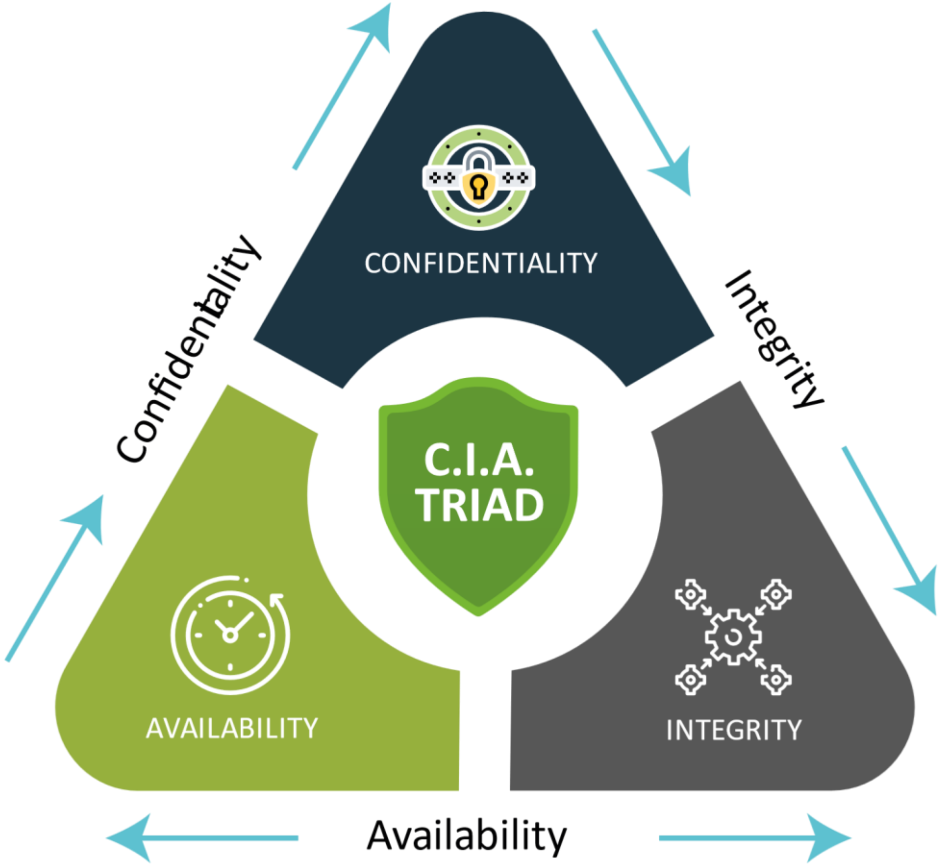
BASICS OF CYBERSECURITY PRINCIPLES AND CONCEPTS

THREAT VECTORS AND TYPES OF CYBERATTACKS.
SECURITY POLICIES AND BEST PRACTICES

CYBERSECURITY PRINCIPLES(1)



- CIA (Confidentiality, Integrity, Availability)
- Confidentiality ensures that only authorized individuals have access to information and resources
  - o Snooping
  - o Dumpster diving
  - o Eavesdropping
  - o Wiretapping
  - o Social engineering.

CYBERSECURITY PRINCIPLES (2)

- Integrity. There aren’t any unauthorized changes to information
- Integrity attacks:
  - Unauthorized modification of information
  - Impersonation attacks
  - Man-in-the-middle (MitM) attacks
  - Replay attacks

CYBERSECURITY PRINCIPLES (3)

- Availability. Information and systems remain available to authorized users when needed.

- Risks:
    - Denial-of-service attacks
    - Power outages
    - Hardware failures
    - Destruction of equipment
    - Service outages

AUTHENTICATION AND AUTHORIZATION

- Access control:
  - Identification. An individual makes a claim about their identity
  - Authentication. An individual proves their identity to the satisfaction of the access control system
  - Authorization.The access control system also needs to be satisfied that you are allowed to access the system

ATTACK AND THREAT VECTORS

- An attack vector is a method of gaining unauthorized access to a network or computer system.
- An attack surface is the total number of attack vectors an attacker can use to manipulate a network or computer system or extract data.
- Threat vector can be used interchangeably with attack vector and generally describes the potential ways a hacker can gain access to data or other confidential information.

https://www.upguard.com/blog/attack-vector

THREAT VECTORS IN AI PROJECTS

- **Social Engineering**: Phishing, baiting
- **Insider Threats**: Employee misuse or negligence
- **External Threats**: Hackers, malware, DDoS
- **Supply Chain Attacks**: Vulnerabilities in third-party libraries or APIs
- **AI-Specific Threats**: Model poisoning, adversarial attacks

SOCIAL ENGINEERING

- Where is the danger and why is it so effective?
  - Authority and trust
  - Intimidation
  - Consensus and social proof
  - Scarcity
  - Urgency
  - Familiarity and liking

TYPES OF CYBERATTACKS

Phishing: Deceptive emails to steal information

Malware: Viruses, worms, ransomware

DDoS Attacks: Overloading systems to deny service

Data Breaches: Unauthorized access to sensitive information

Zero-Day Exploits: Attacks on unknown vulnerabilities

Adversarial AI Attacks: Manipulating models to give incorrect outputs

SECURITY POLICIES FOR AI PROJECTS

- Data Governance: Define who owns, accesses, and manages data
- Access Control: Role-based permissions for sensitive data
- Incident Response Plans: Preparedness for breaches and attacks
- Regular Audits: Ensure compliance and detect vulnerabilities
- Vendor Assessment: Vet third-party services and tools

BEST PRACTICES IN AI PROJECT MANAGEMENT

- Secure Development Lifecycle (SDLC): Build security into AI development phases
- Data Encryption: Protect data at rest and in transit
- AI Model Security: Validate input data and monitor for anomalies
- Ethical AI Practices: Ensure fairness and accountability
- Continuous Monitoring: Real-time threat detection and response