Overview

Learning Outcomes

- Students demonstrate awareness of the ethical dimension of AI and technologies, able to reason about the issues and communicate the issues to others.
- Students make relevant, valuable original contributions to discussion forums that identify and address ethical issues portrayed in world films.

Module Organization and Timeline

The following segments were covered within weeks 6 to 9 in a typical 16-week class.

Assignment 1: Class discussion and analysis of Russell and Norvig's chapter on Philosophy, Ethics, and Safety of AI (Week 6): Students will read the chapter at home and come to class with a summary of the chapter. Also, they will analyze AI Ethics Principles/guidelines according to Montreal ("The Montréal Declaration for a Responsible Development of Artificial Intelligence") and UNESCO ("Recommendations on the Ethics of Artificial Intelligence").

Assignment 2: Class discussion and analysis of the movie What Happened to Monday? (Week 7): Class discussion on AI and Ethics regarding human life; article analysis of "The Cruel Optimism of Anthropocene Technologies: Suspicion and Fascination of Technology in Okja, What Happened to Monday, and Geostorm."; class discussion on the film "What Happened to Monday?"

Assignment 3: Class discussion and analysis of the K-Pop series My Holo Love (Week 8-9): My Holo Love (South Korea, 2020) has twelve episodes in the series. Students will discuss 3 per class period, watch a few select scenes and discuss ethical issues as well as some positive aspects of having a AI hologram helping with day-to-day activities. There will be a class discussion of the article "Virtual Technology in Netflix K-Drama: Augmented Reality, Hologram, and Artificial Intelligence."

AI Ethics Principles/Guidelines

- UNESCO (https://www.unesco.org/en/artificial-intelligence/recommendation-ethics)

Ten core principles lay out a human-rights centered approach to the Ethics of AI.

1. Proportionality and Do No Harm

The use of AI systems must not go beyond what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such uses.

2. Safety and Security

Unwanted harms (safety risks) as well as vulnerabilities to attack (security risks) should be avoided and addressed by AI actors.

3. Right to Privacy and Data Protection

Privacy must be protected and promoted throughout the AI lifecycle. Adequate data protection

frameworks should also be established.

4. Multi-stakeholder and Adaptive Governance & Collaboration

International law & national sovereignty must be respected in the use of data. Additionally, participation of diverse stakeholders is necessary for inclusive approaches to AI governance.

5. Responsibility and Accountability

AI systems should be auditable and traceable. There should be oversight, impact assessment, audit and due diligence mechanisms in place to avoid conflicts with human rights norms and threats to environmental wellbeing.

6. Transparency and Explainability

The ethical deployment of AI systems depends on their transparency & explainability (T&E). The level of T&E should be appropriate to the context, as there may be tensions between T&E and other principles such as privacy, safety, and security.

7. Human Oversight and Determination

Member States should ensure that AI systems do not displace ultimate human responsibility and accountability.

8. Sustainability

AI technologies should be assessed against their impacts on 'sustainability', understood as a set of constantly evolving goals including those set out in the UN's Sustainable Development Goals.

9. Awareness & Literacy

Public understanding of AI and data should be promoted through open & accessible education, civic engagement, digital skills & AI ethics training, media & information literacy.

10. Fairness and Non-Discrimination

AI actors should promote social justice, fairness, and non-discrimination while taking an inclusive approach to ensure AI's benefits are accessible to all.

- Montreal (https://montrealdeclaration-responsibleai.com/the-declaration/)

1. Well-being

The development and use of artificial intelligence systems (AIS) must permit the growth of the well-being of all sentient beings.

2. Autonomy

AIS must be developed and used while respecting people's autonomy, and with the goal of increasing people's control over their lives and their surroundings.

3. Privacy and Intimacy

Privacy and intimacy must be protected from AIS intrusion and data acquisition and archiving systems (DAAS).

4. Solidarity

The development of AIS must be compatible with maintaining the bonds of solidarity among people and generations.

5. Democratic Participation

AIS must meet intelligibility, justifiability, and accessibility criteria, and must be subjected to democratic scrutiny, debate, and control.

6. Equity

The development and use of AIS must contribute to the creation of a just and equitable society.

7. Diversity Inclusion

The development and use of AIS must be compatible with maintaining social and cultural diversity and must not restrict the scope of lifestyle choices or personal experiences.

8. Prudence

Every person involved in AI development must exercise caution by anticipating, as far as possible, the adverse consequences of AIS use and by taking the appropriate measures to avoid them.

9. Responsibility

The development and use of AIS must not contribute to lessening the responsibility of human beings when decisions must be made.

10. Sustainable Development

The development and use of AIS must be carried out to ensure a strong environmental sustainability of the planet.

Preliminary Bibliography

Primary:

- My Holo Love. Dir. Sang-yeop, Lee and Yoon Jong-Ho. Netflix. Studio Dragon. South Korea, 2020.
- What Happened to Monday? Dir. Tommy Wirkola. Netflix. 2017.

Secondary:

- The Montréal Declaration for a Responsible Development of Artificial Intelligence Foreword Principles Consult the Glossary of the Declaration, montrealdeclaration-responsibleai.com/the-declaration/. Accessed 3 Dec. 2023.
- Nielsen, Esben B., and Gregers Andersen. "The Cruel Optimism of Anthropocene Technologies: Suspicion and Fascination of Technology in Okja, what Happened to Monday, and Geostorm." Journal of Popular Culture, vol. 55, no. 4, 2022, pp. 735-754.
- Park, Jinhee. "Virtual Technology in Netflix K-Drama: Augmented Reality, Hologram, and Artificial Intelligence." International Journal of Communication (Online), vol. 17, 2023, pp. 130.

- "Recommendations on the Ethics of Artificial Intelligence." 23 Nov. 2021, unesdoc.unesco.org/ark:/48223/pf0000381137. Accessed 3 Dec. 2023.
- Russell, Stuart J. and Peter Norvig. Artificial Intelligence: A Modern Approach. 4th. Edition. Englewood Cliffs, N.J., Prentice Hall, 2021.

Assessment Rubric

The student's work will be assessed using the AAC&U Ethical Reasoning VALUE Rubric available at VALUE Rubrics - Ethical Reasoning | AAC&U (aacu.org)

For this course, three evaluation criteria of the ER Value Rubric were utilized: Understand Perspectives/Concepts, Issue Recognition, and Apply Perspectives / Concepts. For convenience, the AACU Value Rubric is also provided below:

ETHICAL REASONING VALUE RUBRIC

for more information, please contact value@aacu.org

Association of American Colleges and Universities

Definition

Ethical Reasoning is reasoning about right and wrong human conduct. It requires students to be able to assess their own ethical values and the social context of problems, recognize ethical issues in a variety of settings, think about how different ethical perspectives might be applied to ethical dilemmas, and consider the ramifications of alternative actions. Students' ethical self-identity evolves as they practice ethical decision-making skills and learn how to describe and analyze positions on ethical issues.

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

	Capstone	Milestones		Benchmark
	4	3	2	1
Ethical Self-Awareness	Student discusses in detail/ analyzes both core beliefs and the origins of the core beliefs and discussion has greater depth and clarity.	Student discusses in detail/ analyzes both core beliefs and the origins of the core beliefs.	Student states both core beliefs and the origins of the core beliefs.	Student states either their core beliefs or articulates the origins of the core beliefs but not both.
Understanding Different Ethical Perspectives/Concepts	Student names the theory or theories, can present the gist of said theory or theories, and accurately explains the details of the theory or theories used.	Student can name the major theory or theories she/he uses, can present the gist of said theory or theories, and attempts to explain the details of the theory or theories used, but has some inaccuracies.	Student can name the major theory she/he uses, and is only able to present the gist of the named theory.	Student only names the major theory she/he uses.
Ethical Issue Recognition	Student can recognize ethical issues when presented in a complex, multilayered (gray) context AND can recognize cross-relationships among the issues.	Student can recognize ethical issues when issues are presented in a complex, multilayered (gray) context OR can grasp cross-relationships among the issues.	Student can recognize basic and obvious ethical issues and grasp (incompletely) the complexities or interrelationships among the issues.	Student can recognize basic and obvious ethical issues but fails to grasp complexity or interrelationships.
Application of Ethical Perspectives/Concepts	Student can independently apply ethical perspectives/ concepts to an ethical question, accurately, and is able to consider full implications of the application.	Student can independently (to a new example) apply ethical perspectives/ concepts to an ethical question, accurately, but does not consider the specific implications of the application.	Student can apply ethical perspectives/ concepts to an ethical question, independently (to a new example) and the application is inaccurate.	Student can apply ethical perspectives/ concepts to an ethical question with support (using examples, in a class, in a group, or a fixed-choice setting) but is unable to apply ethical perspectives/ concepts independently (to a new example.).
Evaluation of Different Ethical Perspectives/Concepts	Student states a position and can state the objections to, assumptions and implications of and can reasonably defend against the objections to, assumptions and implications of different ethical perspectives/ concepts, and the student's defense is adequate and effective.	Student states a position and can state the objections to, assumptions and implications of, and respond to the objections to, assumptions and implications of different ethical perspectives/ concepts, but the student's response is inadequate.	Student states a position and can state the objections to, assumptions and implications of different ethical perspectives/ concepts but does not respond to them (and ultimately objections, assumptions, and implications are compartmentalized by student and do not affect student's position.)	Student states a position but cannot state the objections to and assumptions and limitations of the different perspectives/ concepts.