

Branch Name:	M.Sc. (Artificial Intelligence & Data Science)
Program Code:	ET214
Course Title:	Responsible AI
Course Code:	1ET2140105
Pre-requisite Course:	Machine Learning, Data Mining, Cyber Security Fundamentals

Course Objectives:

1. To understand ethical foundations and societal implications of AI systems.
2. To analyze fairness, accountability, and transparency in AI models.
3. To examine AI security threats and privacy-preserving mechanisms.
4. To study governance frameworks and AI regulatory standards.
5. To design secure, fair, and responsible AI solutions.

Teaching and Examination Scheme:

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory (Marks)		Practical (Marks)		Total (Marks)
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
2	-	-	2	50	50	-	-	100

Course Contents:

Unit No	Topic	Total Hours	Weightage (%)
1	Foundations of AI Ethics & Responsible AI: Ethical theories, Principles of Responsible AI (Fairness, Accountability, Transparency), AI lifecycle risks, Human-centered AI, Societal impact	7	25
2	Bias, Fairness & Explainable AI: Types of bias, Algorithmic discrimination, Fairness metrics, Bias mitigation techniques, Explainable AI (XAI), Model interpretability tools	7	25
3	Privacy & AI Security: Data governance, Privacy-preserving ML, Differential Privacy, Federated Learning, Adversarial attacks, Model inversion, Data poisoning, Secure ML lifecycle	7	25
4	AI Governance & Risk Management: AI risk assessment frameworks, AI auditing, Compliance standards, Global AI policies & regulations, Ethical AI implementation roadmap	5	15
5	Case Studies in Responsible & Secure AI: Healthcare AI ethics, Financial AI bias issues, Facial recognition controversies, Autonomous systems safety, Generative AI risks, AI system failure analysis and responsible AI design framework	4	10

Text Books:

1. Stuart Russell & Peter Norvig – Artificial Intelligence: A Modern Approach, Pearson.
2. Mark Coeckelbergh – AI Ethics, MIT Press.
3. Ronald T. Kneusel – Practical Deep Learning Security, O'Reilly.

Reference Books:

1. Cathy O'Neil – Weapons of Math Destruction, Crown Publishing.
2. Barocas, Hardt & Narayanan – Fairness and Machine Learning, Cambridge (Open Access).
3. OECD – OECD AI Principles & Policy Observatory.
4. Solove – Understanding Privacy, Harvard University Press.

Course Learning Outcomes (CLO):

CO	Statement	Bloom's Level
CO1	Explain ethical principles and responsible AI frameworks	L2 – Understand
CO2	Analyze bias, fairness, and explainability issues in AI systems	L4 – Analyze
CO3	Evaluate privacy risks and security vulnerabilities in AI models	L5 – Evaluate
CO4	Apply governance, auditing, and risk management frameworks for AI	L3 – Apply
CO5	Design responsible, fair, and secure AI solutions based on real-world case studies	L6 – Create

Mapping of CLO with PO and PSO

Course	Program Outcomes (POs)												Program Specific Outcomes (PSOs)			
Learning Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CLO1	–	M	–	–	–	H	H	H	L	M	–	M	L	L	L	H
CLO2	–	H	M	M	M	H	M	H	–	M	–	M	M	M	M	H
CLO3	–	M	M	H	H	H	M	H	–	M	M	M	M	H	M	H
CLO4	–	M	H	H	M	H	H	H	M	H	H	M	M	H	M	H
CLO5	–	H	H	H	H	H	H	H	H	H	H	H	H	H	M	H

H:High, M:Medium, L:Low