3 (Sep 15)	**Generative AI Magic:** Hands-on with ChatGPT	Quiz 2
Week 4 (Sep 22)	**Data's Ethical Maze:** Privacy, ownership, and control	Annotation 1 Quiz 3 Resp. Paper 1
Week 5 (Sep 29)	**Algorithmic Pitfalls:** Bias, discrimination, and justice	Quiz 4 Resp. Paper 2

Week 6 (Oct 6)	**Social Media's Dilemma:** Digital well-being vs. manipulation	Annotation 2 Quiz 5 Resp. Paper 3
Week 7 (Oct 13)	**AI for Social Good:** Human-machine interaction and moral enhancements	Quiz 6 Resp. Paper 4
Week 8 (Oct 20)	**Values in AI:** human-centered approach and value-sensitive design	Quiz 7 Resp. Paper 5 Problem Identification (due Oct 24)
Week 9 (Oct 27)	**The ethics of automation:** interpretability, responsibility, and the future of work.	Quiz 8 Resp. Paper 6
Week 10 (Nov 3)	**Global AI Ethics:** East vs. West.	Quiz 9 Resp. Paper 7 Research Proposal (due Nov 7)
Week 11 (Nov 10)	**AI in Governance:** Surveillance, Misinformation, Policy.	Quiz 10 Resp. Paper 8 Research Write-up (due Nov 16)
Week 12 (Nov 17)	**Group Project Workshop**	Quiz 11 Presentation recording (due Nov 23)
Week 13 (Nov 24)	**Group Project Presentation**	Quiz 12

TEAM-BASED LEARNING

This course employs a teaching approach known as Team-Based Learning (TBL). Unlike traditional courses, where students learn mostly individually through a combination of listening to lectures and independent study, TBL emphasizes active learning through preparatory reading and in-class activities. The central features of this approach are:

- ✓ **Teams:** Students are assigned to small teams for the entire semester. Teams are built by the instructor to include students with diverse skills and competencies.
- ✓ **Accountability:** Students are expected to complete reading and assignments for each unit in time.
- ✓ **Application:** While many sessions will have a brief lecture expanding on the material you are responsible for preparing from the reading, the majority of class time will be spent on team activities that apply the course material. This allows students to engage with the material at a deeper level. These activities will also give you practice formulating arguments in the style expected of you in your essays.
- ✓ **Feedback:** Students receive frequent and timely feedback on their work from both the instructor and their peers.

GROUND RULES FOR DISCUSSIONS

These ground rules form a set of expected behaviors for conduct in discussions and lectures. They are intended to foster an intellectual atmosphere where we work together to achieve knowledge. They are also intended to ensure that discussions are spirited without devolving into bickering and to ensure that everyone has an opportunity to be heard.

- ✓ Respect yourself and others (share your viewpoint and allow others to share theirs).
- ✓ Give each other the benefit of the doubt (be charitable).
- ✓ Listen actively and attentively.
- ✓ Keep an open mind. (Expect to learn something new, or to have your views challenged by ideas, questions, and points of view different than your own.)
- ✓ Do not interrupt one another, even when you are excited to respond.

- ✓ Challenge one another but do so respectfully.
- ✓ Allow others (and yourself) to revise or clarify ideas and positions in light of new information.
- ✓ Do not offer opinions without providing evidence of your claim.
- ✓ Try not to make assumptions; ask questions instead.
- ✓ Take responsibility for the quality of the discussion.
- ✓ Build on one another's comments; work toward shared understanding.
- ✓ Do not monopolize discussion.
- ✓ If you notice patterns that are troubling or might be impeding full engagement by others, please speak to me privately during my office hours or via email. All such discussions will be regarded as strictly confidential. If it is not possible to speak to me, feel free to reach out to the department chair, academic advisor, or a trusted mentor.

EVALUATION

Low-stake weekly assignments
They will measure individual learning outcomes on key concepts and theories in the course and provide feedback to students. You are **prohibited** from using generative artificial intelligence (AI) to produce any materials or content related to the assessment task.

- ✓ **Quizzes (1-12) (30%)**
  There will be weekly, open-book, low-stake, individual quizzes. It will assess your understanding of the week's assigned reading material, as well as the understanding of the response paper presentation (see below). They will add up to 120 points throughout the semester; you need 100 points to get a full score.
- ✓ **Annotations & reflections (1-2) (5%)**
  You will read challenging multidisciplinary texts in this course. Two annotation and reflection assignments will help you develop active reading skills.

Response papers
There are eight response papers throughout the semester; **each student will author one of them** and have more limited responsibilities with regard to the others (as explained below). In this assessment, you are **allowed** to use generative artificial intelligence (AI) to **aid** you in any manner (**except for directly generating the final product of any of the following stages**). In addition, you must document, reflect, and evaluate your use of generative AI.

- ✓ **Outline (1-8) (15%, group grade)**
  Each group will draft an outline for **each of the eight** response papers during the lecture.
- ✓ **Response paper (500 words) (1) (15%)**
  Each student will be assigned to author **one** response paper, based on the group outline. The grade will be based not only on the final product, but also the process of drafting, as well as the effective and appropriate usage of assisted technology, such as ChatGPT.
- ✓ **Response paper presentation (1) (5%)**
  Each student will present their response papers **once**. This assesses your ability to communicate effectively. As part of this presentation, you will also reflect on the entire writing process, including the use of assisted technology.
- ✓ **Peer review of response paper (1) (5%)**
  Each student will be assigned to peer-review **one** response paper and presentation.

Final group research project
The research project is broken down into several stages. These stages are designed to help you go through the important formative stages of a research project incrementally. They also help make the research process truly collaborative, which often significantly improves the quality of the final product, because it is enriched by the feedback of your instructor and peers. In this assessment, you are **allowed** to use generative artificial intelligence

(AI) to **aid** you in any manner (**except for directly generating the final product of any of the following stages**). In addition, you must document, reflect, and evaluate your use of generative AI.

- ✓ **Problem specification (2.5%, group grade)**
  Each group will analyze a real-world problem related to course themes, and define the problem, context, and significance.
- ✓ **Research proposal (2.5%, group grade)**
  The research proposal will outline the research and provide a detailed plan, including division of labor among teammates.
- ✓ **Research write-up (10%)**
  Each team member will consolidate and link their research findings in this interim group report. Individual grades will be based on both group performance and the evaluation from the interview with the instructor.
- ✓ **Presentation recording + Q&A (10%, group grade)**
  The group will record their presentation into a short video to be viewed and discussed in the final class.

Rubrics: Detailed rubrics will be provided for each assessment task to guide you on what is expected and how you will be graded. Please refer to these rubrics while working on your tasks.

Group grade policy:
Excellent (100)
Good (90)
Adequate (80)
Developing (70)
Inadequate (60)
Missing the assignment (0%)

Final Grading scale
97 or higher = A+
93-96.9 = A
90-92.9 = A-
87-89.9 = B+
83-86.9 = B
80-82.9 B-
77-79.9 = C+
73-76.9 = C
70-72.9 = C-
67-69.9 = D+
60-66.9 = D
Below 60 = F

WEEKLY RHYTHM

My hope is that, in a world of uncertainty, a predictable weekly schedule can help you consistently learn. There may be one or two exceptions, but this is the general structure of a week in this course.

- ✓ **Friday**: The class convenes to delve into the core concepts of the readings. Each session encompasses:
  - **Quizzes**: A open-book, low-stake assessment to gauge understanding of the material.
  - **Mini lectures and group discussions**: A blend of instructor-led insights and interactive group discussions on key concepts that are required to write the week's response paper.
  - **Response paper outline**: Student groups collaboratively craft a structured blueprint for the week's response paper.
- ✓ **Tuesday**: **Response papers,** encapsulating students' reflections on and analysis of the week's readings, are due **by 10pm**. These papers are pivotal in honing students' analytical and articulation skills.

- ✓ **Wednesday:**
  - **Peer reviews** for the response papers are due **by 3pm.** This exercise not only provides valuable feedback to the authors but also cultivates critical evaluation skills among reviewers.
  - **Interviews** are scheduled among the instructor, the author, and the peer reviewer. It will facilitate constructive feedback, clarification of confusions, and a deeper understanding.
- ✓ **Thursday**
  - The presentation slides are due **by 10pm.**
- ✓ **The following Friday**
  - **Presentations**: Students present the response papers.

POLICIES

Communication
Please check your email at least **once a day** for communication about the course. While I am happy to answer short questions via email, attending my office hours would be best for substantive discussions. I will standardly be available after class meetings, and you can also make an appointment.

Accessibility
If the course delivery, readings, or assessment methods aren't working for you because of linguistic, disability, health, financial or other barriers, please do consult with me. Reasonable accommodation will be made wherever possible.

Attendance
Given the tight schedule of quizzes and presentations (see above), missing too many class meetings without a documented reason is not advisable.

Use of Technology in class
Laptops, tablets, etc. are allowed in class; however, please be mindful that what is on your screen may be distracting to those around you.

Late Work
No late assignments will be accepted unless accompanied by prior arrangement or in exceptional circumstances, such as unforeseeable emergencies.

If you have any concerns at any point throughout the course, I encourage you to email me or come to my office hours to discuss. In general, if a special condition or circumstance in your life may affect your performance, please let me know about it as soon as possible. It will be treated with the strictest confidence. Do not wait until the condition or circumstance is imminent or has already happened before telling me about its impact on you. If something unanticipated occurs, bring it to my attention and we will work out a way of dealing with it together.

Plagiarism
Accurate citation of the materials you've used in your papers is extremely important. The goal of accurate citation is to acknowledge the sources of your thinking, and to make clear when ideas are not your own. This has two purposes. First, to alert your reader as to the intellectual pedigree of your arguments. Second, to demonstrate your engagement with literature. You must cite your sources if you use a direct quote or draw an idea from a paper/book/online source.

Make sure you've listed your sources in a bibliography at the end of your papers. For guidelines on how to construct an academic bibliography, see this. I prefer the APA citation style, but I'm not picky. The only strict style requirement is that direct quotes must be in quotation marks, with author and page number supplied directly after. Here's an example in APA style: It has been said that "philosophy is awesome" (Nado, 2017, p. 106).

Papers will be run through the Turnitin plagiarism detector. If you are found to have plagiarized any portion of your paper, you will automatically receive a score of 0 on that assignment and may be subject to further disciplinary action. Information about student conduct and academic integrity at HKUST can be found here.