COURSE TITLE	The Ethics of Artificial Intelligence		CREDITS		3
COURSE CODE	GLAE 1202	COURSE CATEGORY	CORE	L-T-P	2-0-0
VERSION	1.0	APPROVAL DETAILS	ACM	LEARNING LEVEL	BTL-2

ASSESSMENT SCHEME

During Semester Assessment (DSA)			Semester End Examination
MSE	DSA Components	Attendance	SEE
20%	25%	5%	50%

Course Description	Artificial intelligence (AI) increasingly executes tasks that previously only humans could do, such as driving cars, diagnosing illnesses, and even selecting job applicants. What is more, in some domains, AI also outperforms humans. For these and other reasons, some people conjecture that AI could have a greater impact on humanity than fire and electricity. But in any case, AI will continue to change and disrupt many areas of our lives, posing intricate social, political, and ethical questions. This series of lectures is an introduction to the ethics of AI. The objectives of the lectures are (1) to equip students with detailed knowledge about applications of AI across several domains, (2) to help students learn moral theories and ethical approaches in relation to AI, and (3) enhance students' analytical skills in the context of ethical challenges more generally. For each lecture, I recommend one article or chapter from recommended book to be read in advance. The lectures themselves will include various opportunities for interaction and discussion.
Course Objectives	The objective of this course is to provide students with a comprehensive understanding of the ethical implications and challenges posed by artificial intelligence (AI). By the end of the course, students will be able to critically analyze and evaluate the ethical considerations surrounding AI development, deployment, and impact on society.
Course Outcome	1. Analyze the fundamental principles of ethics, moral values and their application to the field of artificial intelligence. 2. Evaluate the ethical implications of AI technologies in various domains, such as healthcare, finance, autonomous vehicles, and medicine. 3. Understand the ethical considerations in data collection, privacy, and security in AI systems. 4. Examine the social and economic impacts of AI technologies, including the effects on employment, inequality, and power dynamics. 5. Critically evaluate the ethical frameworks and guidelines proposed for AI development and deployment.

Prerequisites: None

CO, PO AND PSO MAPPING

CO	PO - 1	PO-2	PO-3	PO-4	PO-5	PO-6	PO-7	PO-8	PO-9	P0-10	PO-11	PO-12
CO-1	3	1	-	3	-	-	2	-	-	-	-	-
CO-2	3	3	-	1	3	-	2	2	-	-	-	-
CO-3	3	3	-	2	-	-	1	3	3	-	-	-
CO-4	3	3	-	3	3	-	1	3	3	-	-	-
CO-5	3	3	-	3	-	-	1	-		-	-	-

1: Weakly related, 2: Moderately related and 3: Strongly related

MODULE 1: The Ethics of AI: Introduction and Overview	
The first lecture introduces students to the basics of artificial intelligence and to ethics as an academic discipline. On this basis, the lecture also discusses the aims of AI Ethics more specifically.	**CO-1**
MODULE 2: Intelligent machines' values and human intelligence	
The second lecture explains underlying values of AI, the nature of human intelligence in connection with the nature of AI. The lecture also discusses the topic of augmenting human intelligence with AI.	**CO-2**
MODULE 3: Responsibility	
The third lecture explains how the use of AI led to difficult questions about responsibility and the challenge of so-called 'responsibility gaps', i.e. situations in which AI cause harm but no one seems to be responsible for it.	**CO-2**
MODULE 4: Autonomy and Manipulation	
The fourth lecture explains how AI systems can undermine and promote human autonomy, with a special emphasis on the threat from manipulation.	**CO-2**
MODULE 5: Algorithmic Discrimination	
The fifth lecture focuses on discrimination through algorithms. It explains how and why AI systems can be biased and explores several approaches to fight bias and discrimination.	**CO-3**
MODULE 6: Privacy	

The sixth lecture focuses on the collection of personal data through new technologies such as AI and discusses both the dangers and potential promise of AI in this area	**CO-3**
MODULE 7: Cybersecurity	
The seventh lecture elaborates on the increased risk of cyberattacks with new digital technologies and discusses how AI might not only exacerbate these risks but also be a critical tool in enhancing cybersecurity.	**CO-4**
MODULE 8: Models of personhood and AI	
The eight lecture lays out how AI, and digital technologies more generally, can transform the person as well as all inner world of human. We discuss how usage of AI-technology influence personhood.	**CO-4**
MODULE 9: Explainability and Transparency	
The ninth lecture focuses on a central problem in AI ethics, namely that the best AI systems tend to be the least transparent and least explainable.	**CO-5**
MODULE 10: AI and the Future of Work	
The tenth lecture shows how AI has automated an increasing number of tasks and thereby revolutionised production and work force. But this tendency also creates questions about justice and equality.	**CO-5**
MODULE 11. AI and the Personalisation of Medicine	
This eleventh lecture showcases some of the most impressive advances in AI in the context of healthcare. In the discussion of several case studies, the lecture asks what fundamental values are at stake and how we might need to change our ethical frameworks in response to AI.	**CO-5**
MODULE 12. Future perspective of AI: AGI, superintelligence, strong AI	
This twelth lecture shows how humanity can live with AI in future. The future of AI, especially AGI, superintelligence and strong AI is discussed from teleological point of view.	**CO-5**
TEXT BOOKS	
1.	• Lawrence N.D.The Atomic Human: Understanding Ourselves in the Age of AI. New-Your: Public affairs, 2024. 260 p. • Chalmers, D.J. Reality +: Virtual Worlds and. the Problems of Philosophy. New-York. 2022. p.505 • He, Jianxing, et al. "The practical implementation of artificial intelligence technologies in medicine." Nature medicine 25.1 (2019): 30-36 • Peah Moradi and Karen Levy, "The Future of Work in the Age of AI: Displacement or Risk-Shifting", in M. Dubber, • Pasquale F., and Das S. (eds), The Oxford Handbook of Ethics of AI (OUP, 2020)

	• Mittelstadt, Brent, Chris Russell, and Sandra Wachter. "Explaining explanations in AI." Proceedings of the conference on fairness, accountability, and transparency. 2019 • Manheim, Karl, and Lyric Kaplan. "Artificial intelligence: Risks to privacy and democracy." Yale JL & Tech. 21 (2019): 106 • Kaloudi, Nektaria, and Jingyue Li. "The ai-based cyber threat landscape: A survey." ACM Computing Surveys (CSUR) 53.1 (2020): 1-34 • Bostrom, N Superintelligence: Paths, dangers, strategies. Oxford, Oxford University Press..2016 • Wallach, W.; Allen, C. Moral Machines: Teaching Robots Right from Wrong — USA: Oxford University Press, 2008 • Russell, S. Human Compatible: Artificial Intelligence and the Problem of Control. — New-York : Viking, 2019. 345 p.
REFERENCE BOOKS	
1.	Fazelpour, Sina, and David Danks. "Algorithmic bias: Senses, sources, solutions." Philosophy Compass 16.8 (2021): e12760
2.	Bartmann, Marius. "The Ethics of AI-Powered Climate Nudging—How Much AI Should We Use to Save the Planet?." Sustainability 14.9 (2022): 5153
3	Coeckelbergh, Mark. "Artificial intelligence, responsibility attribution, and a relational justification of explainability." Science and engineering ethics 26.4 (2020): 2051-2068
4	Bostrom N. and Yudkowsky E. "The ethics of artificial intelligence". In W. M. Ramsey and K. Frankish, editors, The Cambridge Handbook of Artificial Intelligence,Cambridge University Press, 2014, p. 316– 334
5.	Boddington P. Towards a Code of Ethics for Artificial Intelligence. Artificial Intelligence: Foundations, Theory, and Algorithms, Springer 2017
E BOOKS	
1.	
MOOC	
1.	