Subject Description Form

Subject Code	DSAI5208
Subject Title	Trustworthy AI Systems and Technologies
Credit Value	3
Level	5
Pre-requisite/ Co-requisite/ Exclusion	**Pre-requisite:** DSAI5205 Introduction to Artificial Intelligence; or DSAI5207 Modern Deep Learning
Objectives	• Provide students with knowledge of the foundations of trustworthy AI systems and technologies. • Provide students with advanced knowledge of a wide spectrum of techniques to establish trust and security for AI-driven systems. • Train students to apply advanced AI techniques to ensure trust and safeguard security in contemporary technological environments.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: a) Have a good understanding of the trust, security, and privacy-related challenges in the AI-enabled systems. b) Have a deep understanding of the techniques and requirements for the AI-enabled biometric systems. c) Critically review and consolidate the skills for establishing trustworthy AI systems and security; and d) Conduct a real-world case study and performance analysis in processing biometric data using advanced AI techniques.
Subject Synopsis/ Indicative Syllabus	**1. Foundations of Trustworthy AI:** Principles of Trustworthy AI, Security and Fairness in AI Systems, Threats to Trustworthy AI like Backdoor Attacks, Adversarial Attacks and Mitigation, Deepfakes, Hyper Realistic Deepfakes with Diffusion Models, Face-Swapping, Audio Deepfakes, AI Ethics, Trust Erosion and Manipulation. **2. Defence Mechanisms for AI-generated Content:** CNN and LLM-based Detectors, Watermarking and Cryptographic Solutions, Blockchain for provenance (e.g., video authenticator), and Fake audio detection via spectrogram analysis and linguistic patterns. Legal and Policy Frameworks. **3. Privacy-Preserving AI** Federated Learning, Differential Privacy and Homomorphic Encryption, Global regulations like the EU's AI Act, Challenges in Enforcement and Jurisdiction. **4. Trustworthy Security using Biometrics:** Introduction to Biometrics, Desirable Properties of Biometrics for Practical Systems, Strengths and Weaknesses of Different Biometric Patterns, Behavioral Biometrics, Privacy Concerns and Social Challenges. **5. Performance Evaluation and Applications:** Error Rates and Performance Plots for Trustworthy Security Systems, Performance Evaluation for Real World Systems and Applications, Presentation Attacks and Detection.

<table>
<tr><td></td><td>6. Multimodal Systems:
Advantage of Multimodal Systems over Unimodal Systems for Trustworthy AI, Multimodal and Score Normalization Techniques, Feature, Score, and Rank Level Fusion.</td></tr>
<tr><td>Teaching/Learning Methodology</td><td>The course material will be delivered as a combination of lectures, tutorials, and small group projects. Students will get familiar with basic concepts and technologies of trustworthy AI systems, biometric data processing, security, and applications.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes</td><td>
<table>
<tr><td rowspan="2">Specific assessment methods/tasks</td><td rowspan="2">% weighting</td><td colspan="4">Intended subject learning outcomes to be assessed</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td></tr>
<tr><td>1. Assignments, Quizzes and Projects</td><td>55%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>2. Final Examination</td><td>45%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Total</td><td>100 %</td><td colspan="4"></td></tr>
</table>
Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:
The examination and assignments are designed to evaluate the students' understanding of trustworthy AI concepts and security. The project, on the other hand, is designed to evaluate the students' practical skills in using advanced AI methods and models to ensure security for real-world problems.
</td></tr>
<tr><td>Student Study Effort Expected</td><td>
<table>
<tr><td>Class contact:</td><td></td></tr>
<tr><td>▪ Lectures / Tutorials / Labs</td><td>39 Hrs.</td></tr>
<tr><td>Other student study effort:</td><td></td></tr>
<tr><td>▪ Assignment, Project, Quizzes, and Examination</td><td>66 Hrs.</td></tr>
<tr><td>Total student study effort</td><td>105 Hrs.</td></tr>
</table>
</td></tr>
<tr><td>Reading List and References</td><td>

Ferhat Ozgur Catak and Murat Kuzlu, Trustworthy AI: From Theory to Practices, ISBN-13: 979-8876639592, Feb. 2024.
Ahmed Banafa, Transformative AI: Responsible, Transparent, and Trustworthy AI systems, River Publishers, ISBN 9788770040198, Apr. 2024.
R. Bolle, J. Connel, S. Pankanti, N. Ratha, and A. Senior, Guide to Biometrics, Springer, ISBN 978-1-4757-4036-3, 2013.
A. Kumar, Iris and Periocular Recognition using Deep Learning, Academic Press, 2024.
IEEE Transactions on Pattern Analysis and Machine Intelligence.
IEEE Transactions on Biometrics, Behavior, and Identity Science.

</td></tr>
</table>