Title of the course: AI and Ethics
1. Credit requirement:(L-T-P: 3-0-0, Credit: 3)
2. Please select the committee for Approval: PGPEC
3. Name of the Dept: CSE
4. Please Specify the Level of the Subject: PG level
5. Whether the subject will be offered as compulsory or elective: Elective
6. Prerequisite(s) for the subject, if any
(Please give the subject numbers and names): AI
7. Course Objective

This course will examine some of the deep issues that have become very relevant due to the increasing use of AI in recent times. We will discuss various aspects ranging from the effects of proliferation of algorithmic decision making, autonomous systems (e.g., autonomous weapons), explainability in machine learning, the question of right balance between regulation and innovation, the adversarial role of AI in information dissemination and the questions of individual rights, fairness and discrimination.

While techies like Ray Kurzweil paint a very rosy picture of the future of AI, philosophers like Nick Bostrom caution us about the emergence of "super intelligence". In fact, visionaries like Elon Musk, Bill Gates and Stephen Hawking have repeatedly urged administrations to factor in ethical and engineering standards in the development and deployment of AI systems that are going to somehow "monitor" the future of humanity.

Some of the glaring questions of current times are – will AI systems replace humans in various job sectors? How can we tackle the problem of biased learning in AI behavior; for instance, reinforcing racial and gender informed discriminations in human decision making. How do we allocate responsibility for accidents/errors caused by AI systems such as autonomous vehicles or medical diagnostics? Can we have AI systems that complement rather than displace human intelligence in order to support a more sustainable and just world?

8. Study Materials
In this course, we will use some textbooks only for building up the fundamental concepts. However, majority of the topics will be covered through lectures on important concepts available in the recently published articles, and presentation of the related papers.
Books:

1. Superintelligence: Paths, Dangers, Strategies Paperback, Nick Bostrom, 2016, OUP.
2. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. Virginia Eubanks, 2018, St. Martin's Press
3. Moral Machines: Teaching Robots Right from Wrong. Wendell Wallach and Colin Allen, 2010, OUP.

9. Syllabus:

a. Introduction to Ethics in AI [2L]

Ideas and concepts related to super intelligence. Why singularity is not a singularity?

b. Autonomy, System Design, Agency, and Liability [3L]

The consequences of human-robot interactions, whose lives should an autonomous car save?, how do law interact with machine learning?

c. Algorithmic Bias [8L]

How big data increases inequality and threatens democracy, principle of accountable algorithms, types of bias in machine learning algorithms, mitigating bias in machine learning algorithms and design of "fair" algorithms.

d. Risk Assessment [4L]

Analysis of the COMPAS Recidivism algorithm, should prison sentences be decided based on crimes not committed yet?, the role of machine learning algorithms in bail and sentencing.

e. Predictive Policing [4L]

Predictive analytics and other analytical techniques for law enforcement, algorithmic methods for (i) predicting crimes, (ii) predicting offenders, (iii) predicting perpetrators' identities, (iv) predicting victims of crime.

f. Credit scoring [3L]

Analytics and machine learning algorithms for creditworthiness for of an individual or an organization, credit risk modeling and credit risk analysis.

g. Ownership, control and access [5L]

Spread of fake news, cognitive hacking, hate and abusive content over social media. Role of AI in spread of such information as well as in mitigation (in the form of computational fact checking).

h. Explainability and Accountability [3L]

Algorithmic transparency, why is transparency is crucial, how can one explain the results of a complex ML algorithm – can there principled methods to achieve this?

i. Case Study: Moral Machines and EU's GDPR [3L]

10. Names of the faculty members of the Department/Centers/School who have the necessary expertise and will be the willing to teach the subject (Minimum two faculty members should be willing to teach the subject)

Animesh Mukherjee, Niloy Ganguly

11. Do the contents of the subject have an overlap with any other subject offered in the Institute?

Related Subjects offered by the Institute:

NONE

a) Approximate percentage of overlap: 0%

b) Reasons for offering the new subject in spite of the overlap:

While AI has become one of the most "glamorous" words in the technology world and students, academicians as well as practitioners are in a race to be at the forefront of this AI movement, there is less attention on the repercussions of this movement. This course is geared toward increasing awareness of the immediate moral and legal repercussions of the presence and possibly dominance of AI in our society. The course is meant to investigate the level of truth in the possibility of AI emulating consciousness, cognition, conation and emotion in an “artificial being” in the future, and if so, then what would be the implications of that possible reality. The main aim is to prepare the students to understand the critical capacity of the rapidly evolving AI technology and its societal implications.

All major CS departments across the world have felt the necessity for such a course. A comprehensive list is here:

Computers and Society at Columbia
Computational Ethics for NLP at Carnegie Mellon
Fairness in Machine Learning at Berkeley
AI - Philosophy, Ethics, and Impact at Stanford
Ethical and Social Issues in AI at Cornell
Ethics and Governance of AI at Harvard Law School
Robots and Society at Georgia Tech