Threat Intel Report on

Mitre Atlas Framework

BY

Team SafeNet

Mayur Pawar - 2033

Rahul Gouda - 2069

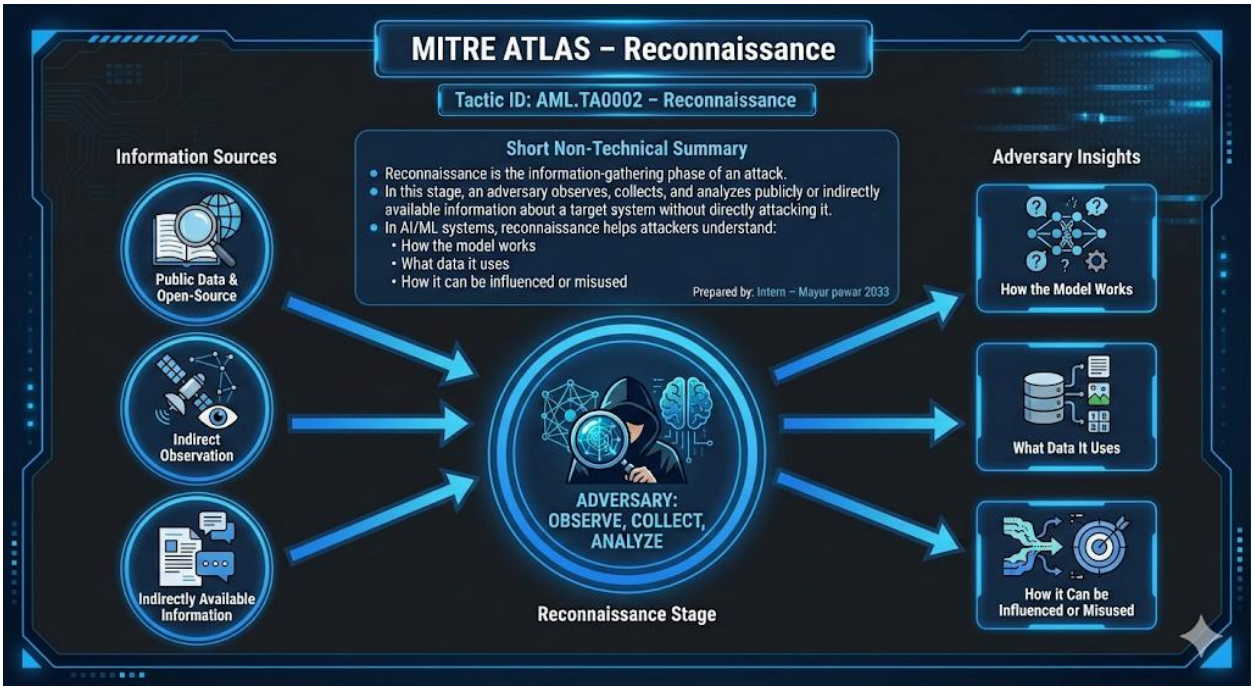
Janhavi Pandey – 2032

Sanika Jankar – 2038

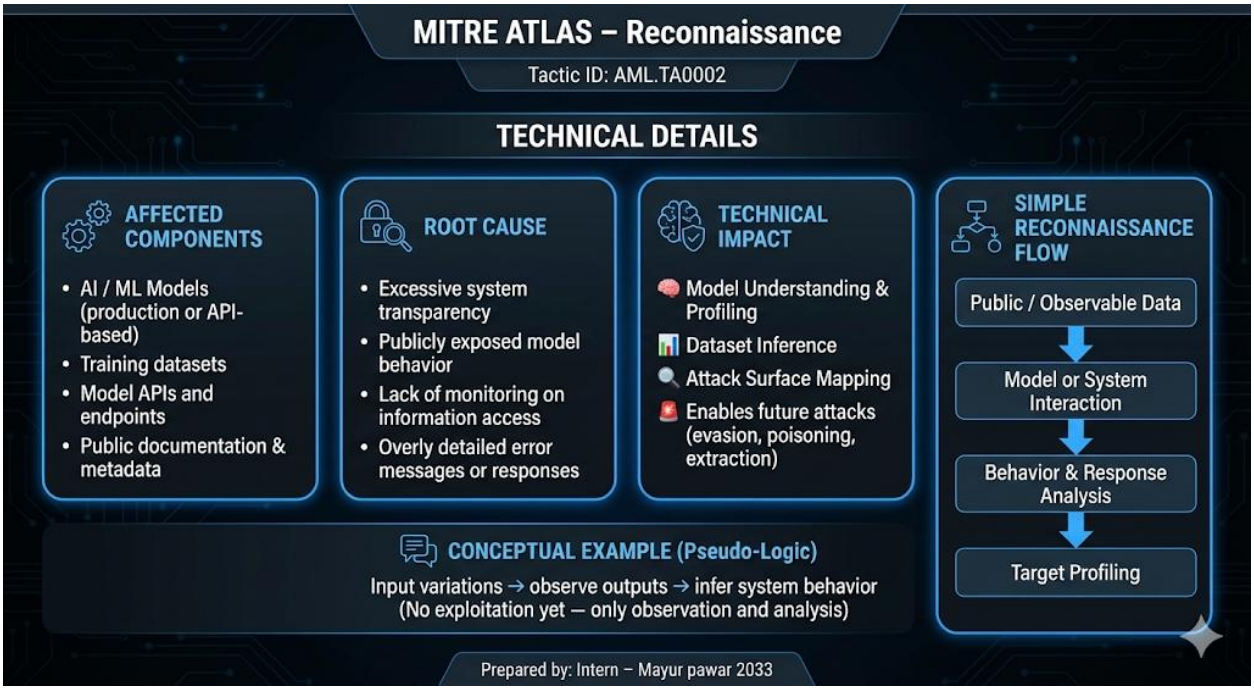
Shruti Rane - 2066

1. Reconnaissance

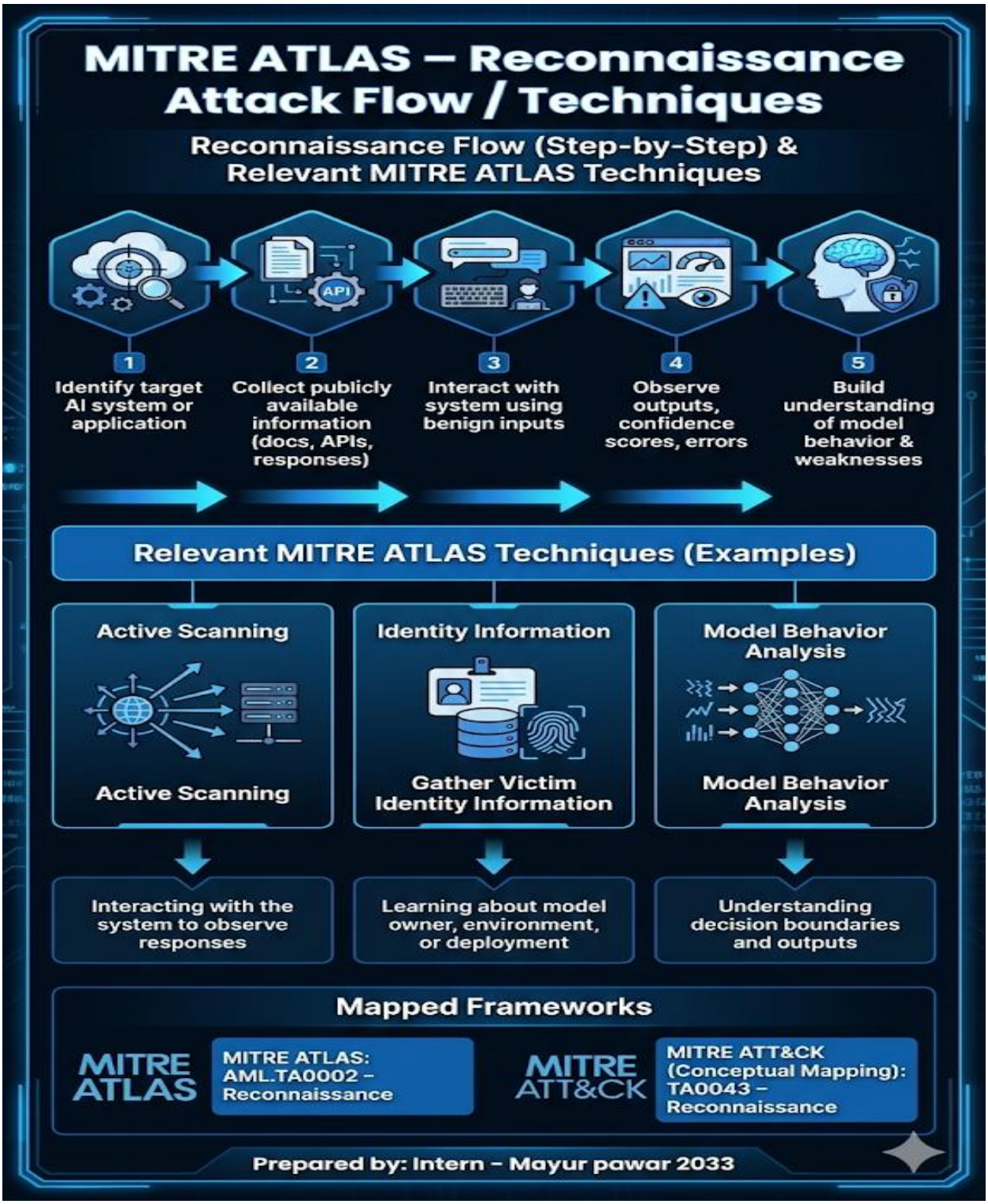
Overview



Technical Details:

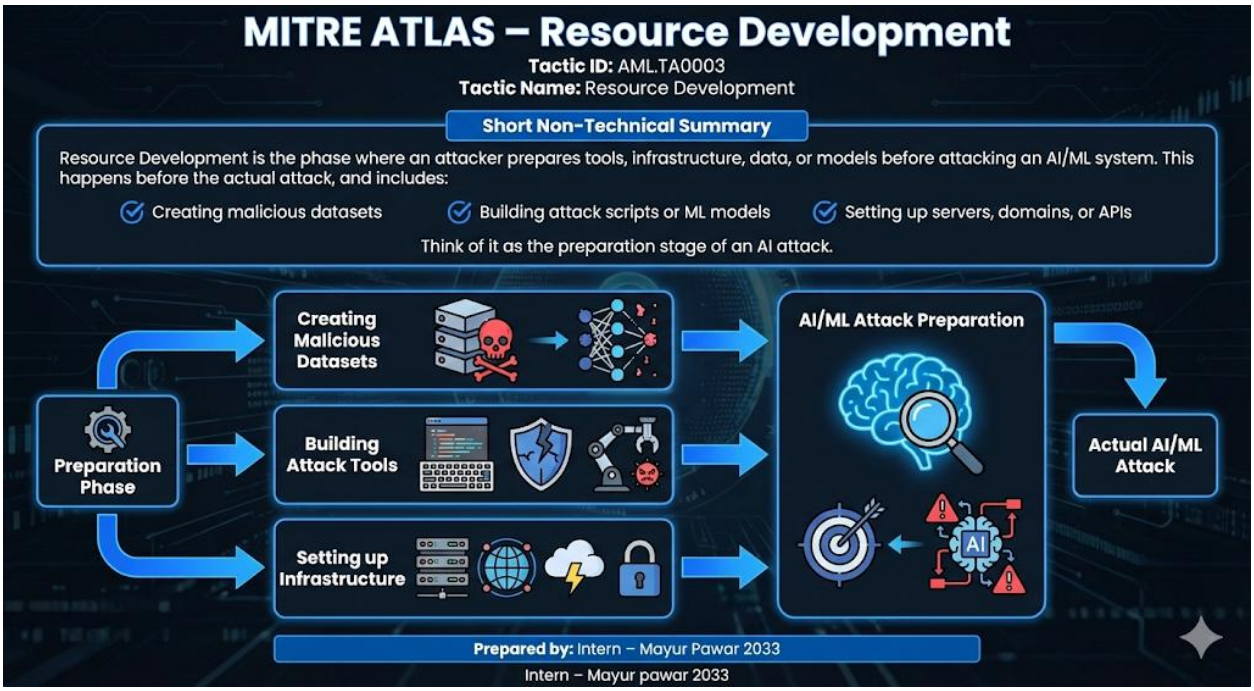


Attack Flow

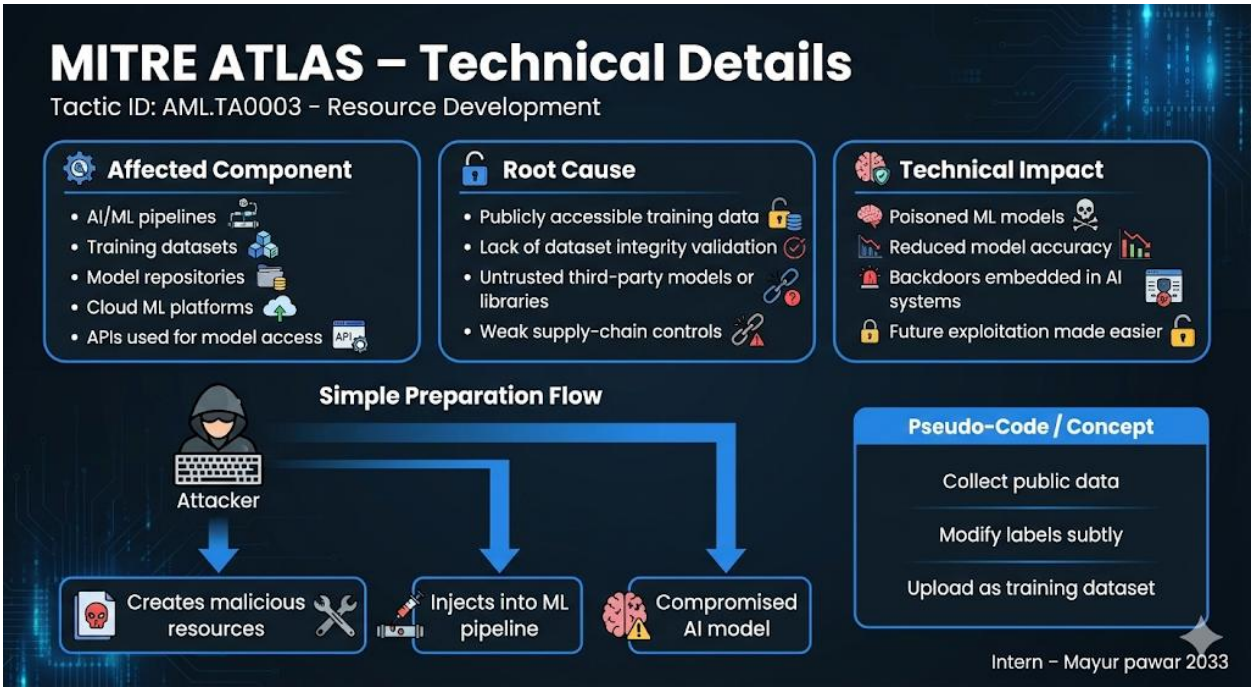

MITRE ATLAS – Reconnaissance Attack Flow / Techniques
Reconnaissance Flow (Step-by-Step) & Relevant MITRE ATLAS Techniques
API
1
Identify target AI system or application
2
Collect publicly available information (docs, APIs, responses)
3
Interact with system using benign inputs
4
Observe outputs, confidence scores, errors
5
Build understanding of model behavior & weaknesses
Relevant MITRE ATLAS Techniques (Examples)
Active Scanning
Active Scanning
Identity Information
Gather Victim Identity Information
Model Behavior Analysis
Model Behavior Analysis
Interacting with the system to observe responses
Learning about model owner, environment, or deployment
Understanding decision boundaries and outputs
Mapped Frameworks
MITRE ATLAS
MITRE ATLAS: AML.TA0002 – Reconnaissance
MITRE ATT&CK
MITRE ATT&CK (Conceptual Mapping): TA0043 – Reconnaissance
Prepared by: Intern – Mayur pawar 2033

2. Resource Development

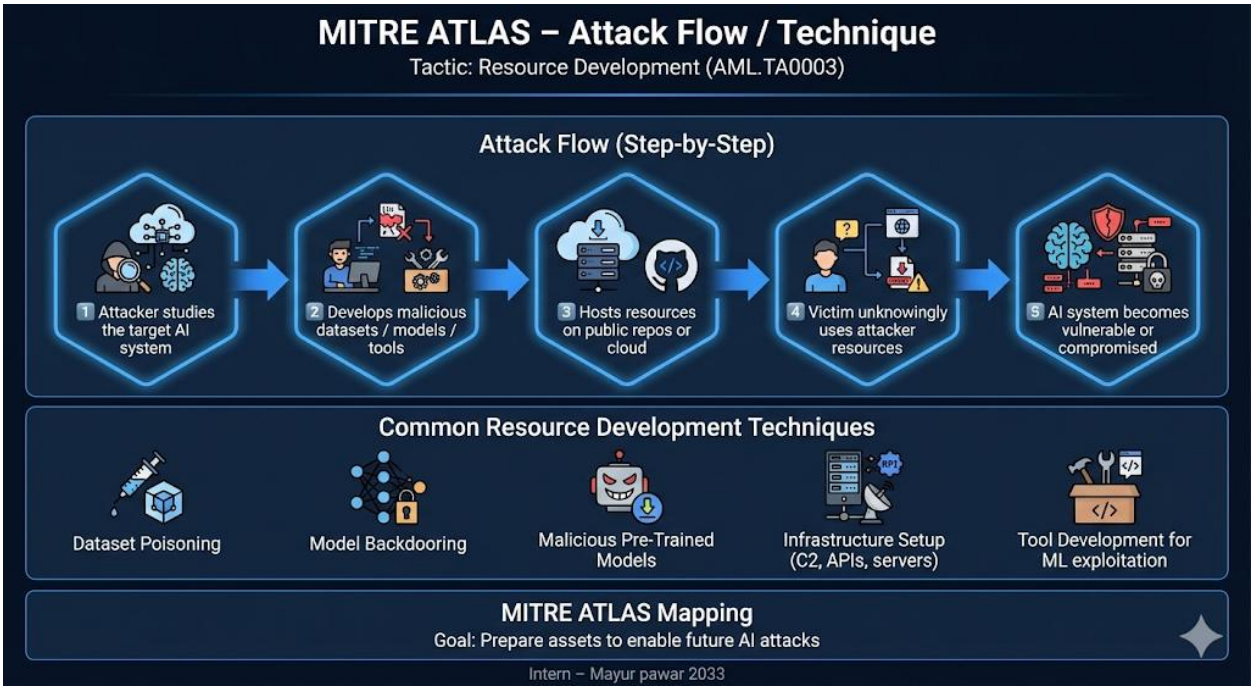
Overview



Technical Details:

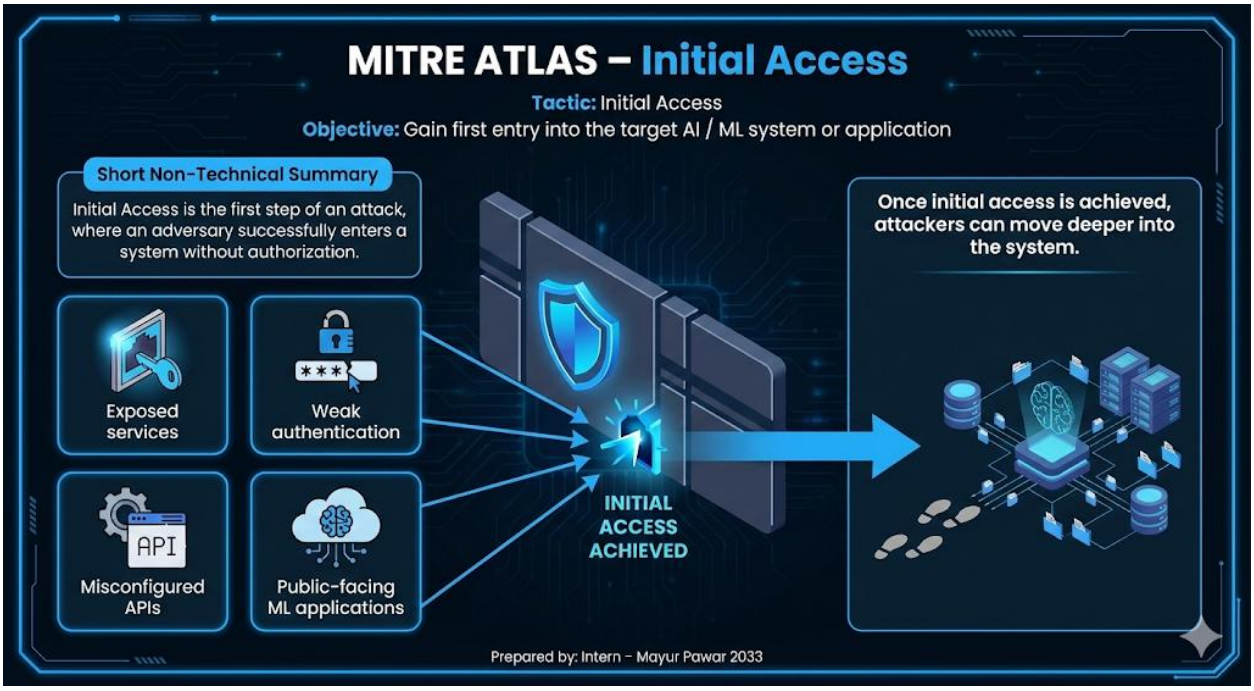


Attack Flow:

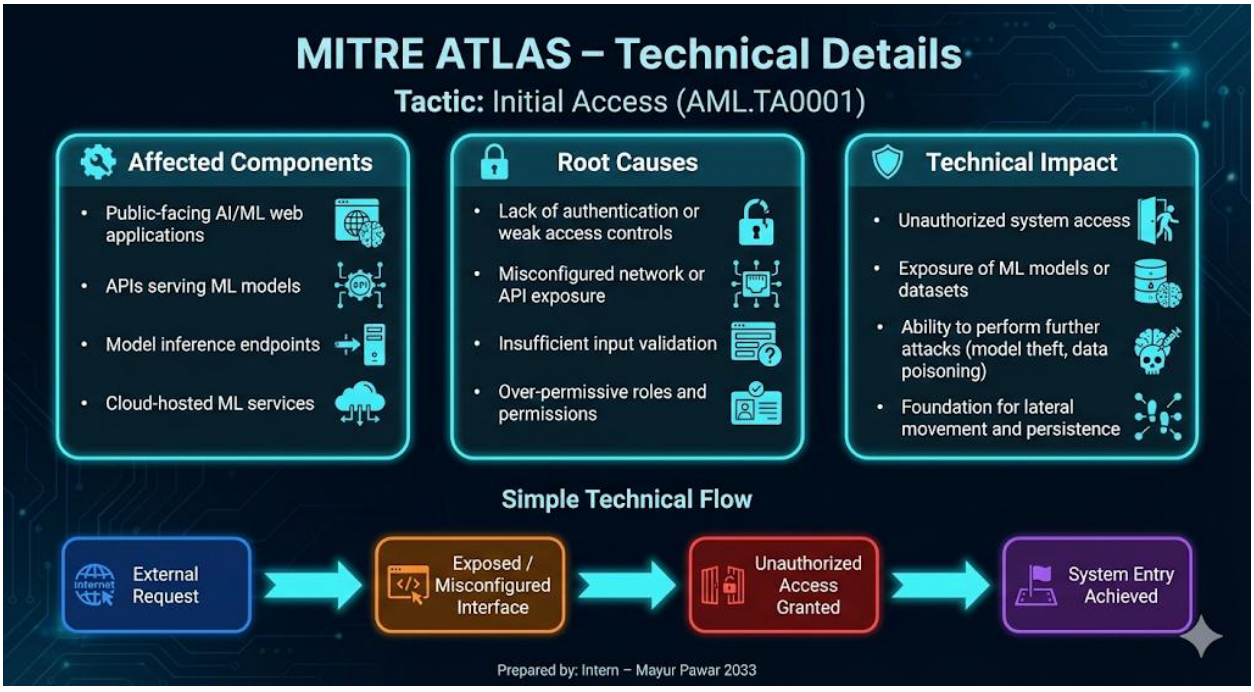

MITRE ATLAS – Attack Flow / Technique
Tactic: Resource Development (AML.TA0003)
Attack Flow (Step-by-Step)
1 Attacker studies the target AI system
2 Develops malicious datasets / models / tools
3 Hosts resources on public repos or cloud
4 Victim unknowingly uses attacker resources
5 AI system becomes vulnerable or compromised
Common Resource Development Techniques
Dataset Poisoning
Model Backdooring
Malicious Pre-Trained Models
Infrastructure Setup (C2, APIs, servers)
Tool Development for ML exploitation
MITRE ATLAS Mapping
Goal: Prepare assets to enable future AI attacks
Intern – Mayur pawar 2033

3. Initial Access

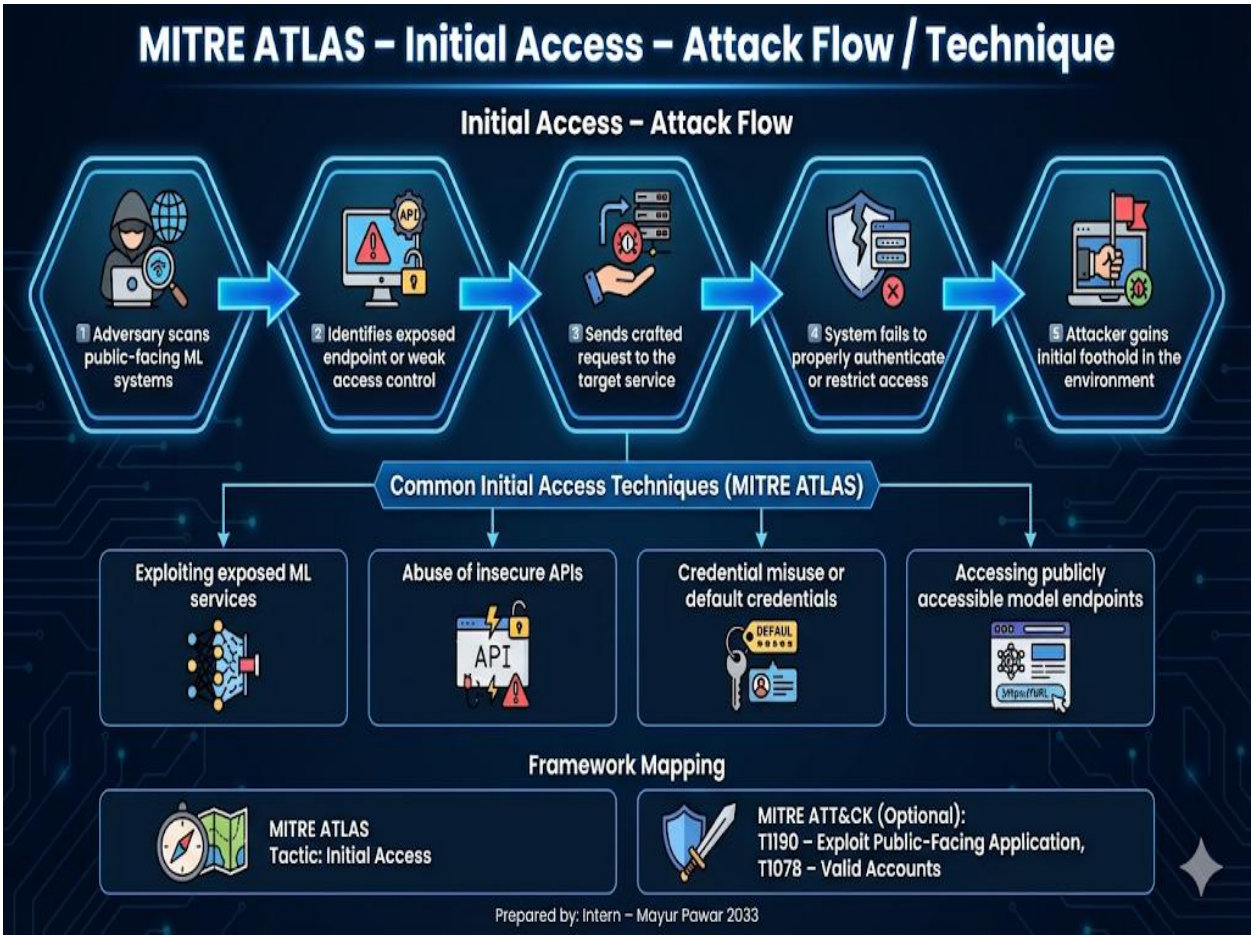
Overview



Technical Details:

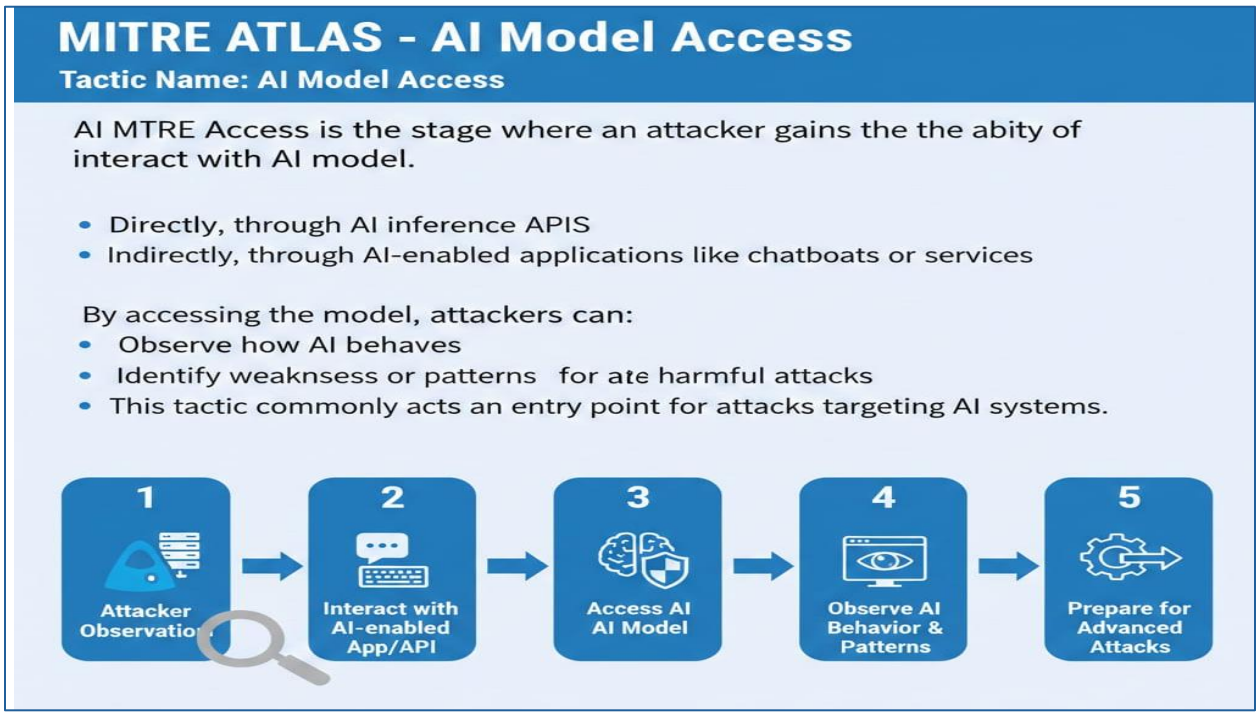


Attack Flow:

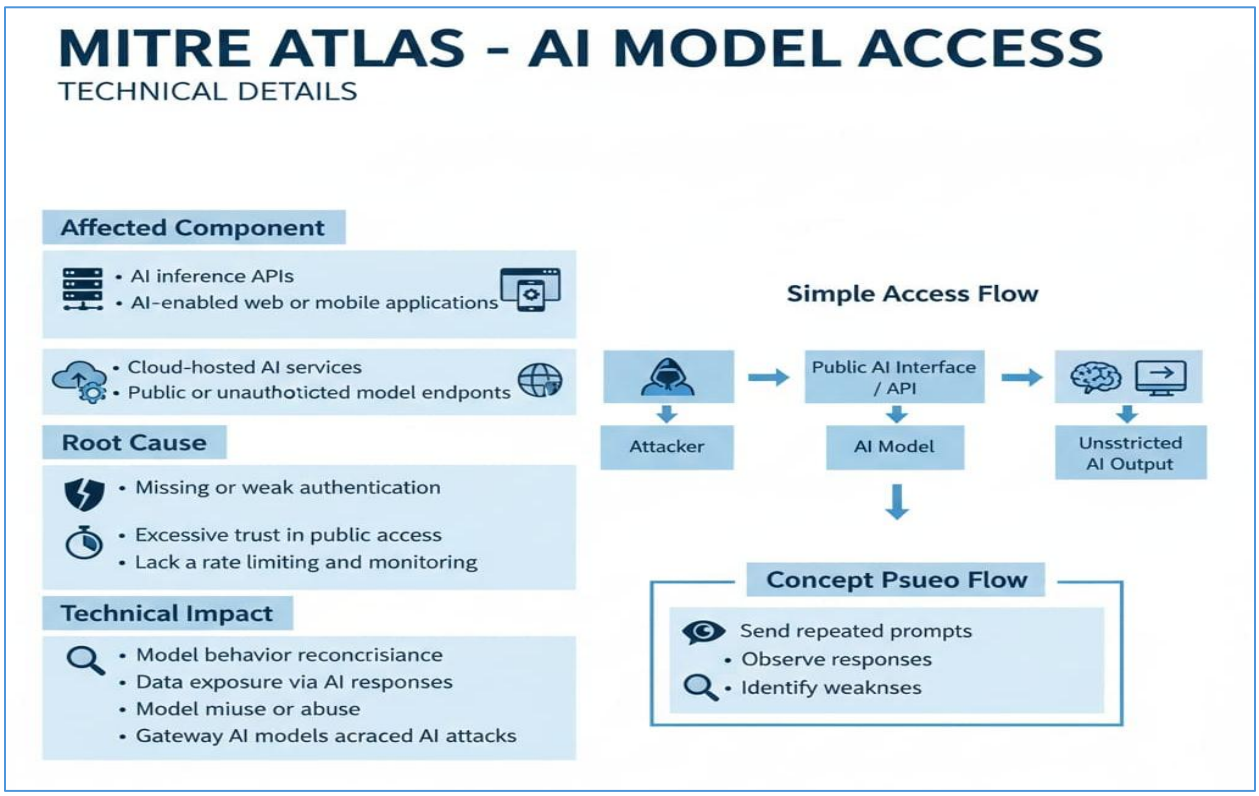

MITRE ATLAS – Initial Access – Attack Flow / Technique
Initial Access – Attack Flow
1 Adversary scans public-facing ML systems
2 Identifies exposed endpoint or weak access control
3 Sends crafted request to the target service
4 System fails to properly authenticate or restrict access
5 Attacker gains initial foothold in the environment
Common Initial Access Techniques (MITRE ATLAS)
Exploiting exposed ML services
Abuse of insecure APIs
API
Credential misuse or default credentials
Accessing publicly accessible model endpoints
Framework Mapping
MITRE ATLAS
Tactic: Initial Access
MITRE ATT&CK (Optional):
T1190 – Exploit Public-Facing Application,
T1078 – Valid Accounts
Prepared by: Intern – Mayur Pawar 2033

4. AI Model Access

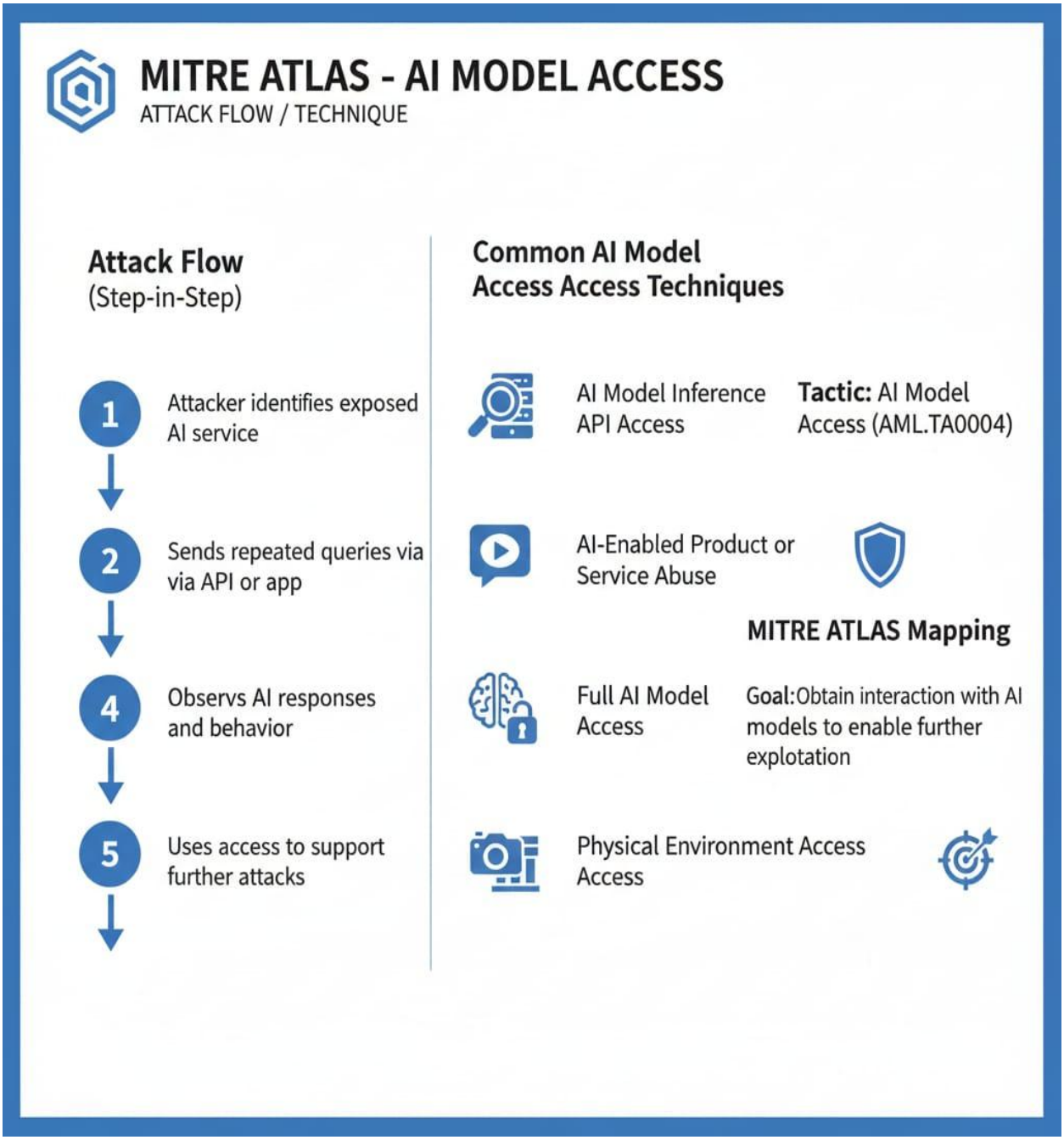
Overview



Technical Details



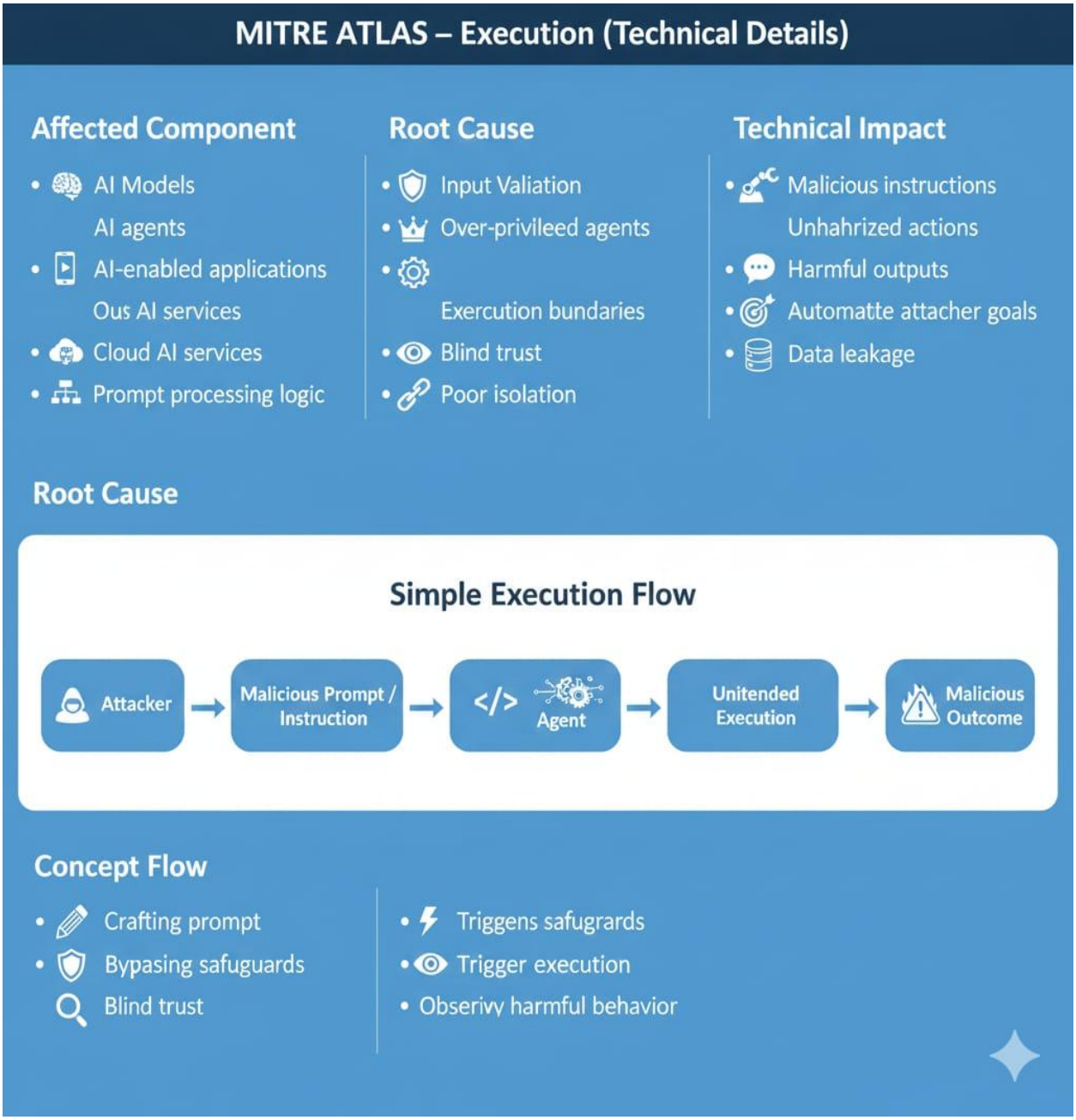
Attack Flow


MITRE ATLAS - AI MODEL ACCESS
ATTACK FLOW / TECHNIQUE
Attack Flow
(Step-in-Step)
1
Attacker identifies exposed AI service
2
Sends repeated queries via via API or app
4
Observs AI responses and behavior
5
Uses access to support further attacks
Common AI Model Access Access Techniques
AI Model Inference API Access
Tactic: AI Model Access (AML.TA0004)
AI-Enabled Product or Service Abuse
MITRE ATLAS Mapping
Full AI Model Access
Goal:Obtain interaction with AI models to enable further explotation
Physical Environment Access Access

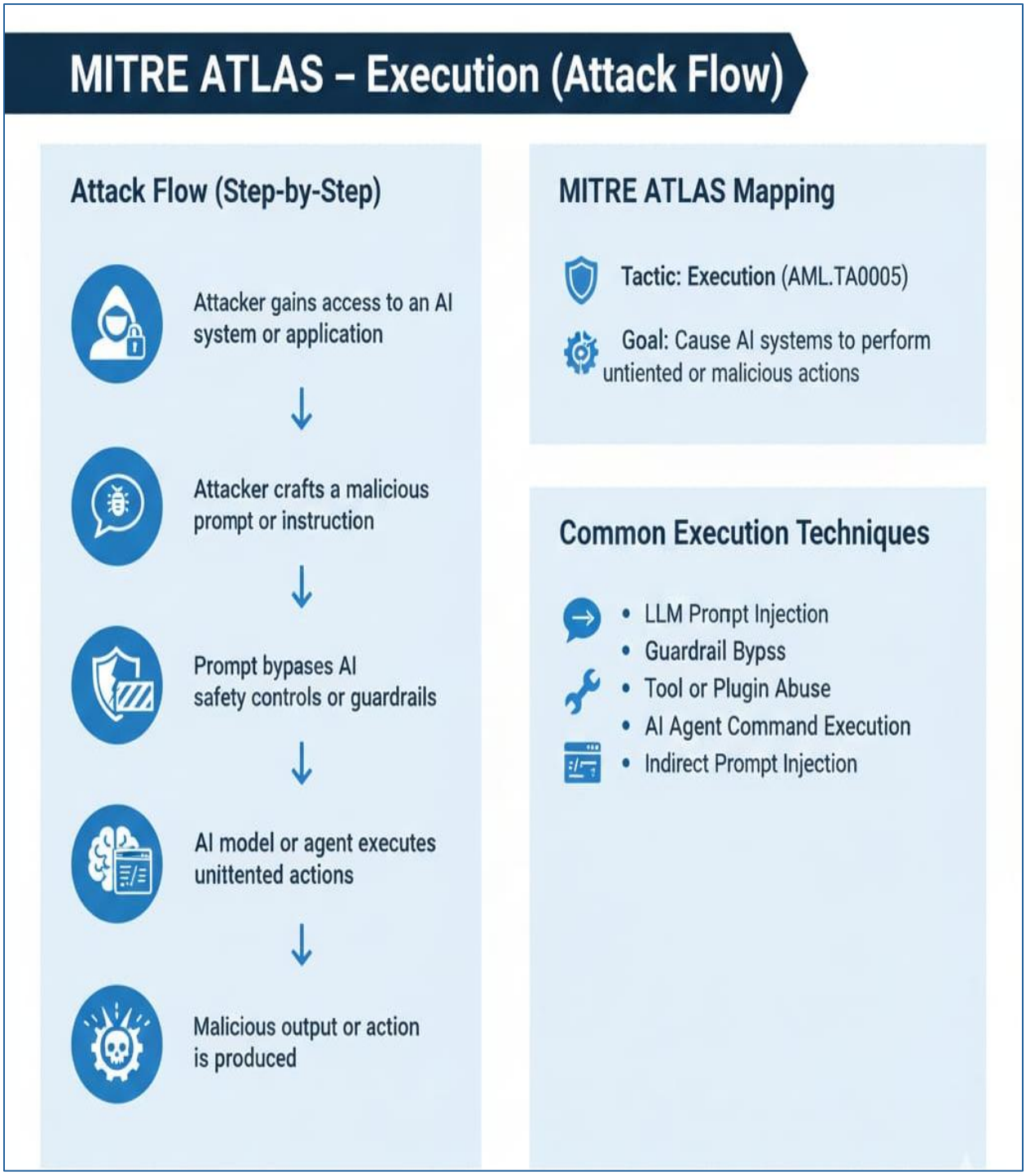
5. Execution

Overview

Technical Details



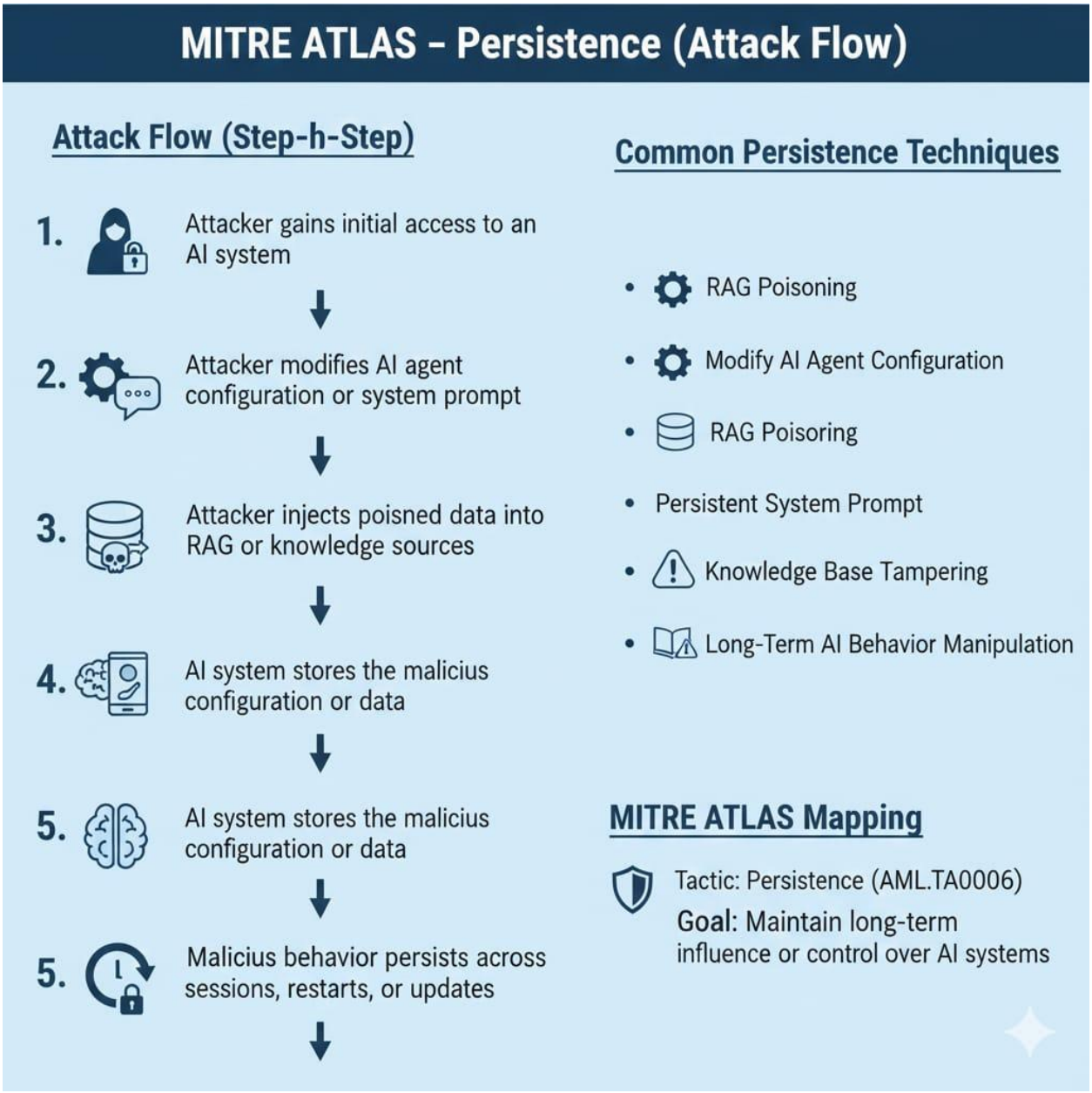
Attack Flow


MITRE ATLAS – Execution (Attack Flow)
Attack Flow (Step-by-Step)
Attacker gains access to an AI system or application
Attacker crafts a malicious prompt or instruction
Prompt bypases AI safety controls or guardrails
AI model or agent executes unittented actions
Malicious output or action is produced
MITRE ATLAS Mapping
Tactic: Execution (AML.TA0005)
Goal: Cause AI systems to perform untiented or malicious actions
Common Execution Techniques
LLM Prompt Injection
Guardrail Bypss
Tool or Plugin Abuse
AI Agent Command Execution
Indirect Prompt Injection

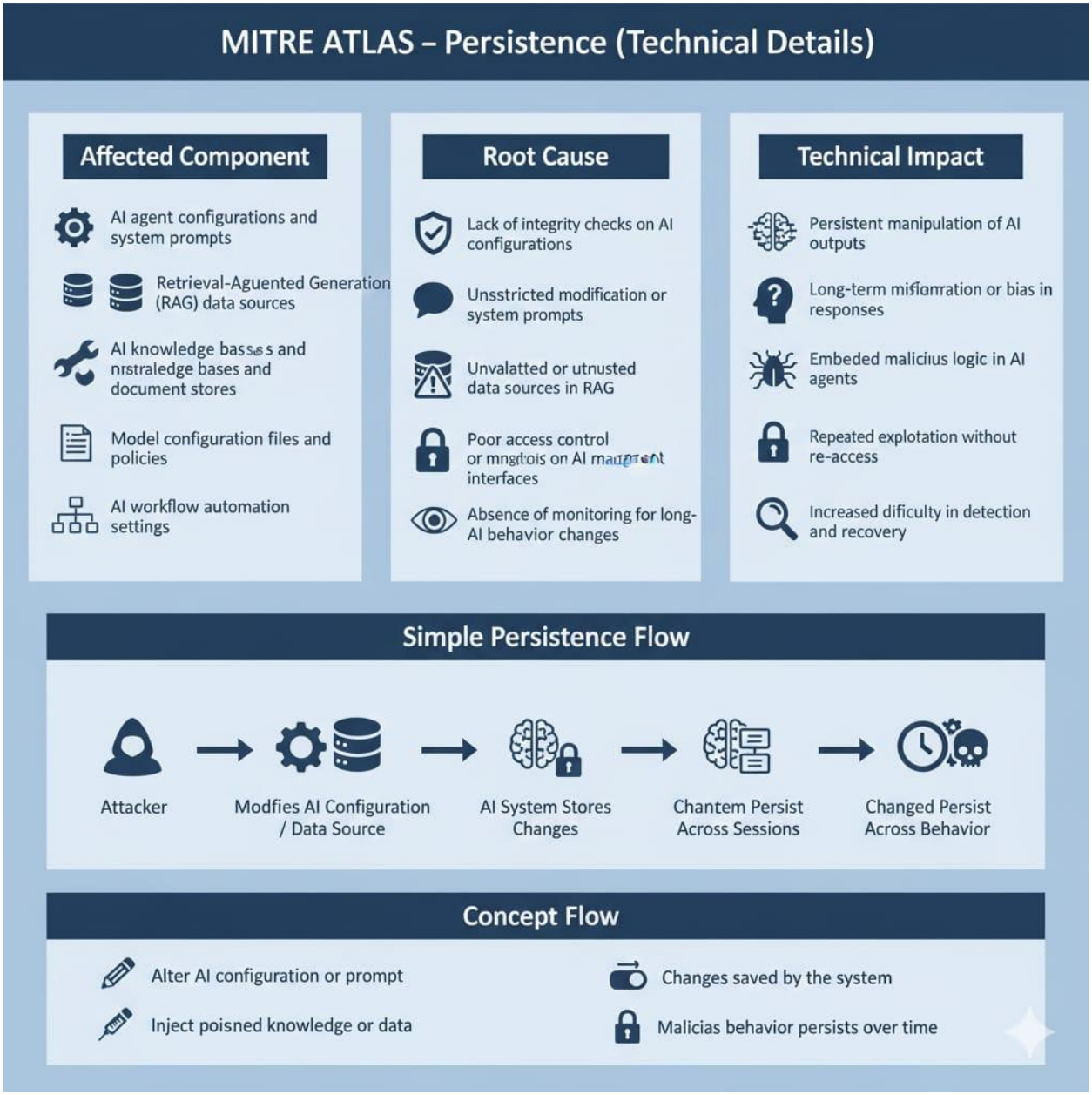
6. Persistence

Overview

Technical Details

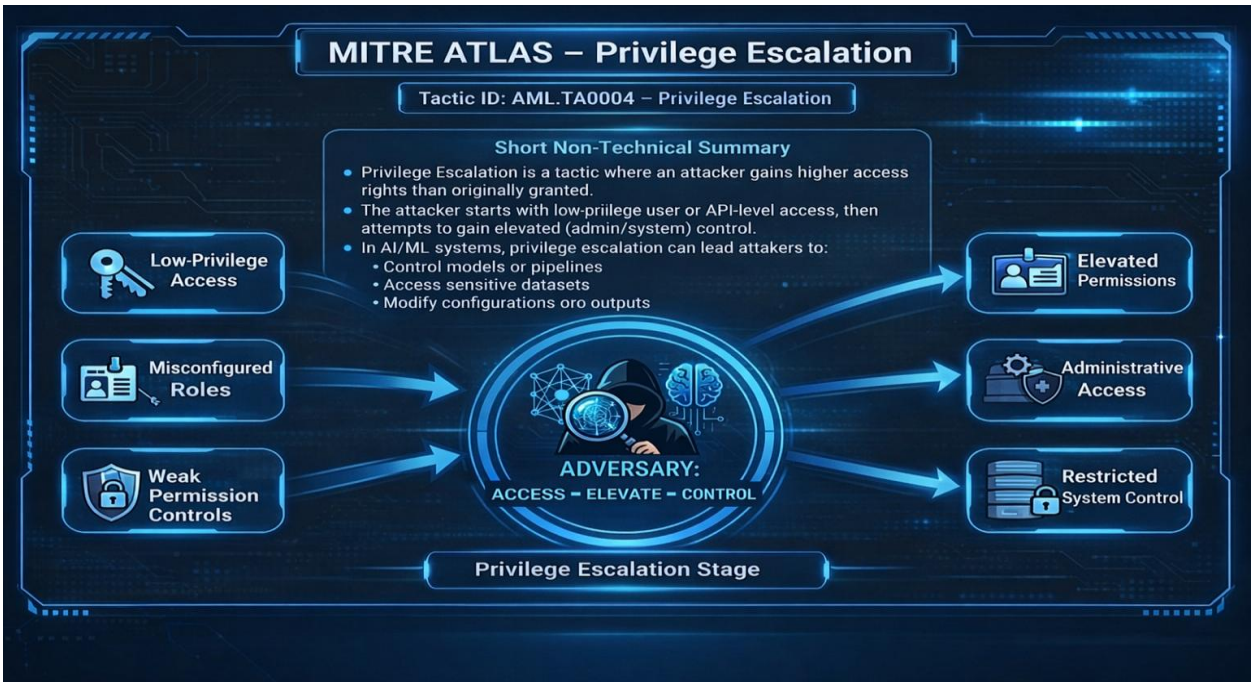


Attack Flow

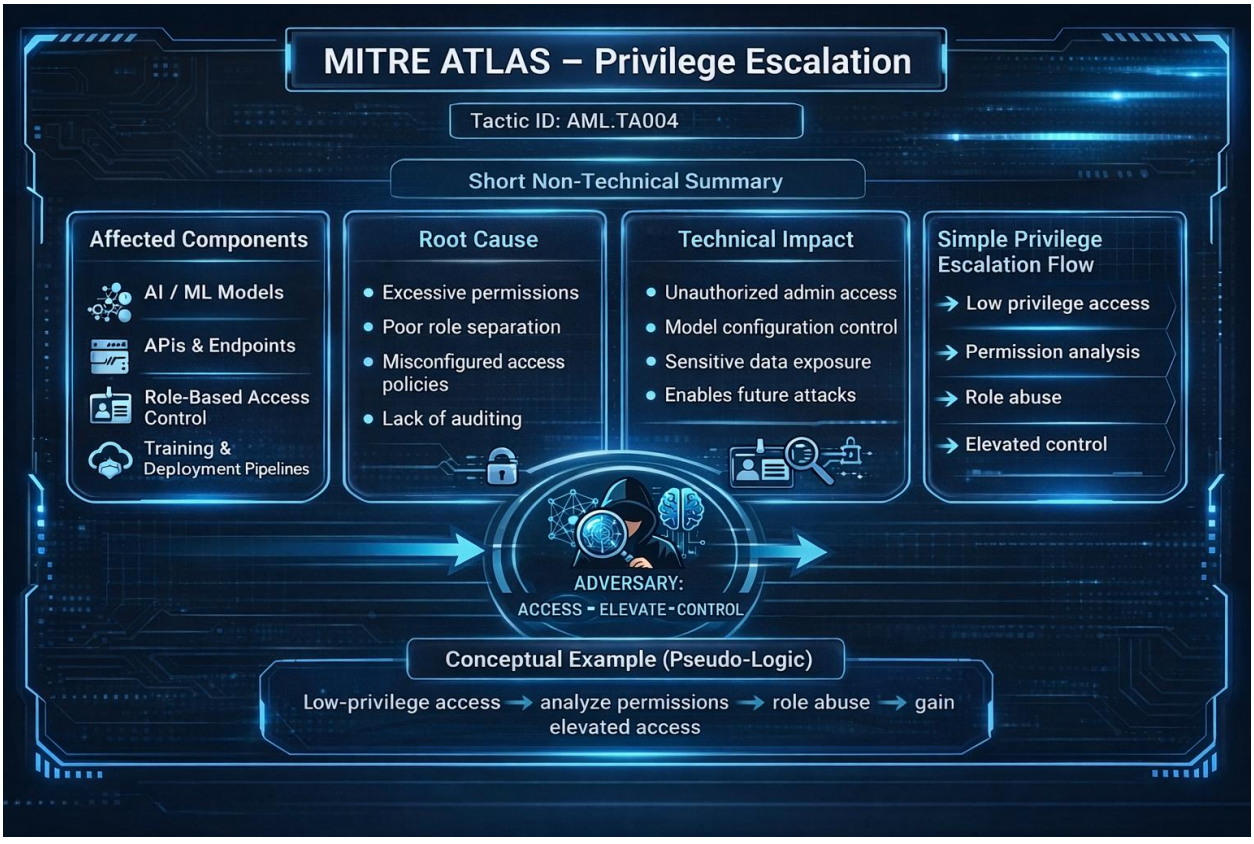

MITRE ATLAS – Persistence (Technical Details)
Affected Component
AI agent configurations and system prompts
Retrieval-Aguented Generation (RAG) data sources
AI knowledge bass s and nrstraledge bases and document stores
Model configuration files and policies
AI workflow automation settings
Root Cause
Lack of integrity checks on AI configurations
Unsstricted modification or system prompts
Unvalatted or utnusted data sources in RAG
Poor access control
interfaces
Absence of monitoring for long-AI behavior changes
Technical Impact
Persistent manipulation of AI outputs
Long-term misiarration or bias in responses
Embeded malicius logic in AI agents
Repeated explotation without re-access
Increased dificulty in detection and recovery
Simple Persistence Flow
Attacker
Modfies AI Configuration / Data Source
AI System Stores Changes
Chantem Persist Across Sessions
Changed Persist Across Behavior
Concept Flow
Alter AI configuration or prompt
Inject poisned knowledge or data
Changes saved by the system
Malicias behavior persists over time

7. Privilege Escalation

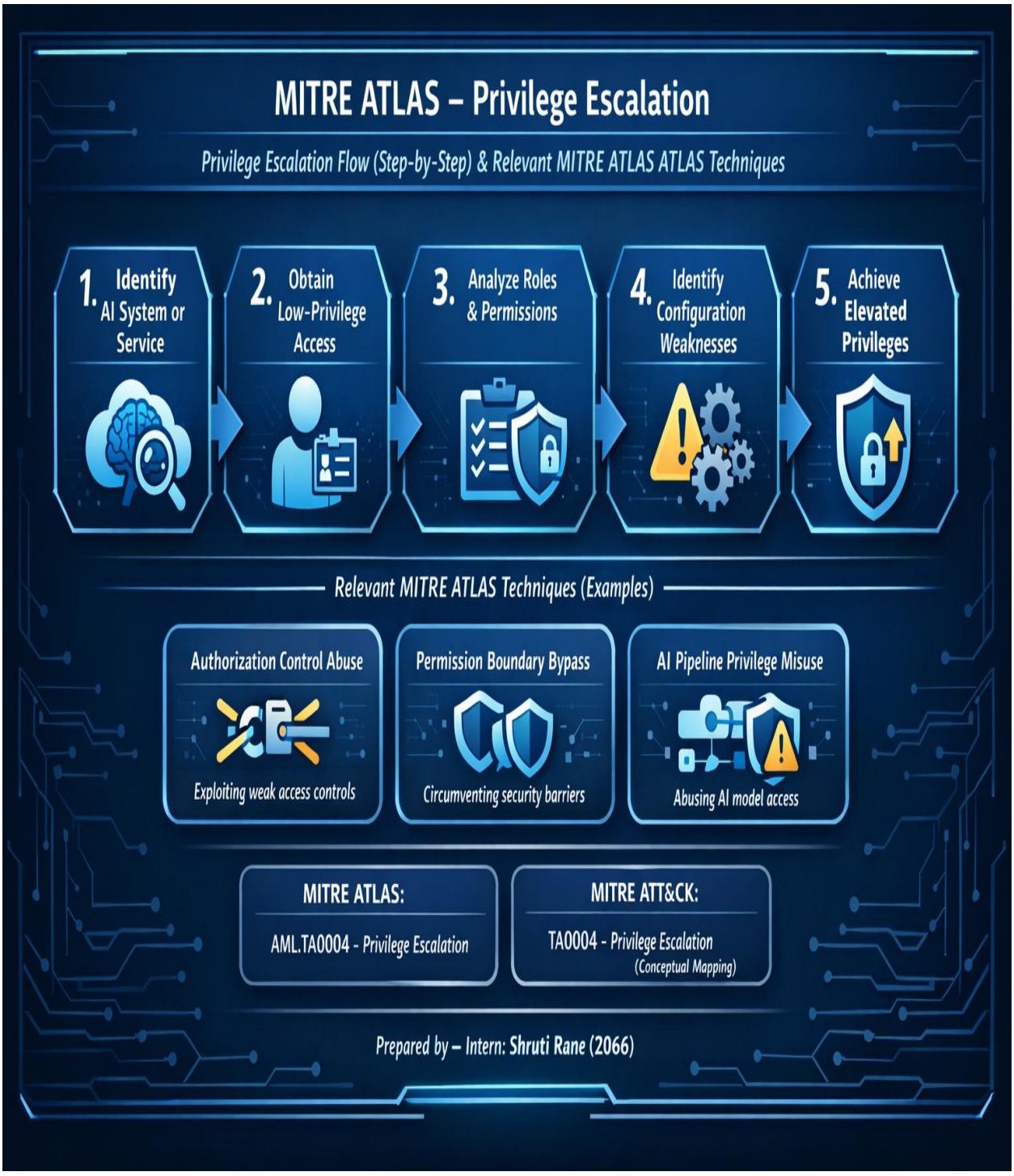
Overview



Technical Details



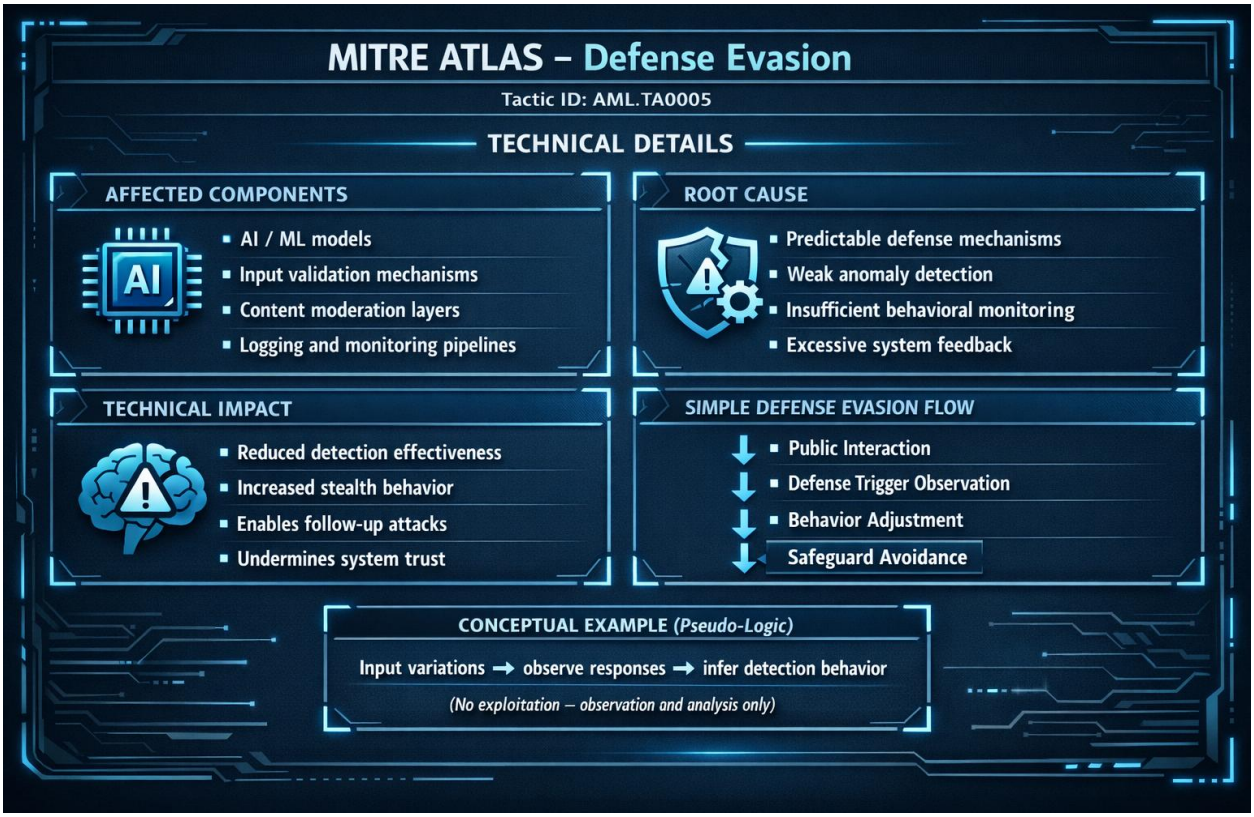
Attack Flow


MITRE ATLAS – Privilege Escalation
Privilege Escalation Flow (Step-by-Step) & Relevant MITRE ATLAS ATLAS Techniques
1. Identify AI System or Service
2. Obtain Low-Privilege Access
3. Analyze Roles & Permissions
4. Identify Configuration Weaknesses
5. Achieve Elevated Privileges
Relevant MITRE ATLAS Techniques (Examples)
Authorization Control Abuse
Exploiting weak access controls
Permission Boundary Bypass
Circumventing security barriers
AI Pipeline Privilege Misuse
Abusing AI model access
MITRE ATLAS:
AML.TA0004 - Privilege Escalation
MITRE ATT&CK:
TA0004 - Privilege Escalation
(Conceptual Mapping)
Prepared by – Intern: Shruti Rane (2066)

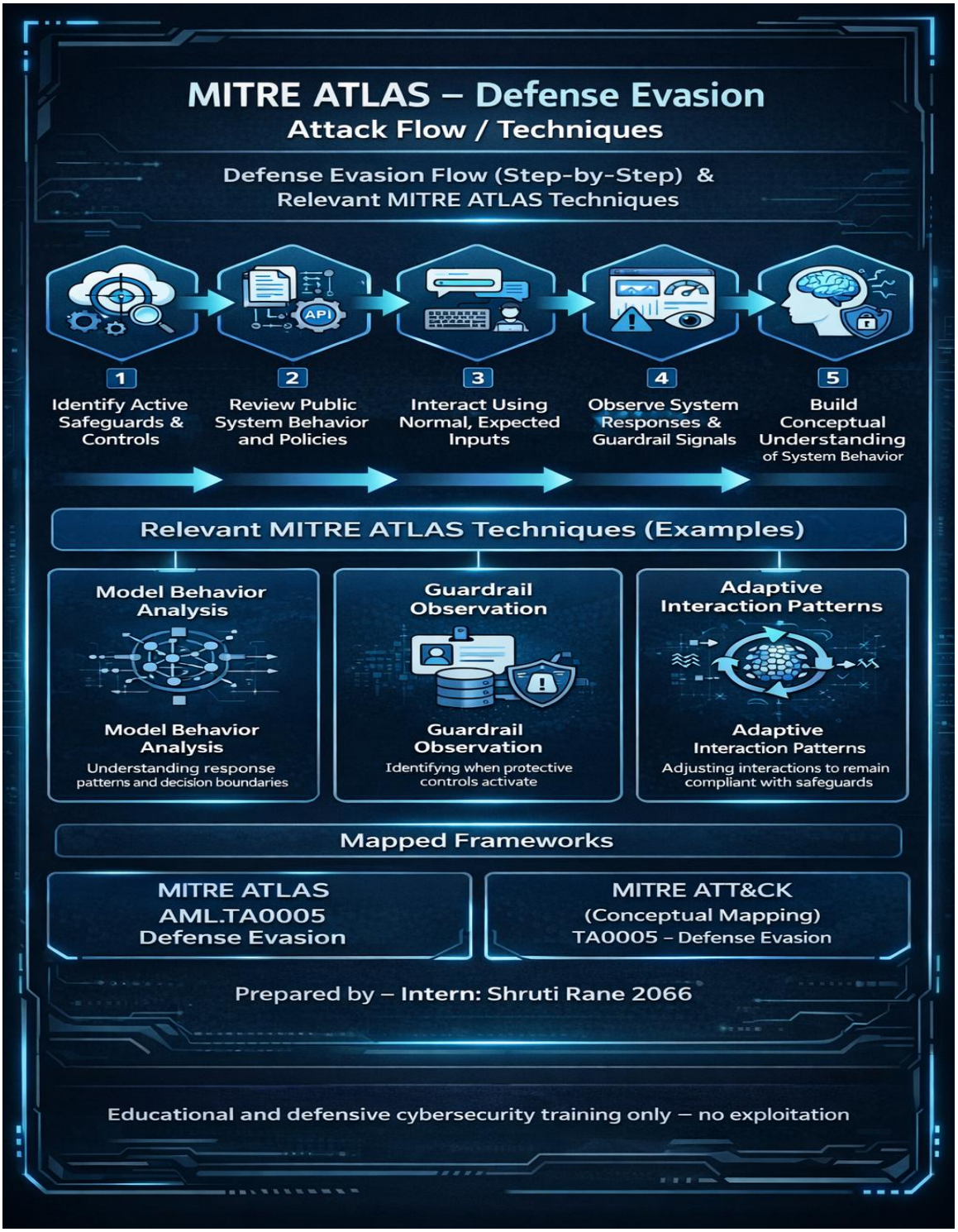
8. Defense Evasion

Overview

Technical Details:

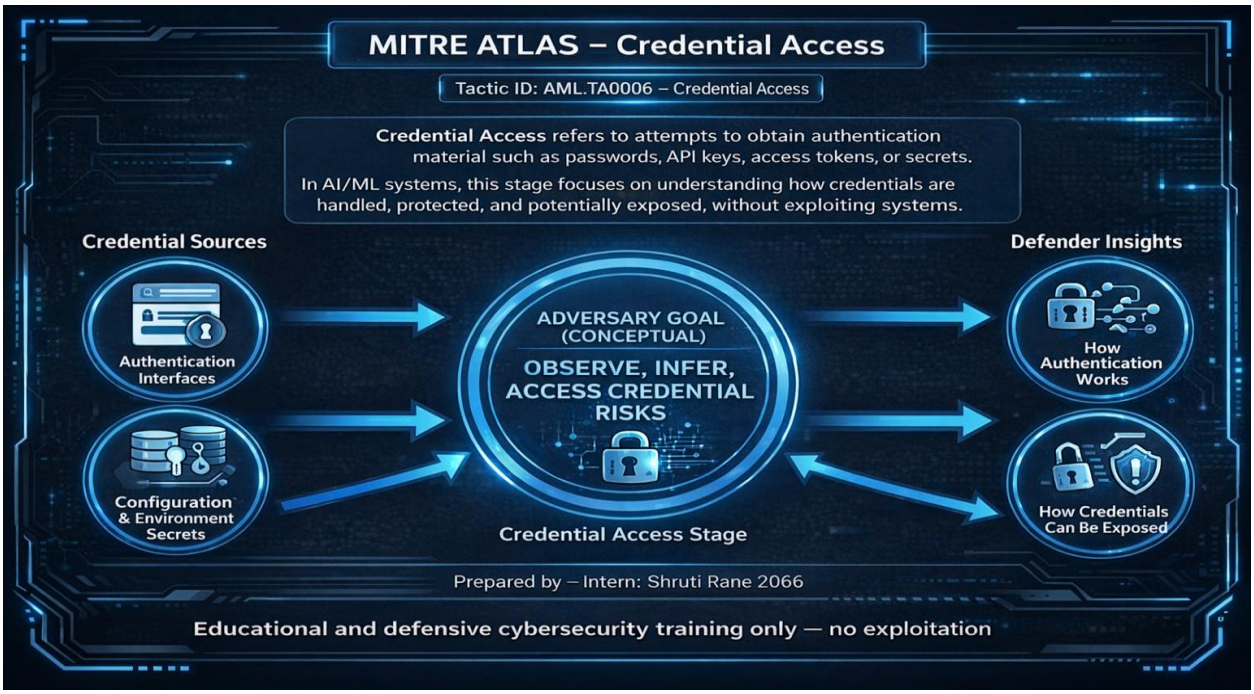


Attack Flow:

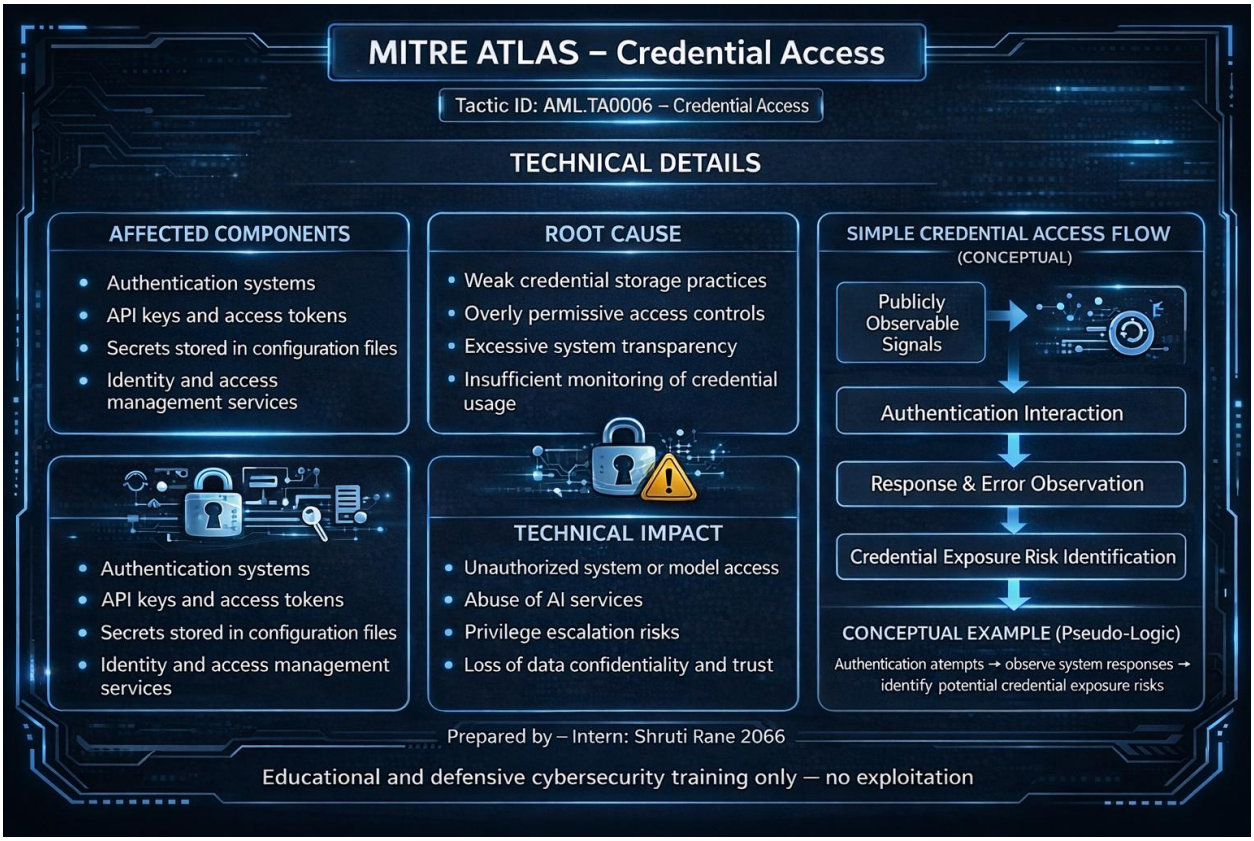

MITRE ATLAS – Defense Evasion
Attack Flow / Techniques
Defense Evasion Flow (Step-by-Step) &
Relevant MITRE ATLAS Techniques
API
1
Identify Active Safeguards & Controls
2
Review Public System Behavior and Policies
3
Interact Using Normal, Expected Inputs
4
Observe System Responses & Guardrail Signals
5
Build Conceptual Understanding of System Behavior
Relevant MITRE ATLAS Techniques (Examples)
Model Behavior Analysis
Model Behavior Analysis
Understanding response patterns and decision boundaries
Guardrail Observation
Guardrail Observation
Identifyng when protective controls activate
Adaptive Interaction Patterns
Adaptive Interaction Patterns
Adjusting interactions to remain compliant with safeguards
Mapped Frameworks
MITRE ATLAS
AML.TA0005
Defense Evasion
MITRE ATT&CK
(Conceptual Mapping)
TA0005 – Defense Evasion
Prepared by – Intern: Shruti Rane 2066
Educational and defensive cybersecurity training only — no exploitation

9. Credential Access

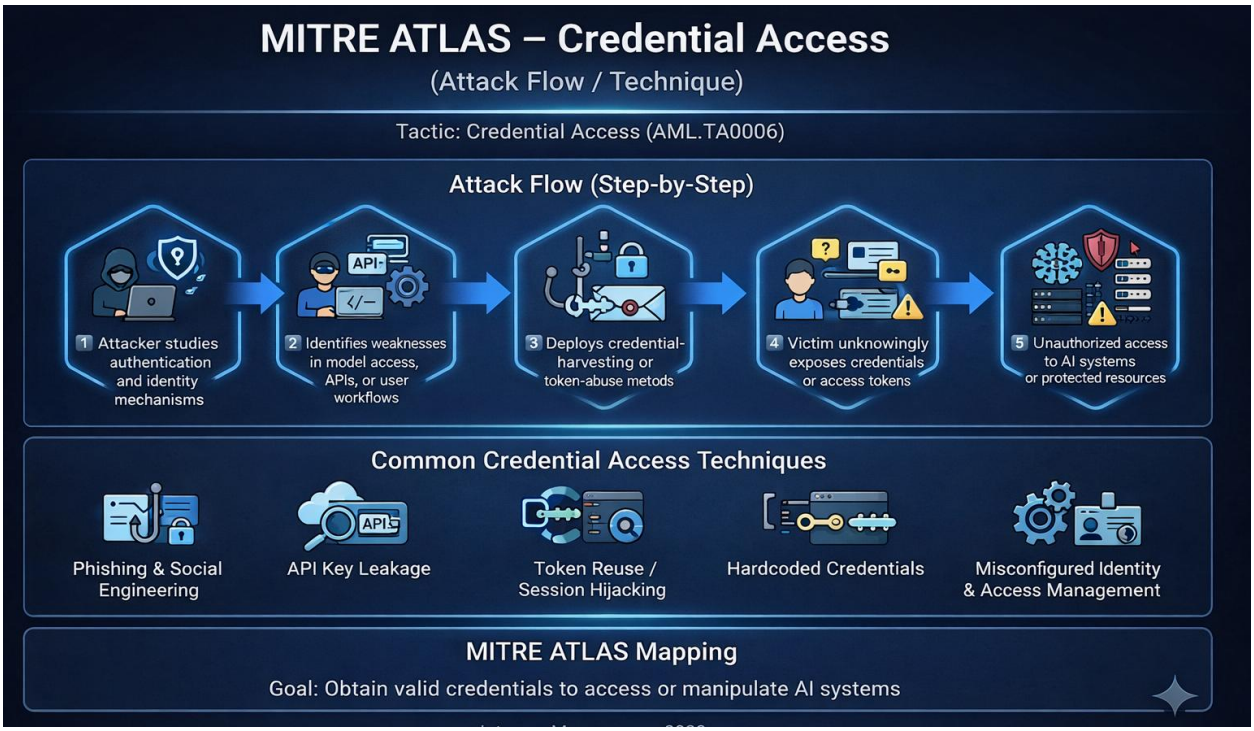
Overview



Technical Details:

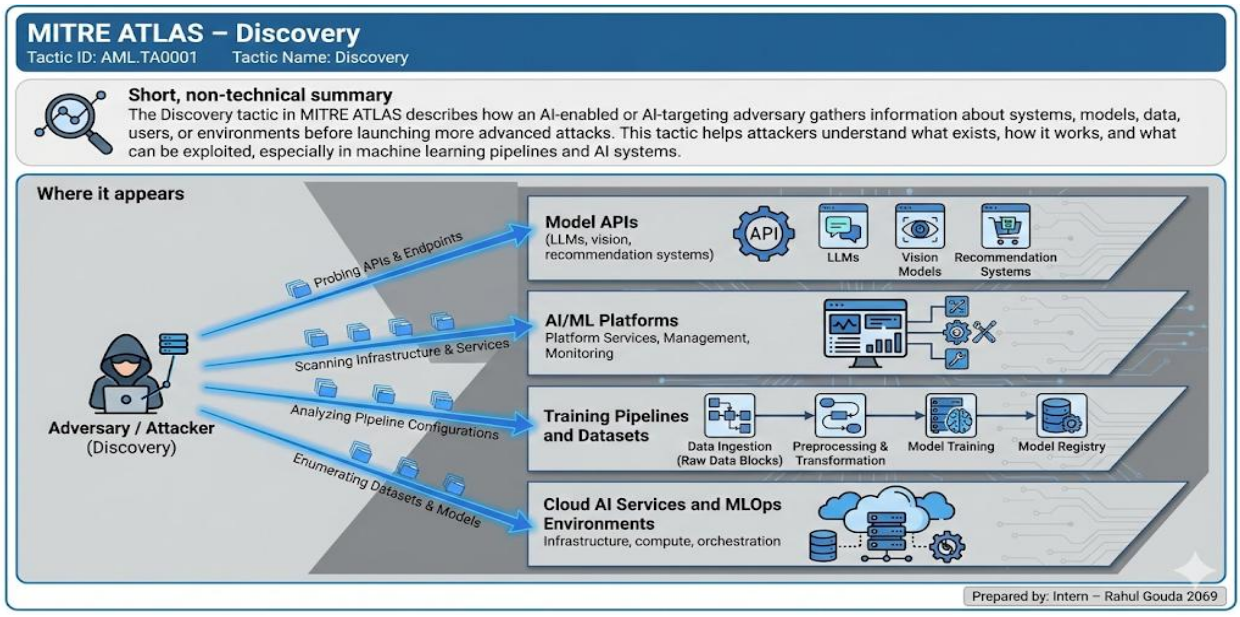


Attack Flow

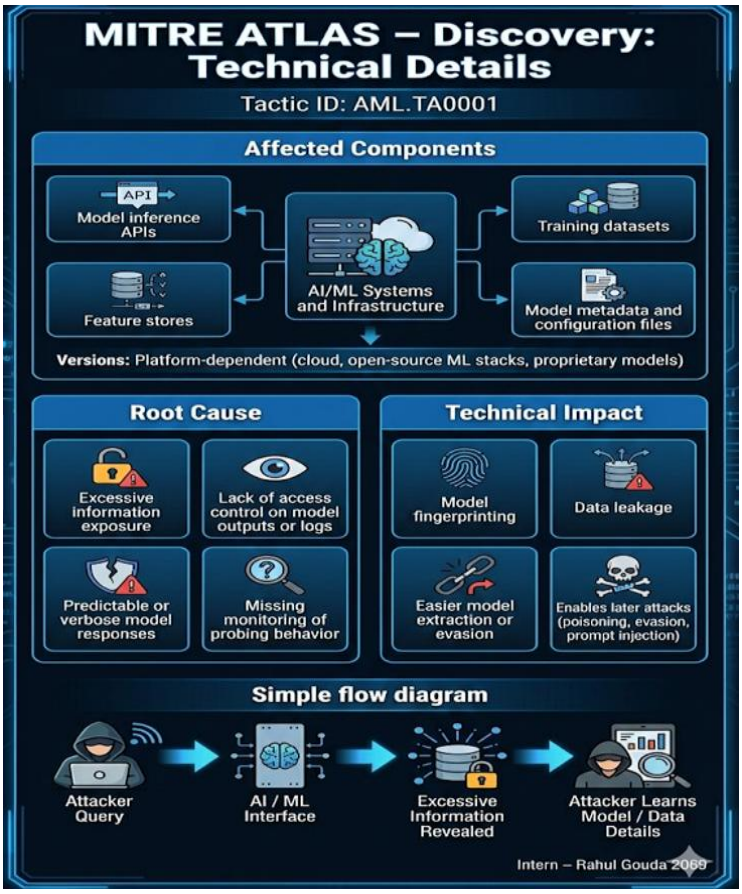

MITRE ATLAS – Credential Access
(Attack Flow / Technique)
Tactic: Credential Access (AML.TA0006)
Attack Flow (Step-by-Step)
1 Attacker studies authentication and identity mechanisms
2 Identifies weaknesses in model access, APIs, or user workflows
3 Deploys credential-harvesting or token-abuse metods
4 Victim unknowingly exposes credentials or access tokens
5 Unauthorized access to AI systems or protected resources
Common Credential Access Techniques
Phishing & Social Engineering
API Key Leakage
Token Reuse / Session Hijacking
Hardcoded Credentials
Misconfigured Identity & Access Management
MITRE ATLAS Mapping
Goal: Obtain valid credentials to access or manipulate AI systems

10. Discovery

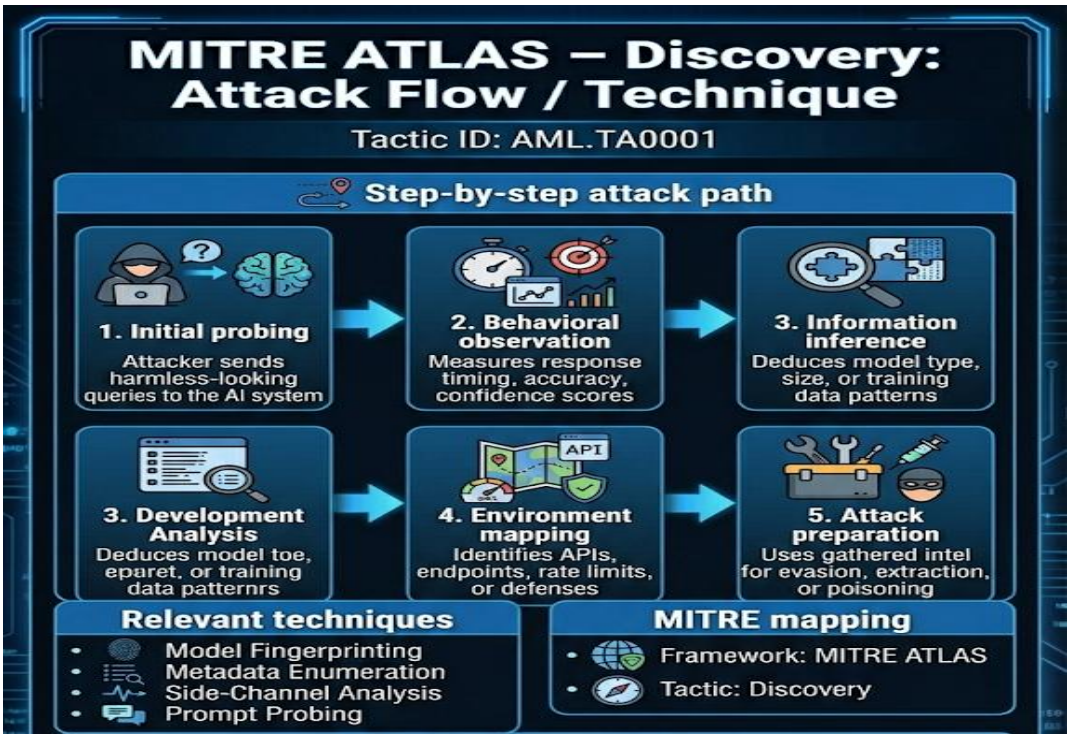
Overview



Technical Details:

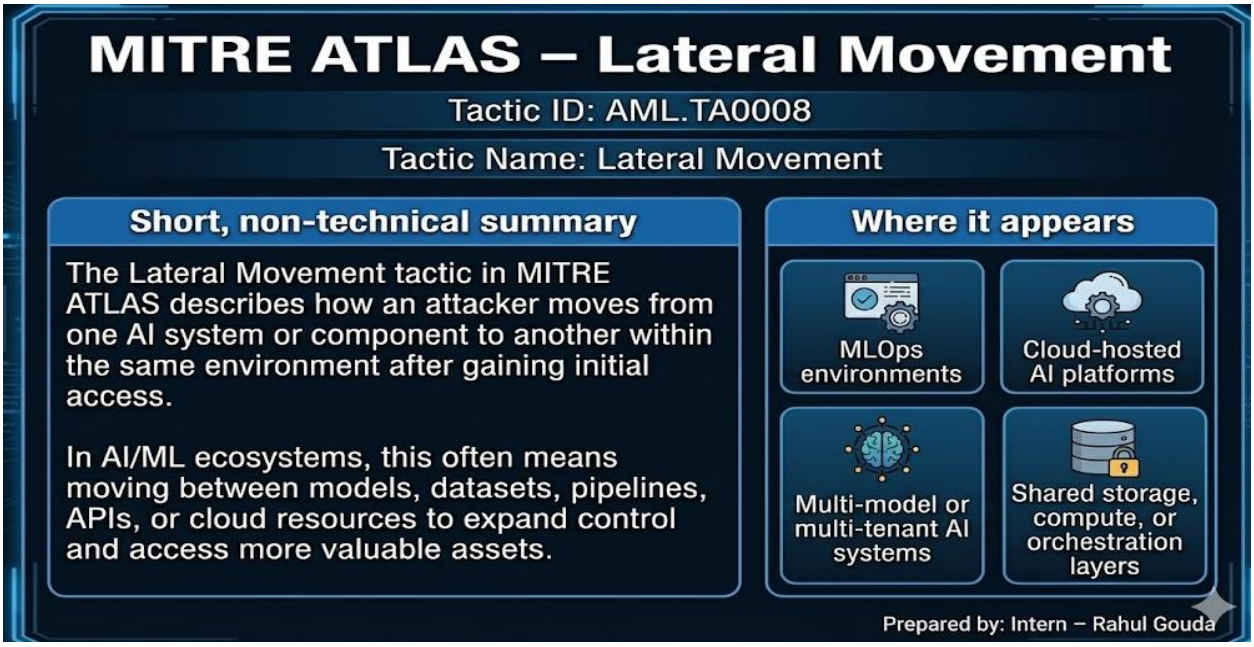


Attack Flow:

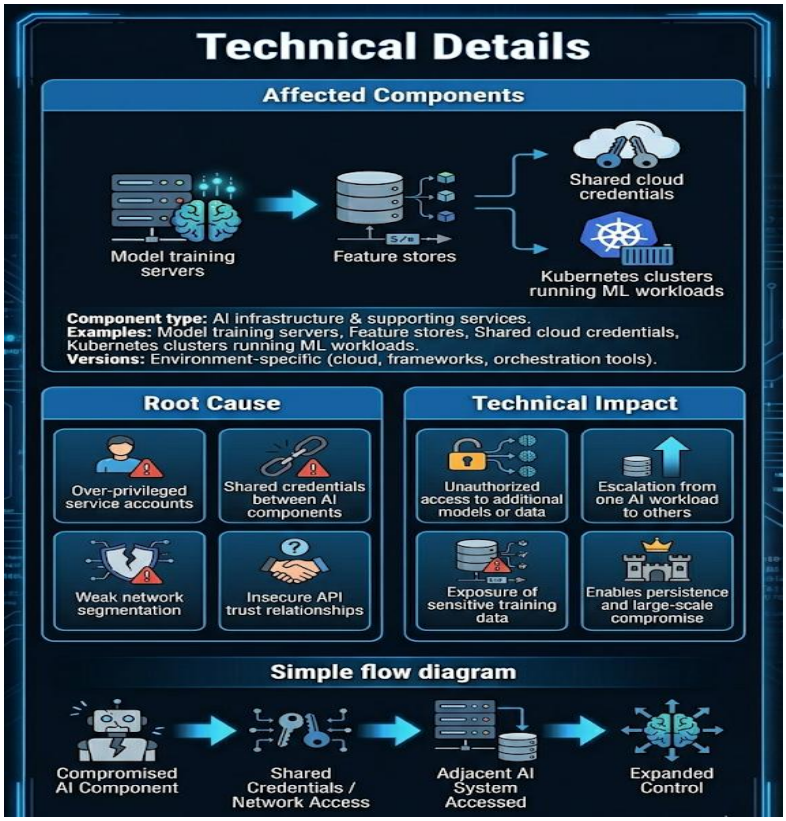


11. Lateral Movement

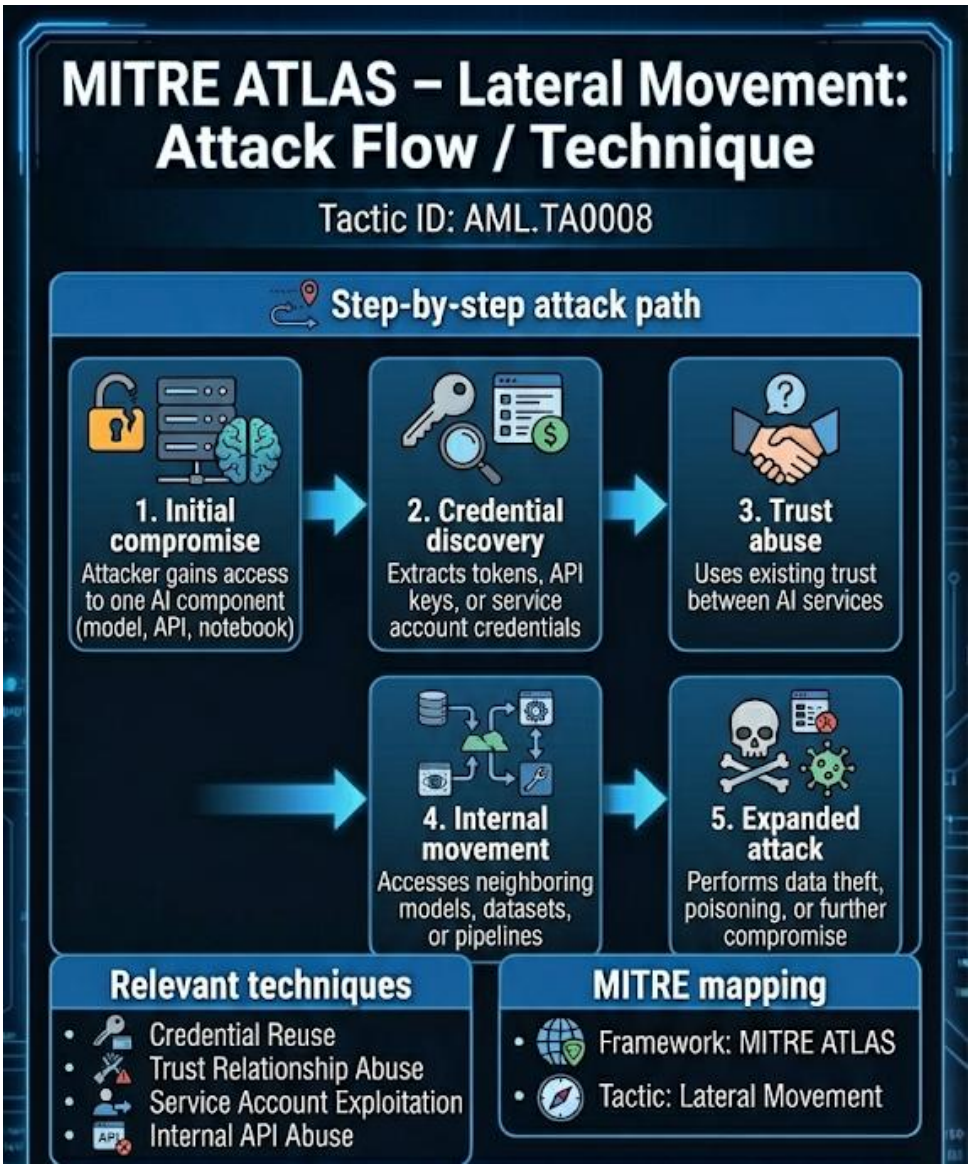
Overview



Technical Details:

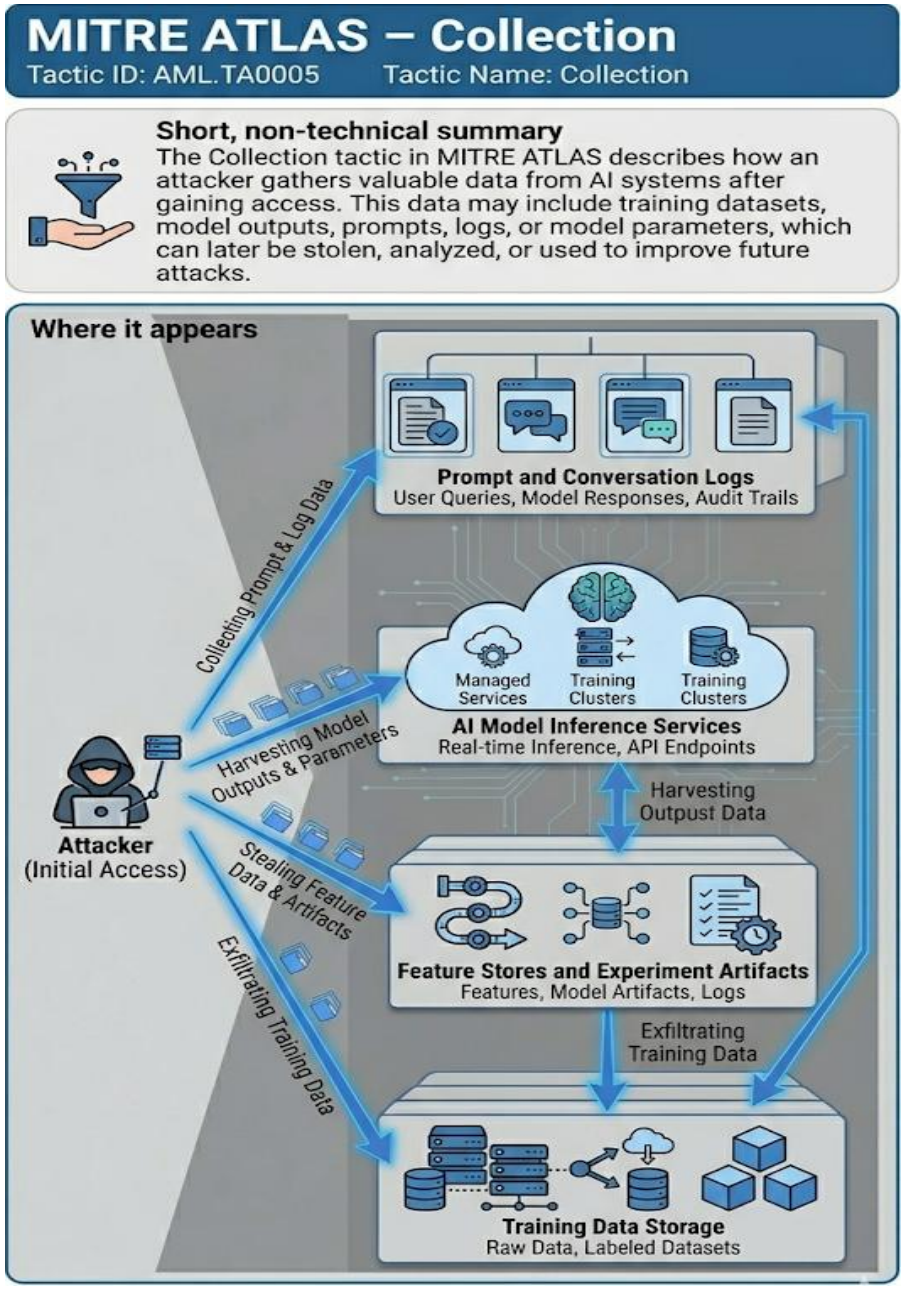


Attack Flow:

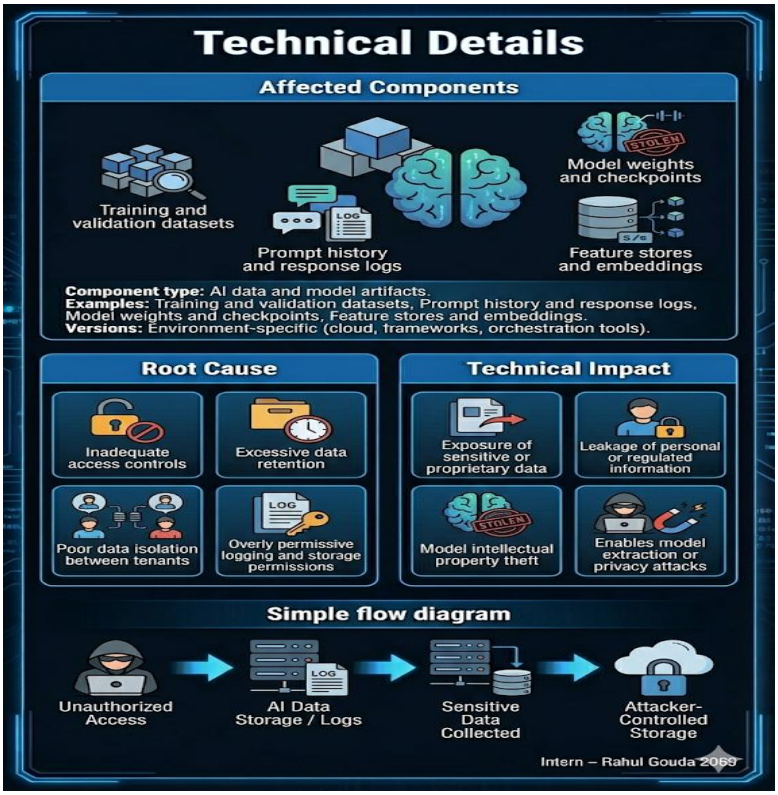

MITRE ATLAS – Lateral Movement: Attack Flow / Technique
Tactic ID: AML.TA0008
Step-by-step attack path
1. Initial compromise
Attacker gains access to one AI component (model, API, notebook)
2. Credential discovery
Extracts tokens, API keys, or service account credentials
3. Trust abuse
Uses existing trust between AI services
4. Internal movement
Accesses neighboring models, datasets, or pipelines
5. Expanded attack
Performs data theft, poisoning, or further compromise
Relevant techniques
Credential Reuse
Trust Relationship Abuse
Service Account Exploitation
Internal API Abuse
MITRE mapping
Framework: MITRE ATLAS
Tactic: Lateral Movement

12. Collection

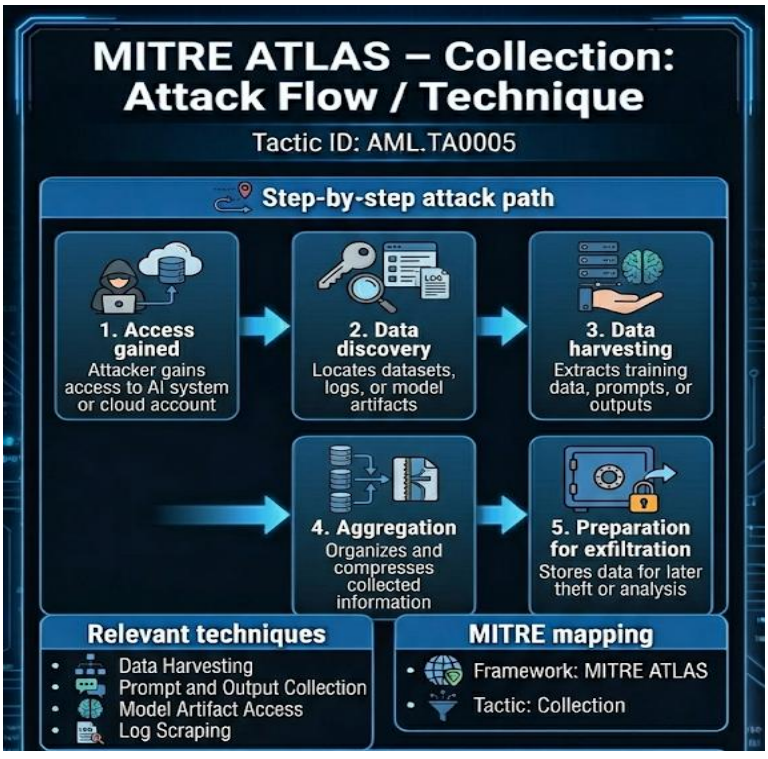
Overview



Technical Details:

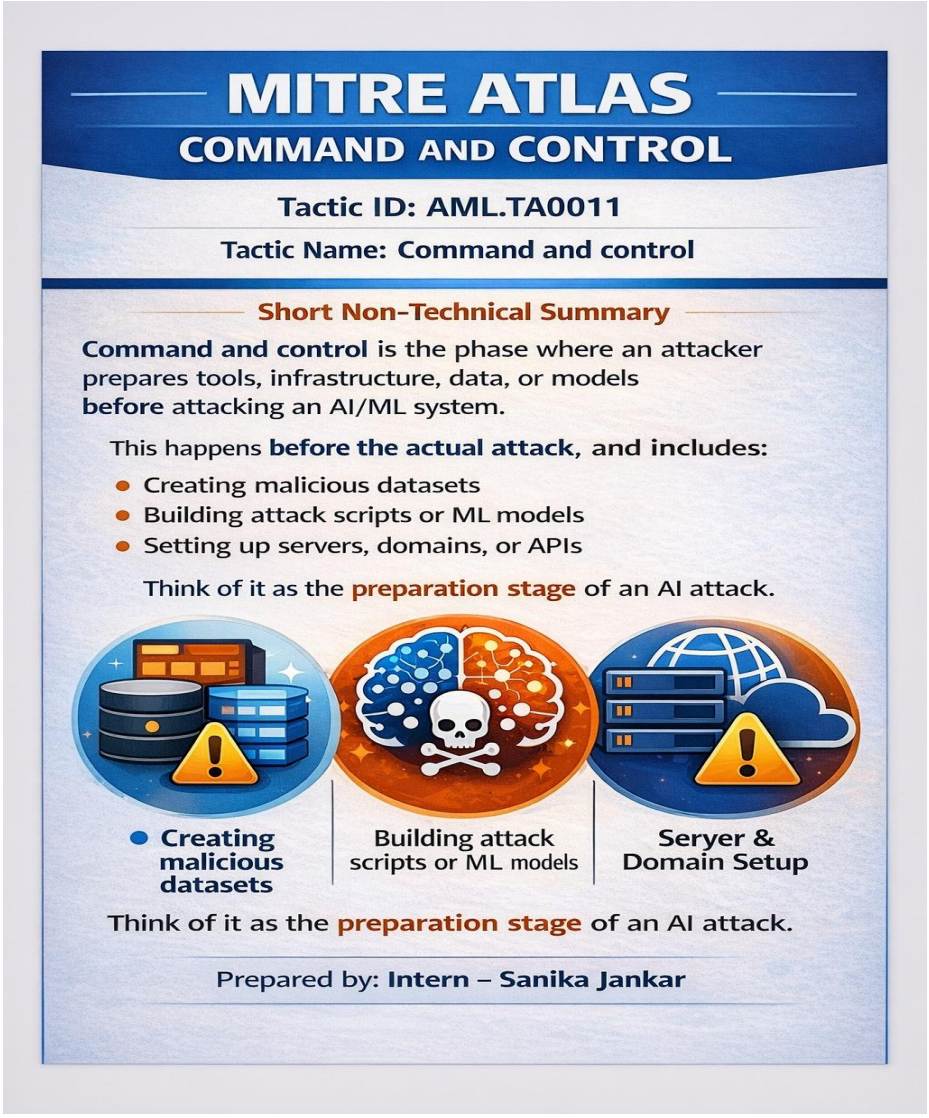


Attack Flow:



13. Command and Control

Overview



Technical detail:

Attack Flow:


MITRE ATLAS
ATTACK FLOW
COMMAND AND CONTROL
Attack Flow (Step-by-Step)
1 Attacker studies the target AI system
2 Develops malicious datasets / models / tools
3 Hosts resources on public repos or cloud
4 Victim unknowingly uses attacker resources
5 AI system becomes vulnerable or compromised
ML
Common Command and Control Techniques
Dataset Poisoning
Model Backdooring
Malicious Pre-Trained Models
Infrastructure Setup (C2, APIs, servers)
Tool Development for ML exploitation
MITRE ATLAS Mapping
Tactic: Command and control (AML.TA0011)
Goal: Prepare assets to enable future AI attacks
Prepared by: Sanika Jankar 2038

14. Exfiltration

Overview

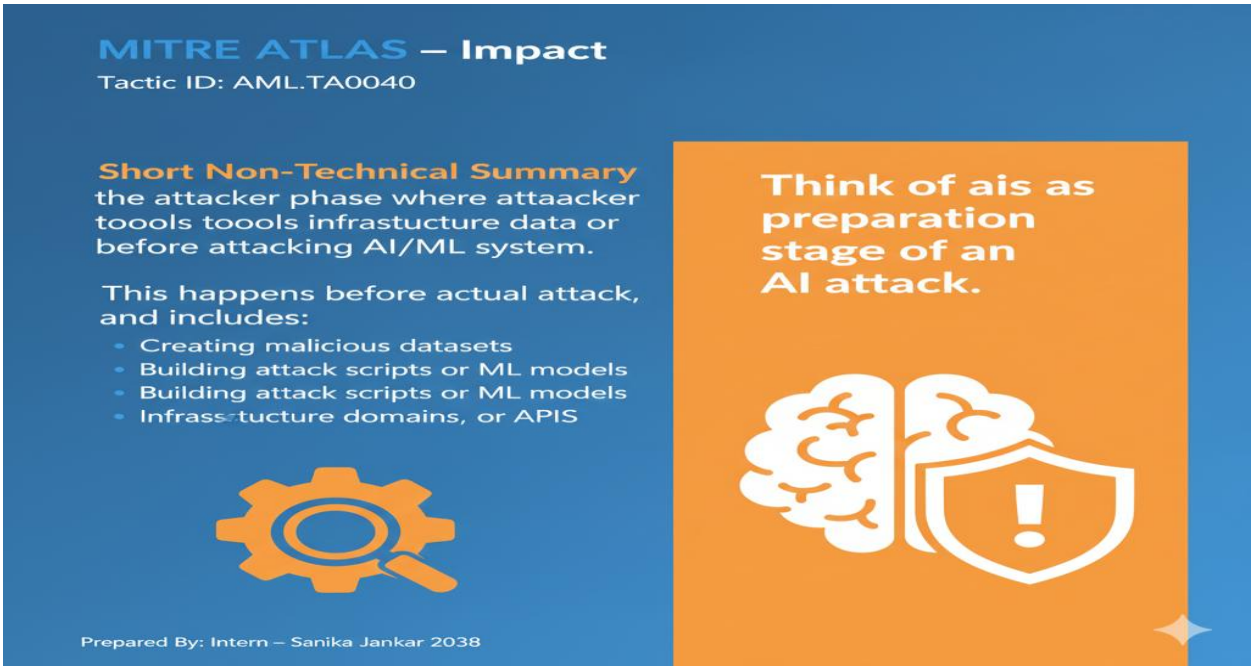
Technical detail:

Attack flow:

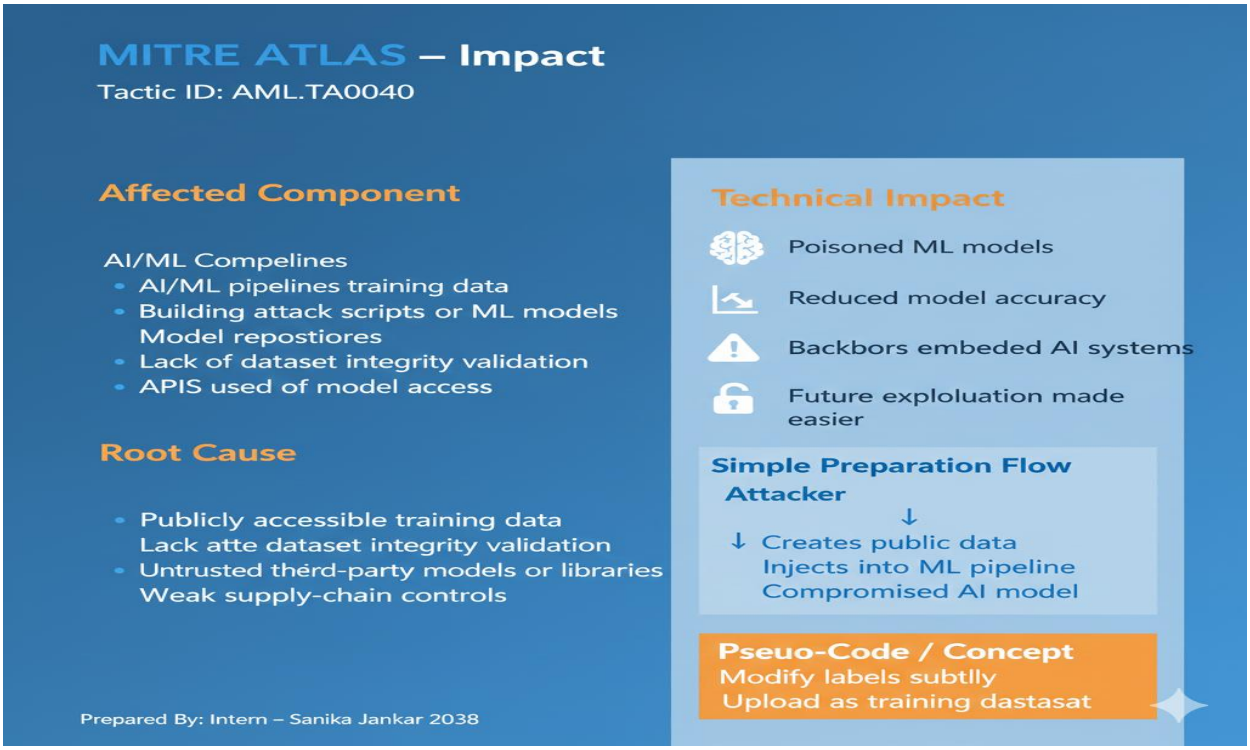

MITRE ATLAS
ATTACK FLOW
EXFILTRATION
Attack Flow (Step-by-Step)
1 Attacker studies the target AI system
2 Develops malicious datasets / models / tools
3 Hosts resources on public repos or cloud
4 Victim unknowingly uses attacker resources
5 AI system becomes vulnerable - or compromised
ML
Common Exfiltration Techniques
Dataset Poisoning
Model Backdooring
Malicious Pre-Trained Models
Infrastructure Setup (C2, APIs, servers)
Tool Development for ML exploitation
MITRE ATLAS Mapping
Attacker
Creates malicious resources
Injects into ML pipeline
Compromised AI model
Pseudo-Code / Concept
Collect public data
Modify labels subtly
Upload as training dataset
Prepared by: Sanika Jankar 2038

15. Impact

Overview



Technical detail:


MITRE ATLAS – Impact
Tactic ID: AML.TA0040
Affected Component
AI/ML Compelines
AI/ML pipelines training data
Building attack scripts or ML models
Model repostiores
Lack of dataset integrity validation
APIS used of model access
Root Cause
Publicly accessible training data
Lack atte dataset integrity validation
Untrusted thérd-party models or libraries
Weak supply-chain controls
Technical Impact
Poisoned ML models
Reduced model accuracy
Backbors embeded AI systems
Future exploluation made easier
Simple Preparation Flow
Attacker
↓
↓ Creates public data
Injects into ML pipeline
Compromised AI model
Pseuo-Code / Concept
Modify labels subtlly
Upload as training dastasat
Prepared By: Intern – Sanika Jankar 2038

Attack flow:

MITRE ATLAS: Impact (AML.TA0040) - Attack Flow

1. Attacker studies the target AI system
2. Develops malicious datssates/models
3. Hosts resources on public repos/cloud
4. Victim unkıowüily uses resources
5. AI system becomes comprmised

Common Impact Techniques

- Dataset Poisoning
- Model Backdooring
- Malicious Pre-Trained Models
- C2 Infrastucture

MITRE ATLAS Mapping

Tactic: Impact (AML.TA0040)

Goal: Strategic preparation of assets to ensure the success a future AI attack

Prepared by: Intern – Sanika Jankar 2038