**Cyber Security in AI and ML (5 Days)**
**By Dr. Vishwanath Rao**

**Introduction to AI and ML Security**:
- Overview of AI and ML technologies
- Importance of security in AI and ML systems
- Distinction between traditional software security and AI/ML security

**Threat Landscape**:
- Common security threats and attacks targeting AI and ML systems
- Adversarial attacks: poisoning, evasion, data integrity attacks
- Privacy risks: data leakage, membership inference attacks
- Model stealing and intellectual property theft

**Adversarial Machine Learning**:
- Understanding adversarial examples
- Techniques for crafting adversarial attacks
- Defense mechanisms against adversarial attacks

**Privacy-Preserving AI and ML**:
- Privacy concerns in AI/ML systems
- Differential privacy and its application to ML
- Federated learning and secure multi-party computation for preserving privacy

**Model Security and Robustness**:
- Model security best practices
- Model explainability and interpretability
- Techniques for improving model robustness and reliability

**Secure Model Deployment**:
- Secure model deployment architectures
- Containerization and isolation for model serving
- Authentication and access control in AI/ML systems

**Data Security and Governance**:
- Data privacy and compliance considerations
- Secure data storage and transmission
- Data governance frameworks for AI/ML projects

**Ethical Considerations**:
- Ethical implications of AI and ML security
- Bias and fairness in AI/ML systems
- Responsible AI principles and guidelines

**Introduction to Secure Development**:

- Overview of software security and its importance
- Introduction to secure development lifecycle (SDLC)
- Understanding the role of Global Product Security (GPS)

**Threat Modeling**:

- Introduction to threat modeling
- Identifying assets, threats, and vulnerabilities
- Techniques for conducting threat modeling exercises

**Secure Coding Practices**:

- Principles of secure coding
- Common vulnerabilities in code (e.g., buffer overflows, injection attacks)
- Secure coding guidelines for different programming languages (e.g., Java, Python, C/C++)

**Authentication and Authorization**:

- Understanding authentication and authorization mechanisms
- Secure implementation of authentication (e.g., password hashing, multi-factor authentication)
- Role-based access control (RBAC) and least privilege principles

**Input Validation and Output Encoding**:

- Importance of input validation and output encoding
- Techniques for sanitizing and validating input data
- Preventing common injection attacks (e.g., SQL injection, XSS)

**Secure Communication**:

- Securing network communication (e.g., HTTPS/TLS)
- Implementing secure APIs and web services
- Transport layer security best practices

**Data Protection**:

- Encryption fundamentals
- Protecting sensitive data at rest and in transit
- Key management and secure storage practices

**Secure Configuration Management**:

- Secure configuration of servers, databases, and applications
- Hardening operating systems and network devices
- Configuration management tools and best practices

**Secure Development Tools and Techniques**:

- Introduction to security testing tools (e.g., static analysis, dynamic analysis)
- Code review best practices
- Automated security testing and continuous integration (CI) pipelines

**Secure Deployment and Operations**:

- Secure deployment strategies (e.g., container security, serverless security)
- Monitoring and logging for security incidents
- Incident response and handling security breaches

**Security Awareness and Training**:

- Importance of security awareness for developers
- Techniques for promoting a security culture within development teams
- Continuous learning and staying updated on security trends

- **Introduction to Risk Management Framework (RMF)**:
  - Overview of NIST RMF and its importance in cybersecurity
  - Evolution of RMF and its relation to NIST Special Publication (SP) 800-3
- 
- **Key Concepts and Principles**:
  - Understanding risk management concepts (e.g., risk, threat, vulnerability)
  - Principles of risk management according to NIST SP 800-37
- 
- **RMF Roles and Responsibilities**:
  - Roles and responsibilities of key stakeholders in the RMF process
  - Responsibilities of the Information System Owner (ISO), Authorizing Official (AO), Security Control Assessor (SCA), and others
- **RMF Process Overview**:
  - Detailed explanation of the six steps in the RMF process:
    - Prepare
    - Categorize
    - Select
    - Implement
    - Assess
    - Authorize
    - Monitor
- **Step 1: Prepare**:
  - Establishing the context for the RMF process

    - Developing the risk management strategy and policies
- **Step 2: Categorize**:
    - Identifying information systems and assets
    - Assigning impact levels based on FIPS 199
- **Step 3: Select**:
    - Selecting security controls based on NIST SP 800-53
    - Tailoring security controls for specific organizational needs
- **Step 4: Implement**:
    - Implementing selected security controls
    - Documentation requirements for control implementation
- **Step 5: Assess**:
    - Conducting security control assessments
    - Performing security testing and evaluation
- **Step 6: Authorize**:
    - Reviewing security assessment results
    - Making authorization decisions
    - Preparing the Authorization Package
- **Step 7: Monitor**:
    - Continuous monitoring of security controls and the information system
    - Reporting security incidents and changes to the system
- **Integration with Other Frameworks and Standards**:
    - Relationship between RMF and other cybersecurity frameworks (e.g., ISO 27001, COBIT)
    - Compliance with regulatory requirements (e.g., FISMA, HIPAA, GDPR)

- **Introduction to CISA and CSA**:
    - Overview of CISA and its role in protecting critical infrastructure in the United States
    - Introduction to the Cloud Security Alliance (CSA) and its mission to promote best practices for secure cloud computing
- **Cloud Computing Fundamentals**:
    - Overview of cloud computing models (IaaS, PaaS, SaaS)
    - Key characteristics of cloud computing (on-demand self-service, broad network access, etc.)
- **CISA's Role in Cloud Security**:
    - CISA guidance and resources for securing cloud environments
    - CISA's involvement in promoting cloud security standards and best practices
- **CSA's Cloud Security Guidance**:
    - Overview of CSA's Cloud Controls Matrix (CCM) and Security, Trust & Assurance Registry (STAR)

    - CSA's research initiatives and publications on cloud security best practices
- **Cloud Security Threats and Risks**:
    - Common security threats and risks associated with cloud computing
    - Understanding shared responsibility model and its implications for security
- **Securing Cloud Infrastructure**:
    - Best practices for securing cloud infrastructure (e.g., virtual networks, compute instances)
    - Identity and access management (IAM) in the cloud
- **Data Protection in the Cloud**:
    - Data encryption and key management
    - Data privacy and compliance considerations (e.g., GDPR, CCPA)
- **Securing Cloud Applications**:
    - Application security considerations in the cloud
    - Secure development practices for cloud-native applications
- **Cloud Incident Response and Forensics**:
    - Incident response planning for cloud environments
    - Forensic investigation techniques in the cloud
- **Cloud Compliance and Governance**:
    - Compliance frameworks and regulations relevant to cloud computing (e.g., NIST, PCI DSS)
    - Cloud governance best practices

- **Introduction to AI Security**:
    - Overview of artificial intelligence (AI) and its applications
    - Understanding the security implications of AI technologies
- **OWASP Top 10 AI Security Risks**:
    - Identification of the top security risks specific to AI systems
    - Examples of vulnerabilities and threats in AI applications
- **Adversarial Machine Learning**:
    - Understanding adversarial attacks on machine learning models
    - Techniques for crafting and defending against adversarial attacks
- **Privacy and Ethical Considerations**:
    - Privacy risks associated with AI systems
    - Ethical considerations in AI development and deployment
- **Data Security in AI Systems**:
    - Importance of data security in AI applications
    - Techniques for securing training data and model outputs
- **Model Explainability and Transparency**:
    - Importance of model explainability in AI systems
    - Techniques for improving the interpretability and transparency of AI models

- **Secure Development Practices for AI**:
  - Secure coding practices for AI applications
  - Integration of security into the AI development lifecycle
- **Secure Deployment and Operations**:
  - Security considerations for deploying AI models in production
  - Monitoring and logging for security incidents in AI systems

**Introduction to Threat Detection and Prevention**:
- Overview of cybersecurity threats and their impact
- Importance of proactive threat detection and prevention measures

**Threat Intelligence**:
- Understanding threat intelligence sources and feeds
- Techniques for collecting, analyzing, and applying threat intelligence

**Vulnerability Management**:
- Identifying and assessing vulnerabilities in systems and applications
- Strategies for prioritizing and remedying vulnerabilities

**Intrusion Detection Systems (IDS)**:
- Introduction to IDS and their role in threat detection
- Types of IDS (e.g., network-based, host-based) and their deployment models

**Intrusion Prevention Systems (IPS)**:
- Overview of IPS and their capabilities
- Techniques for preventing and mitigating intrusions in real-time

**Security Information and Event Management (SIEM)**:
- Understanding SIEM and its role in centralized log management and analysis
- Using SIEM for threat detection, incident response, and compliance reporting

**Behavioral Analysis**:
- Behavioral analysis techniques for detecting abnormal activities and threats

- Machine learning and AI-driven approaches to behavioral analysis

**Network Traffic Analysis**:
- Analyzing network traffic patterns for anomalies and suspicious activities
- Tools and technologies for network traffic analysis

**Endpoint Detection and Response (EDR)**:
- Overview of EDR solutions and their capabilities
- Detecting and responding to threats at the endpoint level

**Web Application Firewall (WAF)**:
- Introduction to WAF and its role in protecting web applications from attacks
- Configuration and tuning of WAF rules for effective threat prevention

**Email Security**:
- Common email-based threats (e.g., phishing, spam, malware)
- Techniques for detecting and preventing email-based attacks

**Cloud Security Monitoring**:
- Monitoring security events and activities in cloud environments
- Best practices for securing cloud workloads and data

**Incident Response and Threat Hunting**:
- Incident response process and procedures
- Threat hunting techniques for proactively identifying and mitigating threats

**Introduction to Data Security in AI and ML**:
- Overview of data security challenges in AI and machine learning projects
- Importance of data protection for maintaining confidentiality, integrity, and availability

**Data Privacy Regulations and Compliance**:

- Understanding data privacy regulations (e.g., GDPR, CCPA) and their impact on AI/ML projects
- Compliance requirements for handling sensitive data in AI/ML systems

**Data Classification and Sensitivity**:
- Techniques for classifying data based on sensitivity and confidentiality levels
- Data labeling and metadata management for tracking data sensitivity

**Data Collection and Acquisition**:
- Best practices for securely collecting and acquiring training data
- Data provenance and traceability to ensure data integrity and authenticity

**Data Storage and Encryption**:
- Secure storage solutions for AI/ML datasets
- Encryption techniques for protecting data at rest (e.g., full disk encryption, database encryption)

**Data Transmission and Network Security**:
- Secure data transmission protocols for transferring data between systems and environments
- Network security measures to protect data in transit (e.g., VPNs, TLS/SSL)

**Data Masking and Anonymization**:
- Techniques for masking and anonymizing sensitive data in AI/ML datasets
- Preserving data utility while protecting individual privacy

**Access Control and Authorization**:
- Role-based access control (RBAC) for controlling access to AI/ML data
- Authorization mechanisms to enforce data access policies and permissions

**Data Governance and Compliance Monitoring**:
- Implementing data governance frameworks to ensure compliance with regulations and internal policies
- Continuous monitoring of data security controls and compliance status

**Secure Data Sharing and Collaboration**:

- Secure mechanisms for sharing and collaborating on AI/ML datasets
- Data sharing agreements and access controls for external collaborators

**Secure Model Training and Development**:

- Securing data pipelines and environments for model training
- Techniques for protecting data during model development and experimentation

**Data Security Testing and Validation**:

- Techniques for testing and validating data security controls in AI/ML systems
- Auditing and logging for tracking data access and usage

**Incident Response and Data Breach Management**:

- Incident response procedures for addressing data breaches and security incidents
- Data breach notification requirements and mitigation strategies

**Introduction to AI Infrastructure Security**:

- Overview of AI infrastructure components (e.g., servers, storage, networking)
- Importance of securing AI infrastructure for protecting sensitive data and intellectual property

**Threat Landscape for AI Infrastructure**:

- Common security threats targeting AI infrastructure (e.g., unauthorized access, data breaches, denial-of-service attacks)
- Understanding the unique risks posed by AI-specific attacks (e.g., model poisoning, evasion attacks)

**Security Architecture for AI Infrastructure**:

- Design principles for building secure AI infrastructure
- Segmentation and isolation of AI workloads to minimize attack surface

**Identity and Access Management (IAM)**:

- Role-based access control (RBAC) for controlling access to AI resources
- Multi-factor authentication (MFA) and strong authentication mechanisms

**Network Security for AI Infrastructure**:

- Network segmentation and micro-segmentation to isolate AI workloads
- Firewall and intrusion detection/prevention systems (IDS/IPS) for monitoring and controlling network traffic

**Data Security in AI Infrastructure**:

- Secure data storage solutions for AI datasets and models
- Encryption techniques for protecting data at rest and in transit

**Secure AI Development Environments**:

- Securing development environments and tools used for AI model training and testing
- Best practices for securing AI development pipelines

**Container Security for AI Workloads**:

- Securing containers and container orchestration platforms (e.g., Kubernetes)
- Container image security and vulnerability scanning

**Cloud Security for AI Infrastructure**:

- Securing AI workloads deployed in cloud environments
- Best practices for cloud security posture management (CSPM)

**Secure Model Deployment and Serving**:

- Security considerations for deploying AI models into production environments
- Techniques for securing AI model serving endpoints

**Continuous Security Monitoring and Incident Response**:

- Implementing security monitoring and logging for AI infrastructure

- Incident response procedures for addressing security incidents in AI systems

**Compliance and Governance**:
- Compliance requirements for AI infrastructure (e.g., GDPR, HIPAA)
- Implementing governance frameworks to ensure compliance and risk management

**Security Testing and Validation**:
- Techniques for testing and validating the security of AI infrastructure
- Penetration testing and vulnerability assessment for identifying weaknesses

**Emerging Technologies and Future Trends**:
- Emerging technologies for enhancing AI infrastructure security (e.g., homomorphic encryption, secure enclaves)
- Future trends and challenges in securing AI infrastructure

**Introduction to GDPR**:

Overview of the GDPR and its objectives
Scope of the regulation and its applicability to organizations handling personal data of EU residents

**Key Principles of GDPR**:

Data protection principles, including lawfulness, fairness, and transparency
Purpose limitation, data minimization, and storage limitation principles

**Individual Rights under GDPR**:

Rights of data subjects, including the right to access, rectification, erasure, and data portability
Understanding the right to be forgotten and its implications

**Data Processing Requirements**:

Requirements for lawful processing of personal data
Conditions for obtaining valid consent

**Data Protection Impact Assessments (DPIA)**:

Understanding DPIA requirements and when they are necessary
Conducting DPIAs and documenting the process

**Data Breach Notification**:
Requirements for data breach notification under GDPR
Timelines and procedures for reporting data breaches to supervisory authorities and affected individuals

**Data Protection Officer (DPO) Role**:
Responsibilities and qualifications of the DPO under GDPR
Role of the DPO in ensuring compliance with GDPR requirements

**Cross-Border Data Transfers**:
Legal mechanisms for transferring personal data outside the EU
Understanding the requirements for adequacy decisions, standard contractual clauses, and binding corporate rules

**GDPR Enforcement and Penalties**:
Powers and responsibilities of supervisory authorities (e.g., the European Data Protection Board)
Administrative fines and penalties for non-compliance with GDPR

**Practical Compliance Strategies**:
Developing GDPR compliance programs and policies
Implementing technical and organizational measures to ensure data protection

**Case Studies and Practical Examples**:
Real-world examples of GDPR compliance challenges and solutions
Analysis of GDPR enforcement actions and fines

**Emerging Trends and Future Developments**:
Emerging trends in data protection and privacy regulation
Future developments in EU data protection law post-GDPR

**Introduction to LLMs and Security**:
Overview of Large Language Models (LLMs) and their applications
Introduction to security risks associated with LLMs

**Injection Attacks**:
- Potential risks of injecting biased or malicious input data into LLMs
- Examples of injection attacks on LLMs and their impact on model behavior

**Broken Authentication**:
- Risks related to unauthorized access to LLM training data or models
- Best practices for securing authentication mechanisms in LLM environments

**Sensitive Data Exposure**:
- Risks of exposing sensitive information through LLM outputs
- Techniques for protecting sensitive data during LLM training and inference

**XML External Entities (XXE)**:
- Risks of XXE attacks in LLM environments processing XML or structured data
- Mitigation strategies for preventing XXE vulnerabilities in LLM systems

**Broken Access Control**:
- Risks associated with inadequate access controls on LLM training data or model parameters
- Implementing robust access control mechanisms for LLM environments

**Security Misconfiguration**:
- Risks stemming from misconfigurations in LLM deployment environments
- Best practices for securing LLM configurations and settings

**Cross-Site Scripting (XSS)**:
- Risks of XSS attacks in LLM-based web applications or interfaces
- Techniques for preventing XSS vulnerabilities in LLM systems

**Insecure Deserialization**:
- Risks associated with insecure deserialization of data in LLM environments
- Implementing secure deserialization practices for LLM systems

**Using Components with Known Vulnerabilities**:
- Risks of using vulnerable components or libraries in LLM frameworks or

dependencies
Strategies for identifying and mitigating vulnerabilities in LLM environments

**Insufficient Logging and Monitoring**:
Risks related to insufficient logging and monitoring of LLM activities
Implementing robust logging and monitoring solutions for detecting and responding to security incidents

**Case Studies and Practical Examples**:
Real-world examples of security risks and incidents involving LLMs
Hands-on exercises and simulations to reinforce security concepts and practices

**Future Trends and Emerging Technologies**:
Emerging trends in LLM security research and development
Implications of new technologies (e.g., federated learning, differential privacy) on LLM security

**Introduction to RAG Applications and AI Agents**:
Overview of RAG applications and AI agents and their role in risk management and assurance processes.
Introduction to the security challenges associated with these systems.

**Data Privacy and Confidentiality Risks**:
Risks related to the handling of sensitive data by RAG applications and AI agents.
Techniques for protecting data privacy and ensuring confidentiality in these systems.

**Data Integrity Risks**:
Risks of data manipulation or tampering within RAG applications or by AI agents.
Methods for ensuring data integrity and preventing unauthorized modifications.

**Authentication and Authorization Risks**:
Risks associated with inadequate authentication and authorization mechanisms in RAG applications.
Best practices for implementing strong authentication and access control

measures.

**Third-Party and Supply Chain Risks**:
Risks arising from the integration of third-party components or dependencies in RAG applications and AI agents.
Strategies for managing third-party risks and ensuring the security of the supply chain.

**Model Security Risks**:
Risks associated with vulnerabilities in AI models used by RAG applications and agents.
Techniques for securing AI models and mitigating risks such as adversarial attacks and model poisoning.

**Compliance and Regulatory Risks**:
Risks related to non-compliance with regulatory requirements and industry standards.
Strategies for ensuring compliance with relevant regulations and standards (e.g., GDPR, financial regulations).

**Insider Threats**:
Risks posed by malicious insiders or negligent employees within organizations using RAG applications and AI agents.
Methods for detecting and mitigating insider threats.

**Social Engineering Attacks**:
Risks of social engineering attacks targeting users or administrators of RAG applications and AI agents.
Techniques for raising awareness and preventing social engineering attacks.

**Incident Response and Crisis Management**:
Strategies and procedures for responding to security incidents and crises involving RAG applications and AI agents.
Developing incident response plans and conducting tabletop exercises.

**Security Testing and Assurance**:
Techniques for assessing the security of RAG applications and AI agents through penetration testing, vulnerability assessments, and security audits.
Ensuring ongoing security assurance through regular testing and

monitoring.

**Case Studies and Practical Examples**:
Real-world examples of security incidents and breaches involving RAG applications and AI agents.
Hands-on exercises and simulations to reinforce security concepts and practices.

**Future Trends and Emerging Technologies**:
Emerging trends in RAG applications and AI security.
Implications of new technologies (e.g., blockchain, federated learning) on security risks and mitigation strategies.