

The Hong Kong University of Science and Technology

UG Course Syllabus

[Course Title] Artificial Intelligence Ethics

[Course Code] COMP1944

[No. of Credits] 3-credit

[Any pre-/co-requisites]

Name: [Instructor(s) Name] Prof. Dekai WU

Email: [Your Email Address] de kai@cse.ust.hk

Course Description

AI is disrupting every sphere of our work and lives, bringing unprecedented risks to society. This introductory course surveys the explosive area of artificial intelligence ethics, illuminating relevant AI concepts with no prior background needed. 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. No area remains untouched. Policy think tanks, governments, and tech companies around the world have started paying serious attention to AI ethics. How will human civilization survive the rise of AI? What are the new rules? What are the ethical frameworks needed to avoid extinction? What are engineers' and entrepreneurs' ethical responsibilities? Previous Course Code(s): COMP 4901M

List of Topics

Overview. Fairness, accountability, and transparency in society, AI and machine learning, the impact of AI and automation upon labor and the job market

Information disorder. Misinformation vs disinformation vs malinformation, deepfakes, chatbots, and how they disrupt society

Algorithmic bias, unconscious bias, and inductive bias. Relationships between three different foundational kinds of bias and the serious social consequences

AI ethics methodologies. Survey of approaches to formulating AI ethics methodologies

Weak AI, strong AI, and superintelligence. Contrasts between different senses and levels of "AI". that impact human-machine interaction and society in very different ways

Conscious AI. Artificial mindfulness and its societal impact

Prescriptive vs descriptive AI ethics, and deontological vs consequentialist vs virtue AI ethics. Relates classic ethics philosophy to the problem of AI ethics, and discusses why purely rule-based AI ethics will fail

Artificial moral cognition. Embedding ethics into AIs themselves, and the illusion of explainability

Privacy, safety, security. Surveillance capitalism, identity theft, artificial gossips

AI-driven framing and narratives. The role of social media, recommendation engines, and search engines in computational propaganda and artificial storytellers

Constructive and creative AI. Are machines more creative than humans? Computational creativity and its impact on society and culture

Empathetic AI. Affective computing and artificial intimacy

AGI safety. Is the future of humans extinction, to be pets in a zoo, to upload, or to merge with AIs?

Assessments:

Assessment Task	Contribution to Overall Course grade (%)
In-class questionnaires/quizzes	20%
midterm exam	10%
Written assignments	20%
Group presentation	25%
Course participation	25%
Total	100%

Required Texts and Materials

Artificial Children, 2022, Dekai Wu

[Optional] Additional Resources

Reference books

IEEE Ethically Aligned Design

Artificial Intelligence: A Modern Approach, 4th ed., 2020, by Stuart Russell and Peter Norvig

The Fundamentals of Ethics, 5th ed, 2020, by Russ Shafer-Landau