| | |
|---|---|
| **COURSE TITLE** | **CYBER SECURITY FOR AI SYSTEMS** |
| **COURSE CODE** | **01AS0506** |
| **COURSE CREDITS** | **4** |

**Objective:**

1 With the rapid adoption of Artificial Intelligence systems across critical domains, securing AI-driven applications has become essential. This course aims to provide students with advanced theoretical and practical knowledge of cyber security concepts specifically tailored for AI systems, including data security, model security, adversarial attacks, privacy preservation, and secure deployment of AI solutions.

**Course Outcomes:** After completion of this course, student will be able to:

1 Understand fundamental concepts of cyber security and their relevance to AI systems.

2 Identify security threats, vulnerabilities, and attacks targeting AI models and data pipelines.

3 Apply cryptographic and data protection techniques in AI-based applications.

4 Analyze adversarial attacks and defense mechanisms for machine learning models.

5 Design secure and privacy-preserving AI systems following ethical and regulatory guidelines.

**Pre-requisite of course:**Basic knowledge of Artificial Intelligence, Machine Learning, and Computer Networks.

**Teaching and Examination Scheme**

| Theory Hours | Tutorial Hours | Practical Hours | ESE | IA | CSE | Viva | Term Work |
|---|---|---|---|---|---|---|---|
| 3 | 0 | 2 | 50 | 30 | 20 | 25 | 25 |

| Contents : Unit | Topics | Contact Hours |
|---|---|---|
| 1 | **Introduction to Cyber Security & AI Systems**<br>Cyber security basics, CIA Triad, Authentication, Authorization, Threat landscape, AI attack surface, Risk assessment, Security challenges in AI applications | 7 |
| 2 | **Data Security & Privacy in AI**<br>Data leakage, Secure storage, Encryption techniques (AES, RSA), Anonymization, Data poisoning, Inference attacks, Privacy leakage, GDPR, Indian DPDP Act | 8 |
| 3 | **Machine Learning Model Security**<br>ML lifecycle security, Secure data preprocessing, Validation, Model theft, Model inversion, Membership inference, Secure pipelines, Audit logging, Secure deployment | 8 |

DR. TARANNUM ALIMAHMAD BLOCH
**Digitally signed by (Name of HOD)**

DR. RAJENDRASINH BAHADURSINH JADEJA
**Digitally signed by (Name of Dean/ Principal)**

| Contents : Unit | Topics | Contact Hours |
|---|---|---|
| 4 | **Adversarial Attacks on AI Systems**<br>Adversarial examples, FGSM, PGD, Carlini-Wagner attack, Evasion attacks, Backdoor attacks, Black-box vs white-box attacks, Robustness testing | 8 |
| 5 | **Defense Mechanisms & Secure AI Design**<br>Adversarial training, Regularization, Differential privacy, Federated learning, Secure coding practices, Explainable AI, Secure cloud deployment | 6 |
| 6 | **Ethics, Governance & Future of AI Security**<br>Responsible AI, AI governance, Cyber laws in India, Compliance frameworks, Trustworthy AI, Risk assessment, Future trends in AI security | 5 |
| **Total Hours** | | **42** |

**Suggested List of Experiments:**

| Contents : Unit | Topics | Contact Hours |
|---|---|---|
| 1 | **Practical 1**<br>Identify cyber security threats and classify them based on CIA Triad (case study-based) | 2 |
| 2 | **Practical 2**<br>Perform basic risk assessment for an AI system (identify threats, vulnerabilities, impact) | 2 |
| 3 | **Practical 3**<br>Demonstrate encryption and decryption using AES/RSA in Python | 2 |
| 4 | **Practical 4**<br>Perform data anonymization techniques (masking, pseudonymization) | 2 |
| 5 | **Practical 5**<br>Identify privacy risks in a dataset (basic inference attack demonstration) | 2 |
| 6 | **Practical 6**<br>Perform secure data preprocessing (handling missing values, normalization) | 2 |
| 7 | **Practical 7**<br>Train a simple ML model and analyze overfitting and underfitting | 2 |
| 8 | **Practical 8**<br>Demonstrate adversarial attack (FGSM) using pre-built library (demo-based) of audit logging and monitoring for an ML pipeline . | 2 |
| 9 | **Practical 9**<br>Analyze the impact of adversarial noise on model predictions | 2 |
| 10 | **Practical 10**<br>Apply basic defense techniques (input normalization, noise filtering) | 2 |

DR. TARANNUM ALIMAHMAD BLOCH
**Digitally signed by (Name of HOD)**

DR. RAJENDRASINH BAHADURSINH JADEJA
**Digitally signed by (Name of Dean/ Principal)**

**Suggested List of Experiments:**

| Contents : Unit | Topics | Contact Hours |
|---|---|---|
| 11 | **Practical 11**<br>Demonstrate differential privacy using noise addition | 2 |
| 12 | **Practical 12**<br>Analyze ethical and legal issues in AI security using a case study | 2 |
| **Total Hours** | | **24** |

**Textbook :**

1 Machine Learning and Security: Protecting Systems with Data and Algorithms, Clarence Chio & David Freeman, O'Reilly Media, 2018

2 Cyber Security and Cyber Laws, Nina Godbole & Sunit Belapure, Wiley, 2011

3 Adversarial Machine Learning, Anthony D. Joseph, B. Nelson, B. Rubinstein, J. D. Tygar, Cambridge University Press / Springer, 2018

**References:**

1 Artificial Intelligence Security, Artificial Intelligence Security, Andy Patel, Packt Publishing, 2023

2 Privacy-Preserving Machine Learning, Privacy-Preserving Machine Learning, Reza Shokri, MIT Press, 2024

**Suggested Theory Distribution:**

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

| Distribution of Theory for course delivery | | | | | |
|---|---|---|---|---|---|
| **Remember / Knowledge** | **Understand** | **Apply** | **Analyze** | **Evaluate** | **Higher order Thinking / Creative** |
| 15.00 | 20.00 | 25.00 | 20.00 | 10.00 | 10.00 |

**Instructional Method:**

1 Classroom lectures with advanced cybersecurity + AI integration concepts.

2 Case studies on real-world AI attacks (ChatGPT misuse, deepfakes, etc.).

3 Hands-on labs for adversarial ML and privacy techniques.

4 Continuous assessment via quizzes, assignments, and mini-projects..

5 Encouragement of research paper reading and implementation.

**Supplementary Resources:**

1 https://www.coursera.org/learn/ai-security

2 https://www.ibm.com/security/artificial-intelligence

3 https://www.nptel.ac.in

DR. TARANNUM ALIMAHMAD BLOCH
**Digitally signed by (Name of HOD)**

DR. RAJENDRASINH BAHADURSINH JADEJA
**Digitally signed by (Name of Dean/ Principal)**