U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: V**

Branch: CSE(DS)/AI & DS **Duration: 3 hrs.**

Course Code: 23DS5PERAI **Max Marks: 100**

Course: Responsible AI

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Analyze how the design and functionality of Virtual Agents and Embodied Systems differ in human-machine interaction with suitable examples.	*CO2*	*PO2*	**06**
	b)	Describe the key differences between Top-Down and Bottom-Up AI Development approaches. Provide examples to illustrate how each approach is implemented and their impact on AI system design.	*CO1*	*PO1*	**06**
	c)	Illustrate the interplay and continuous nature of the elements in the Responsible Research and Innovation Process.	*CO1*	*PO1*	**08**
		OR			
2	a)	Design a comprehensive approach to create a Responsible AI System that integrates Design for Values and Domain Requirements from software engineering. Highlight the steps and considerations needed to ensure the AI system aligns with ethical principles, stakeholder values, and domain-specific requirements.	*CO3*	*PO4*	**08**
	b)	Justify how the principles of Accountability, Responsibility, and Transparency (ART) create a framework for the ethical development and deployment of AI technologies. Provide specific examples and scenarios to illustrate how each principle can be applied in practice, ensuring that AI systems operate ethically and align with societal values.	*CO1*	*PO1*	**08**
	c)	Explain Beliefs, Desires and Intention (BDI) model	*CO1*	*PO1*	**04**
		UNIT - II			
3	a)	Analyze how the four traditional bioethics principle - autonomy, beneficence, non-maleficence and justice are incorporated into the ethical framework for AI. Also, evaluate how the inclusion of the new principle of explicability enhances this framework.	*CO2*	*PO2*	**10**

		b)	Design an organizational framework that addresses the five risks of being unethical (ethics shopping, ethics blue washing, ethics lobbying, ethics dumping and ethics shirking) while implementing principles of digital ethics in AI-driven projects.	*CO1*	*PO1*	**10**
			OR			
	4	a)	Investigate the ethical risks associated with inconclusive evidence when developing machine learning algorithms.	*CO2*	*PO2*	**05**
		b)	Identify the seven essential factors for ensuring the successful deployment of AI for Social Good (AI4SG) and analyse how each factor contributes to ethical AI development and usage.	*CO2*	*PO2*	**07**
		c)	Analyze how an organization can ensure fairness across different demographic groups by using various algorithmic fairness techniques.	*CO2*	*PO2*	**08**
			UNIT - III			
	5	a)	A City Hospital has experienced several security incidents involving patient data over the past year. The hospital administration is concerned about these breaches and has decided to implement various strategies to ensure patient privacy and data security. As a member of the hospital's technical team, outline the key strategies you would implement to ensure patient privacy and data security.	*CO3*	*PO4*	**06**
		b)	Discuss the Risk Identification and Management techniques that can be used to create a balanced approach that safeguards both the firm and its customers.	*CO1*	*PO1*	**05**
		c)	Identify the essential steps to deploy an AI solution, ensuring it is safe to scale, reviewing the feedback mechanism, and implementing a kill switch and business continuity plan.	*CO1*	*PO1*	**09**
			OR			
	6	a)	Identify the need for the usage of Typology and Illustrate the Applied AI Ethics Typology comprising the ethical principles and stages of algorithmic development.	*CO2*	*PO2*	**10**
		b)	A job recruitment platform uses an AI model to screen and shortlist candidates for interviews. How can the job recruitment platform assess Counterfactual Fairness by determining if an applicant's chances of being shortlisted remain the same when their race or gender attributes are altered.	*CO3*	*PO4*	**10**
			UNIT - IV			
	7	a)	Can Audit Studies help to identify and measure discrimination in a system. Justify your answer with example scenarios.	*CO1*	*PO1*	**08**
		b)	Interpret comprehensive approaches to identify and address fairness concerns in an NLP system.	*CO1*	*PO1*	**06**
		c)	A data scientist is working on developing a machine learning model for a social media platform. What actions will he/she has to take to minimize representational harm.	*CO1*	*PO1*	**06**

			OR			
	8	a)	Evaluate the potential harms associated with implementing search and recommendation systems and propose effective mitigation strategies to address them.	*CO2*	*PO2*	**10**
		b)	Identify the potential mechanisms in ad targeting that could lead to disparities and determine the methods to detect the disparities used by the researchers.	*CO2*	*PO2*	**06**
		c)	Interpret the usage of information flow for Algorithmic Audits.	*CO1*	*PO1*	**04**
			UNIT - V			
	9	a)	Differentiate between local and global interpretable models.	*CO2*	*PO2*	**06**
		b)	As a data scientist at DataCorp, you are tasked with using SHAP to interpret the machine learning model's predictions and ensure fairness in the credit scoring process. Interpret how using SHAP can gain insights into the prediction.	*CO3*	*PO4*	**08**
		c)	Explain the usage of Accumulated Local Effect Plot on a model.	*CO1*	*PO1*	**06**
			OR			
	10	a)	Differentiate between Intrinsic and Post Hoc interpretable models.	*CO2*	*PO2*	**06**
		b)	Consider a healthcare heart disease detection application, how does explainable machine learning transform raw patient data (such as medical records and test results) into interpretable explanations that help doctors make informed decisions about a patient's diagnosis.	*CO3*	*PO4*	**08**
		b)	Discuss the different levels of evaluating interpretability.	*CO1*	*PO1*	**06**

****