

Course Title: Ethics in AI: Navigating the Intersection of Technology and Society

Course Description:

This course is designed to explore the ethical implications of artificial intelligence (AI) technologies. This course introduces the ethical, social, and legal implications of artificial intelligence (AI) systems, emphasizing responsible design, development, and deployment of AI technologies. Students will learn about the ethical considerations of AI, including issues of bias, accountability, transparency, privacy, and governance. The goal is to gain an understanding of AI's potential impacts on society and how to address ethical challenges, promote fairness, and ensure AI systems align with human values.

Course Objectives:

Upon completion of this course, students will be able to:

- Understand the ethical principles and frameworks relevant to AI development and deployment.
- Evaluate the social and legal implications of AI systems.
- Apply principles of fairness, accountability, and transparency in AI design and decision-making.
- Develop strategies for addressing biases and promoting diversity, equity, and inclusion in AI technologies.
- Engage in critical reflection on the role of AI in society and personal responsibilities as AI practitioners.

Course Outline:

Week 1: Introduction to Ethics and AI

- Introduction
 - The role of AI in society, including natural language processing, machine learning, computer vision, etc
 - Introduction to ethics, fairness, and accountability
 - Historical and contemporary case studies on AI and social responsibility
- AI ethics: principles and guidelines
- Ethical Frameworks for AI
 - Utilitarianism, deontology, and virtue ethics
 - Rights-based and duty-based ethics
 - Ethical decision-making in AI development and deployment
 - Case studies: ethical dilemmas in AI

Week 2: Fairness and Bias in AI

- Types of bias in AI
- Sources of bias in AI

- Addressing bias in AI: Algorithmic fairness and bias mitigation

Week 3: Privacy and Security in AI

- Risks to privacy and security in AI
- Data Privacy in AI
- Cybersecurity in AI
- Protecting privacy and security in AI

Week 4: Social and Legal Implications of AI

- AI and cross-culture social norm
- AI and discrimination
- Case studies: AI systems with significant social and legal consequences

Week 5: Accountability, and Transparency in AI

- AI explainability and interpretability
- Accountability and responsibility in AI development
- Case studies: AI systems promoting fairness, accountability, and transparency

Week 6: Diversity, Equity, and Inclusion in AI

- Identifying and addressing biases in AI systems
- Promoting diversity in AI research and development
- Creating inclusive AI applications
- Case studies: diversity, equity, and inclusion initiatives in AI organizations

Week 7: AI Safety and Long-term Impacts'

- AI safety research and best practices
- The alignment problem and value alignment
- AI and the common good
- Case studies: AI safety challenges and solutions

Week 8: AI Governance, Policy, and Regulation

- The role of AI in public policy and regulation
- International AI governance and cooperation
- Case studies: AI policy and regulatory initiatives

Week 9: AI for Education

- Overview of AI technologies used in education
- Developing ethical frameworks for AI in education

Week 10: Future Directions

- Personal values and responsibilities as AI practitioners
- Strategies for ethical leadership in AI
- Integrating ethics and responsibility into AI practice
- Developing an action plan for promoting ethics in AI

Assignments:

- Class participation and discussion (20%)
- Case study analysis assignments (40%)
- Final project: Action plan for promoting ethics in AI (40%)

Resources:

Ethical AI Principles by IEEE: <https://ethicsinaction.ieee.org/>

AI Now Institute: <https://ainowinstitute.org/>

AI Ethics Lab: <https://aiethicslab.com/>

The Partnership on AI: <https://www.partnershiponai.org/>

Algorithmic Justice League: <https://www.ajl.org/>

Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) Conference:
<https://www.fatml.org/>