विश्वजीवनामृतं ज्ञानम्

Atal Bihari Vajpayee Indian Institute of Information Technology & Management, Gwalior

IT304: Trustworthy Artificial Intelligence

Major Examination (Session 2024–25)

Maximum Time: 3 Hours   Max Marks: 45

Note: All questions are compulsory. Support your answers with case studies or examples.

1. (a) What are the key challenges in building trustworthy AI systems? (b) Explain the role of international AI ethics guidelines.   (7 Marks)

2. (a) Define robustness in AI models. (b) Explain adversarial attacks with one example of image misclassification.   (8 Marks)

3. (a) Differentiate between model interpretability and model transparency. (b) Explain LIME or SHAP technique for explainability.   (8 Marks)

4. (a) Discuss various fairness metrics used in AI evaluation. (b) Suggest methods to mitigate algorithmic bias.   (7 Marks)

5. (a) What are privacy-preserving ML techniques? (b) Explain federated learning with a neat diagram.   (8 Marks)

6. Write short notes on any two: (i) Societal impact of untrustworthy AI (ii) Role of regulations in AI governance (iii) Human–AI collaboration   (7 Marks)