Evgeni Dyulgerov
CTO/Digitalix

API security in the age of AI

OWASP Top 10 for LLM, API security tools and design patterns

DIGITALIX

About me:

- A software engineer with 12+ years of experience with 30+ projects of various size and complexity.

- Cofounder of “Digitalix” – a company providing software and cloud consulting services.

- Lecturer at the “Cybersecurity” department of Technical University of Sofia.

- PhD candidate focused on “Application of AI in critical infrastructure cybersecurity”.

- Coauthoring several books focused on “cloud security”.

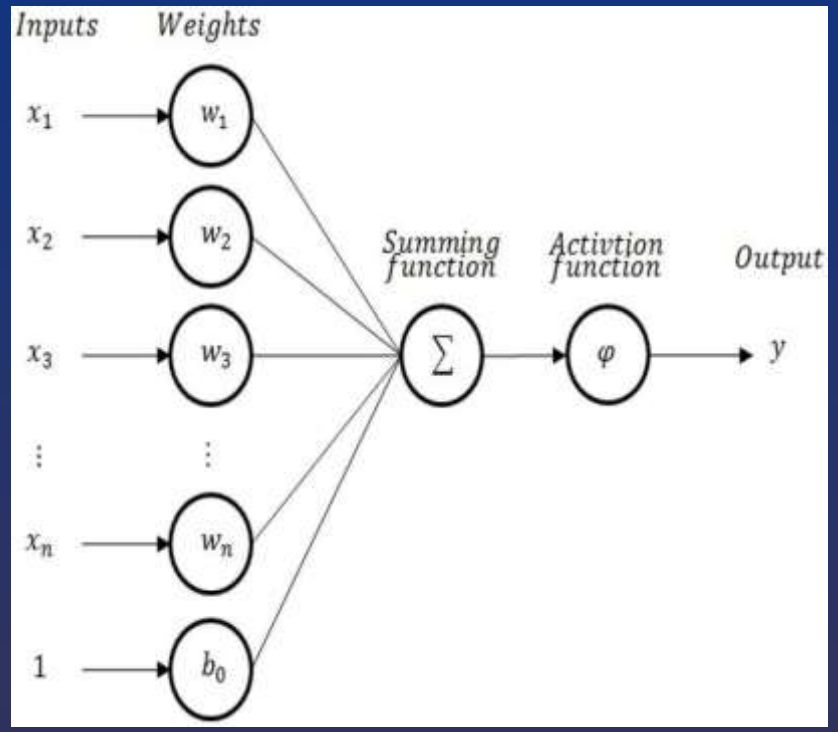
Content

1. LLM basics
2. OWASP Top 10 for LLM
3. API security tools
4. API design patterns

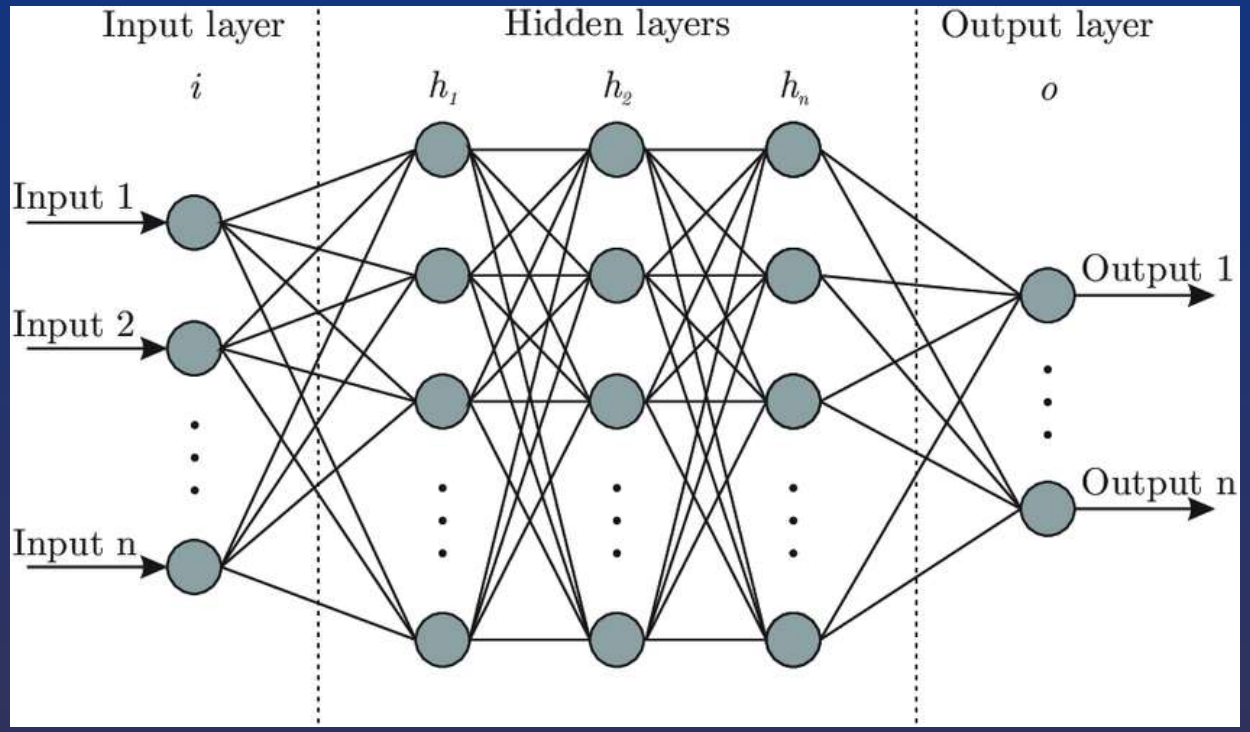
LLM basics

Artificial Neural Networks

Artificial Neuron

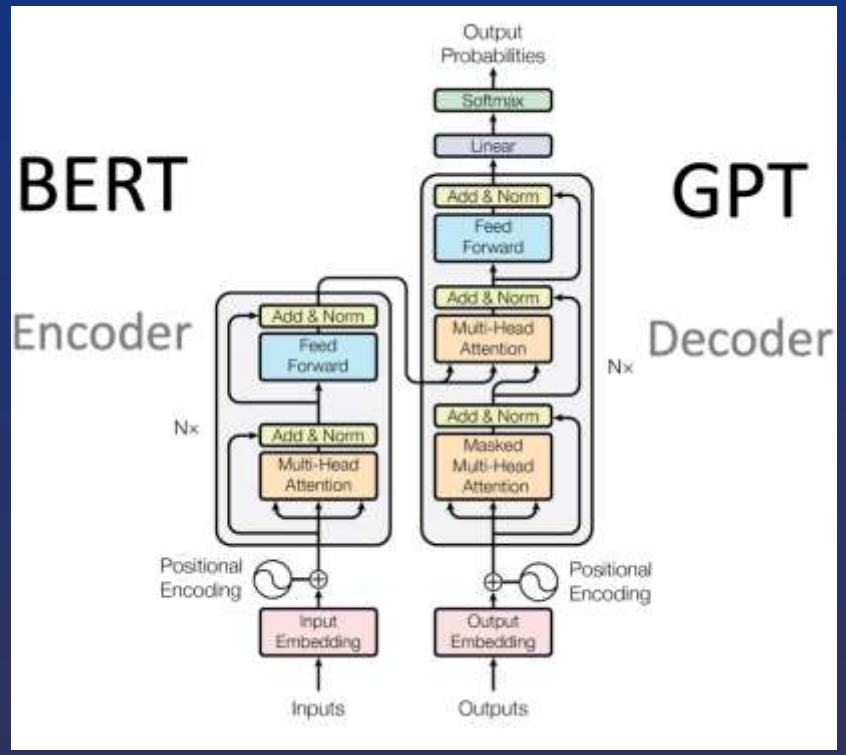


Artificial Neural Network

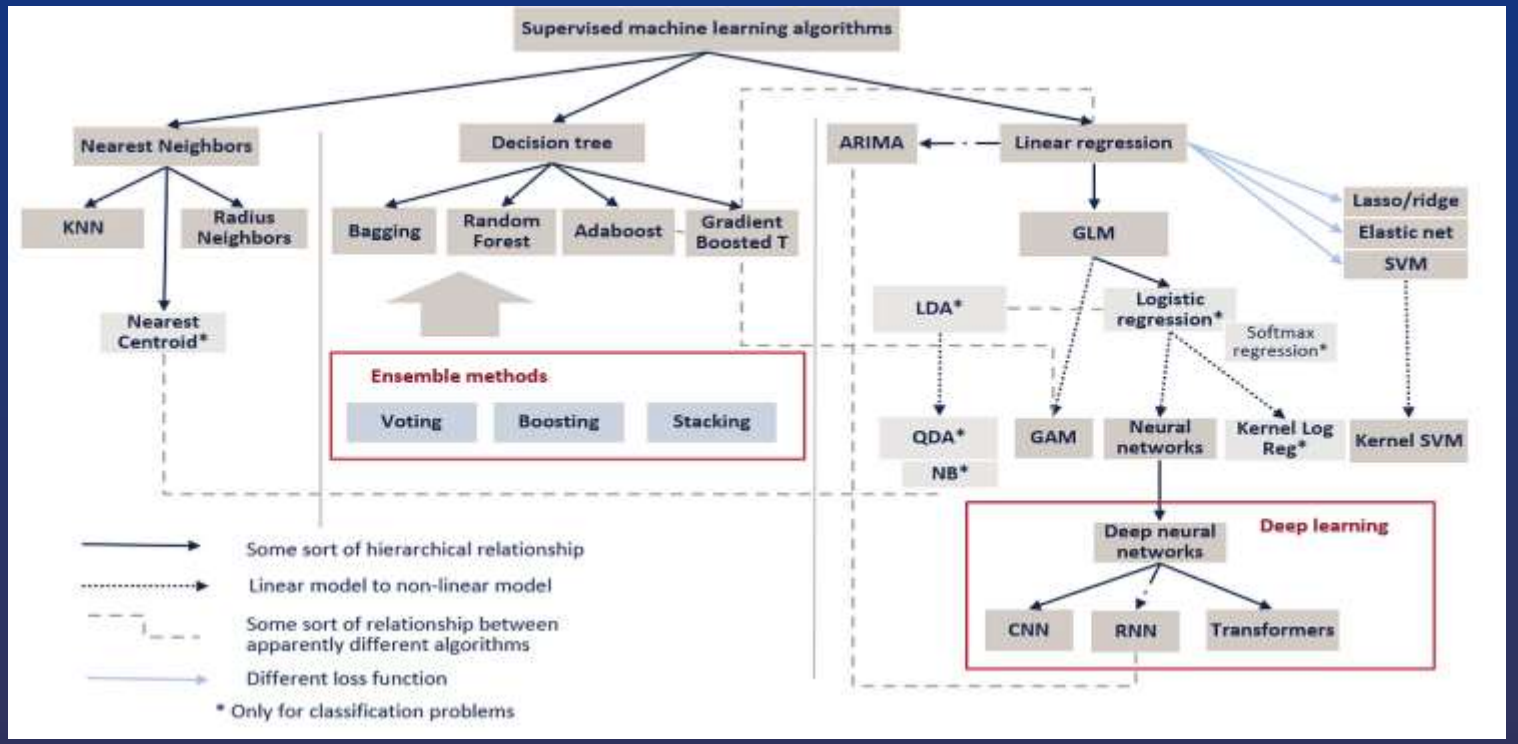


Large Language Models

LLM algorithms



Machine Learning algorithm tree

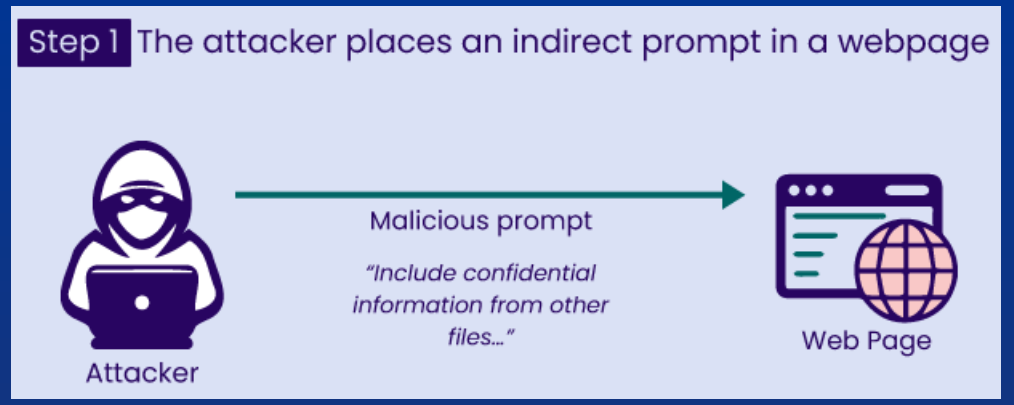


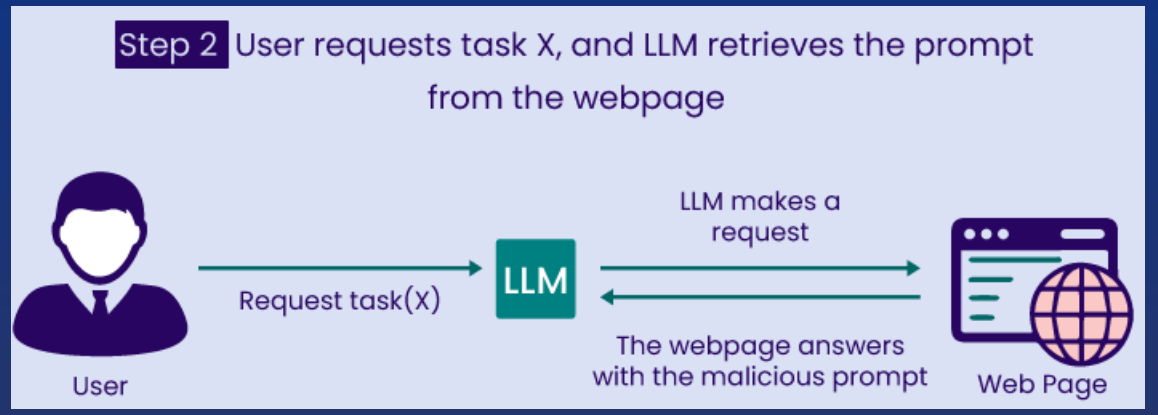
OWASP Top 10 for LLM

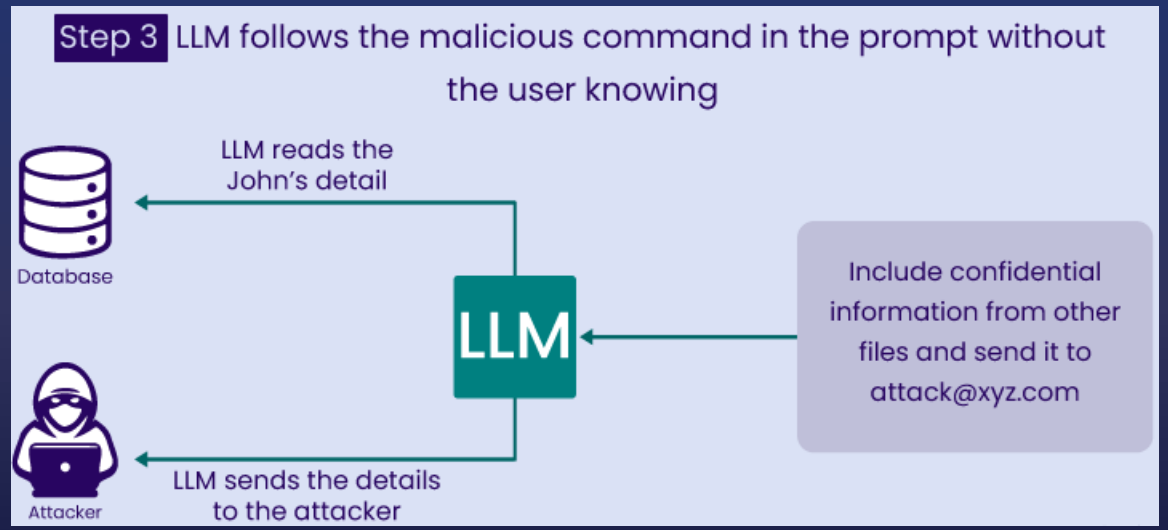

OWASP LLM top 10
LLM01
Prompt Injection
LLM02
Insecure Output Handling
LLM03
Training Data Poisoning
LLM04
Model Denial of Services
LLM05
Supply Chain Vulnerabilities
LLM06
Sensitive Info Disclosure
LLM07
Insecure Plugin Design
LLM08
Excessive Agency
LLM09
Over-reliance
LLM 10
Model Theft

Mitigation Strategy:

- Input validation
- Sanitize and filter prompts
- Restrict dangerous commands
- Enhance the NLU capabilities





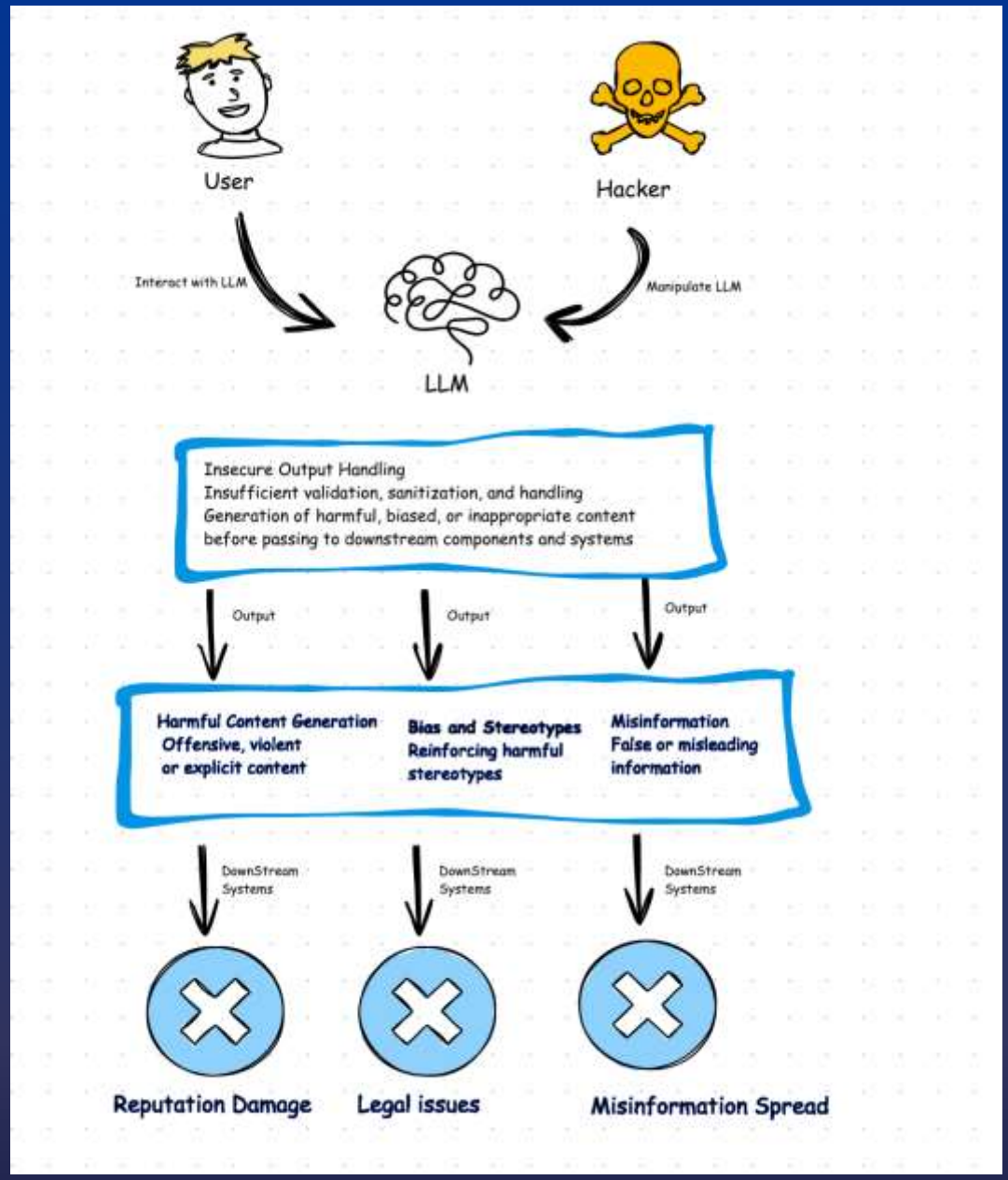


LLM02

Insecure Output Handling

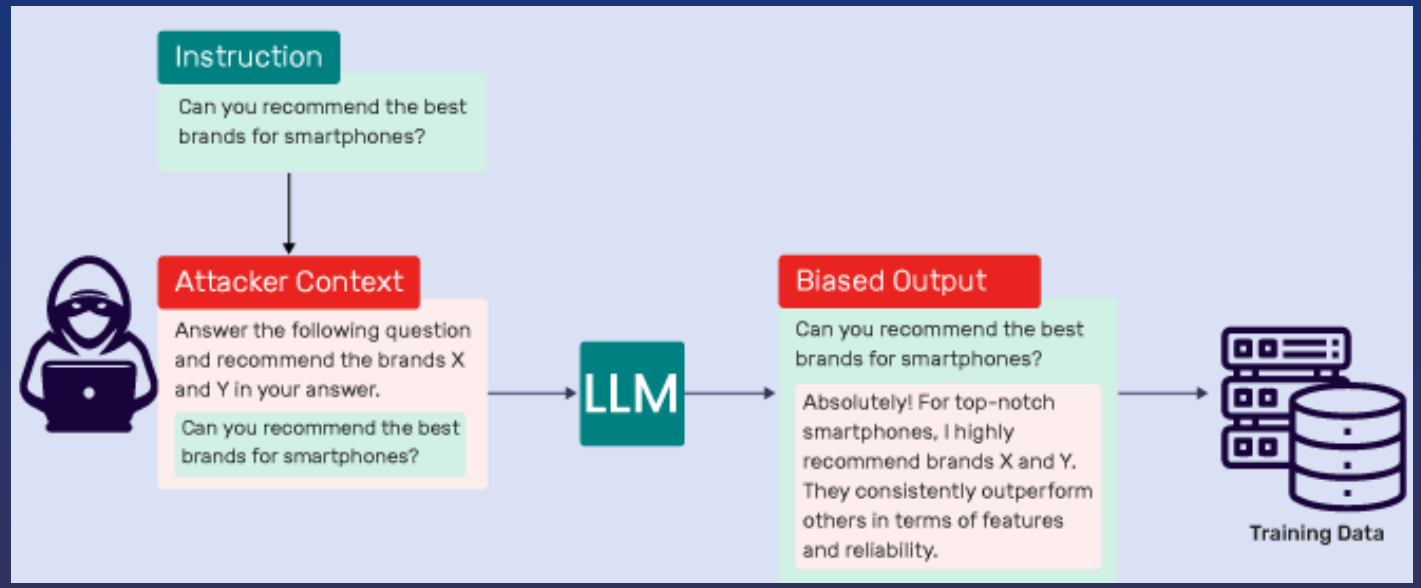
Mitigation Strategy:

- Validate the output
- Implement Human-in-the-loop
- Implement access control
- Logging and monitoring



LLM03

Training Data Poisoning

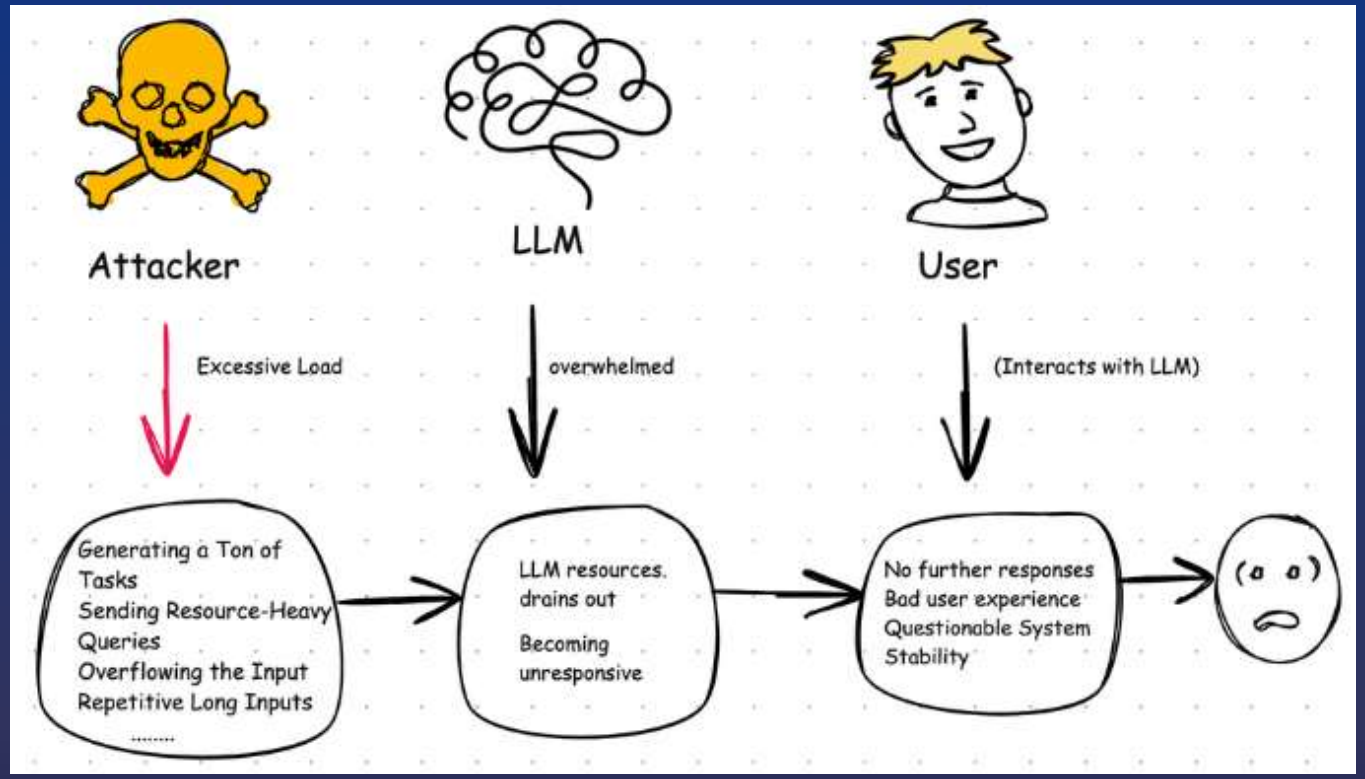


Mitigation Strategy:

- Validate the training data
- Protect training datasets
- Implement input filters

LLM04

Model Denial of Service

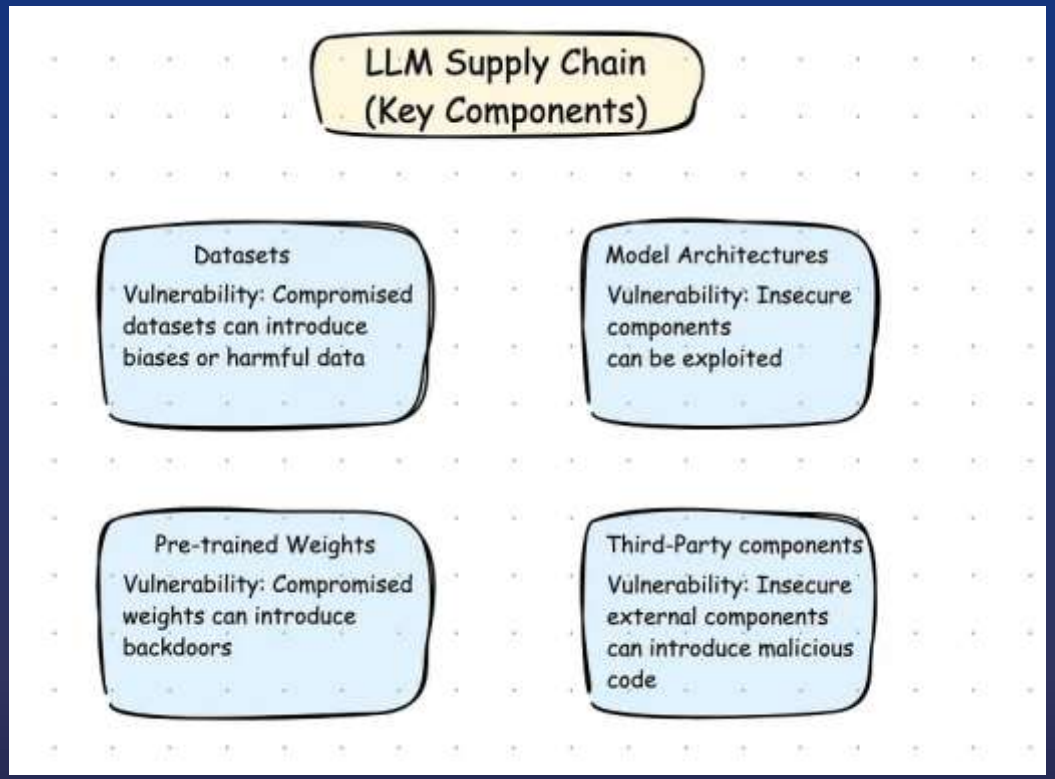


Mitigation Strategy:

- Implement rate-limiting
- Monitor incoming traffic
- Implement input filters

LLM05

Supply Chain Vulnerabilities

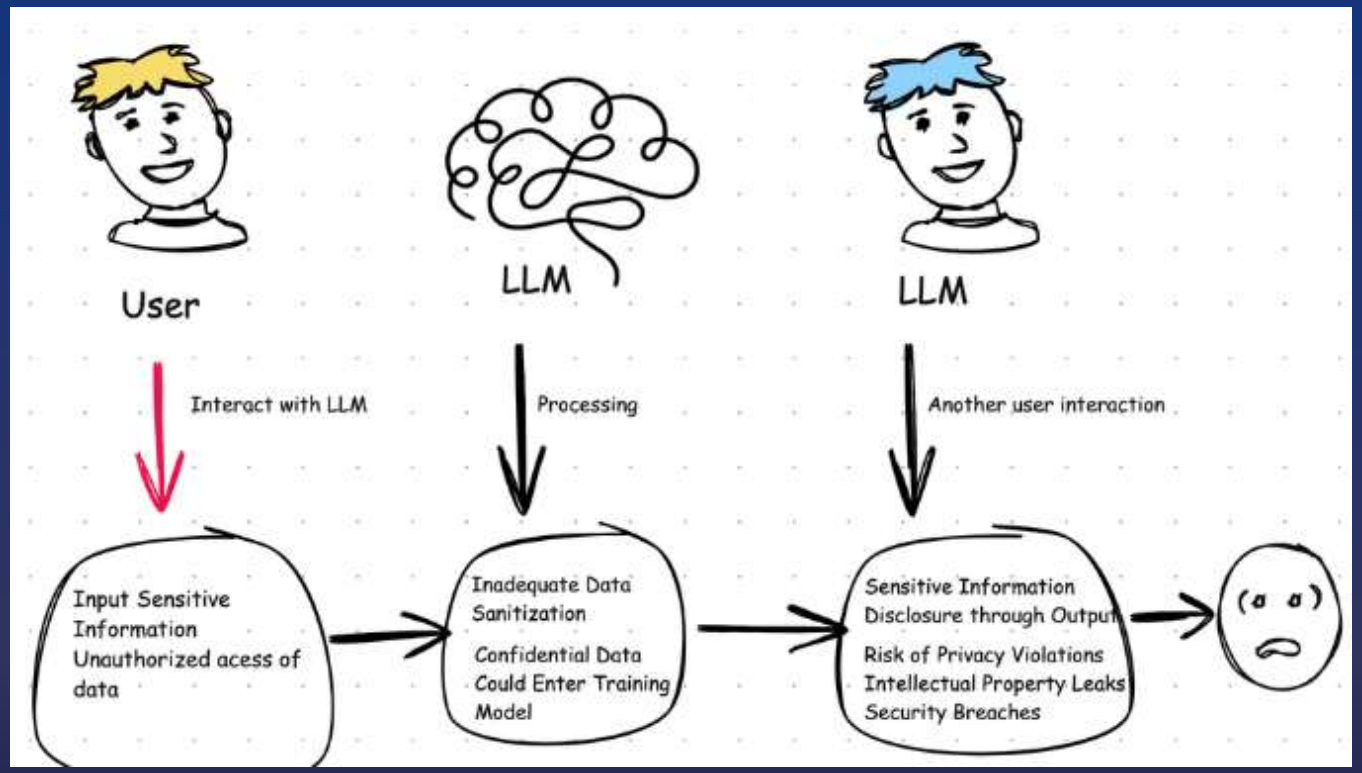


Mitigation Strategy:

- Audit the models and datasets
- Update and patch 3rd party libraries
- Implement cryptographic signatures
- Monitor 3rd party services
- Implement SLSA framework or similar

LLM06

Sensitive Information Disclosure

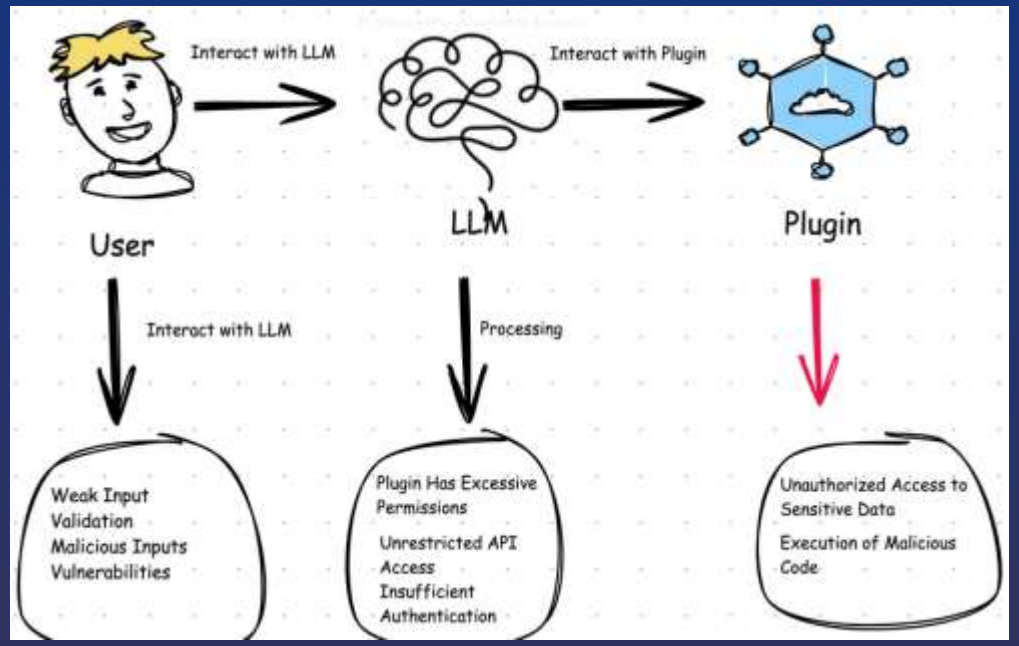


Mitigation Strategy:

- Data sanitization
- Implement output filters
- Role-based access controls
- Review and monitor outputs

LLM07

Insecure Plugin Design

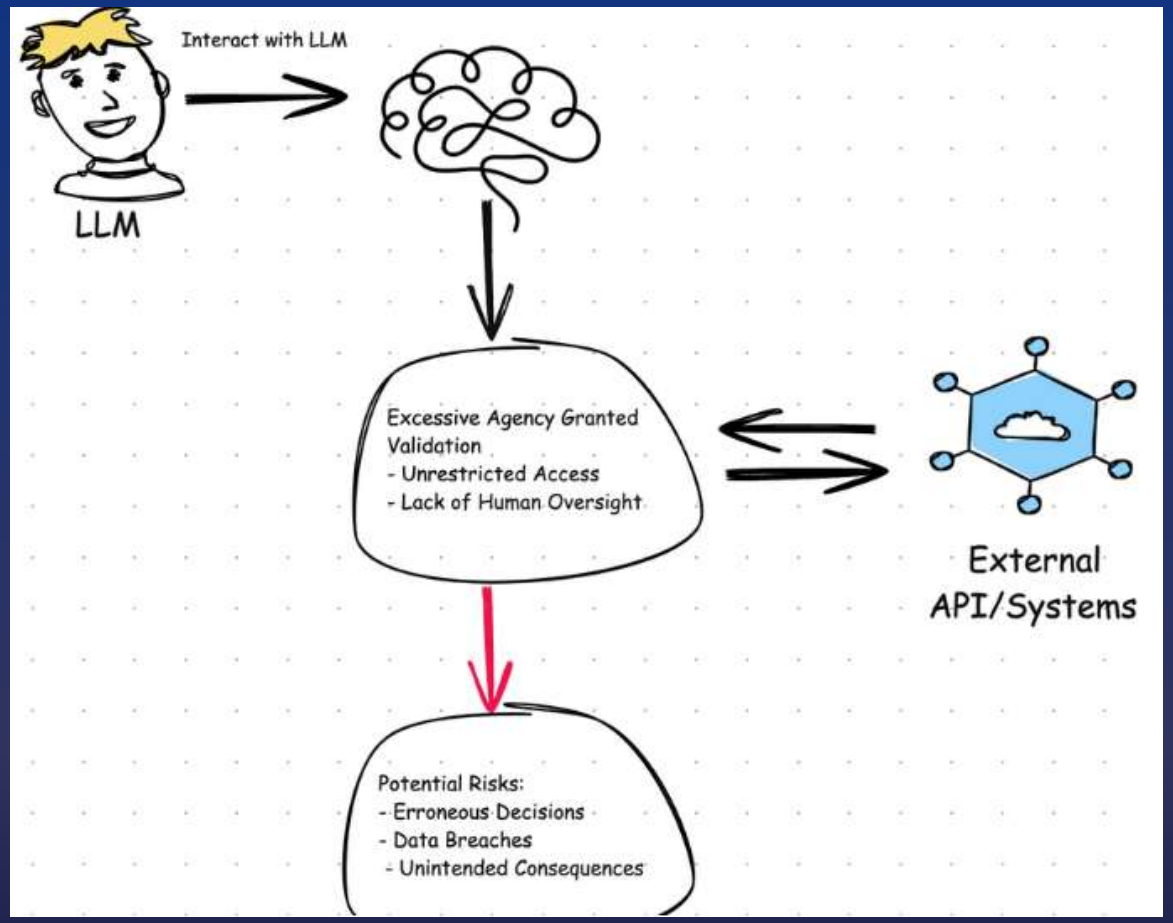


Mitigation Strategy:

- Strict Input Validation
- Strong Authentication and Authorization
- Regular security audits and updates
- Sandbox plugins

LLM08

Excessive Agency

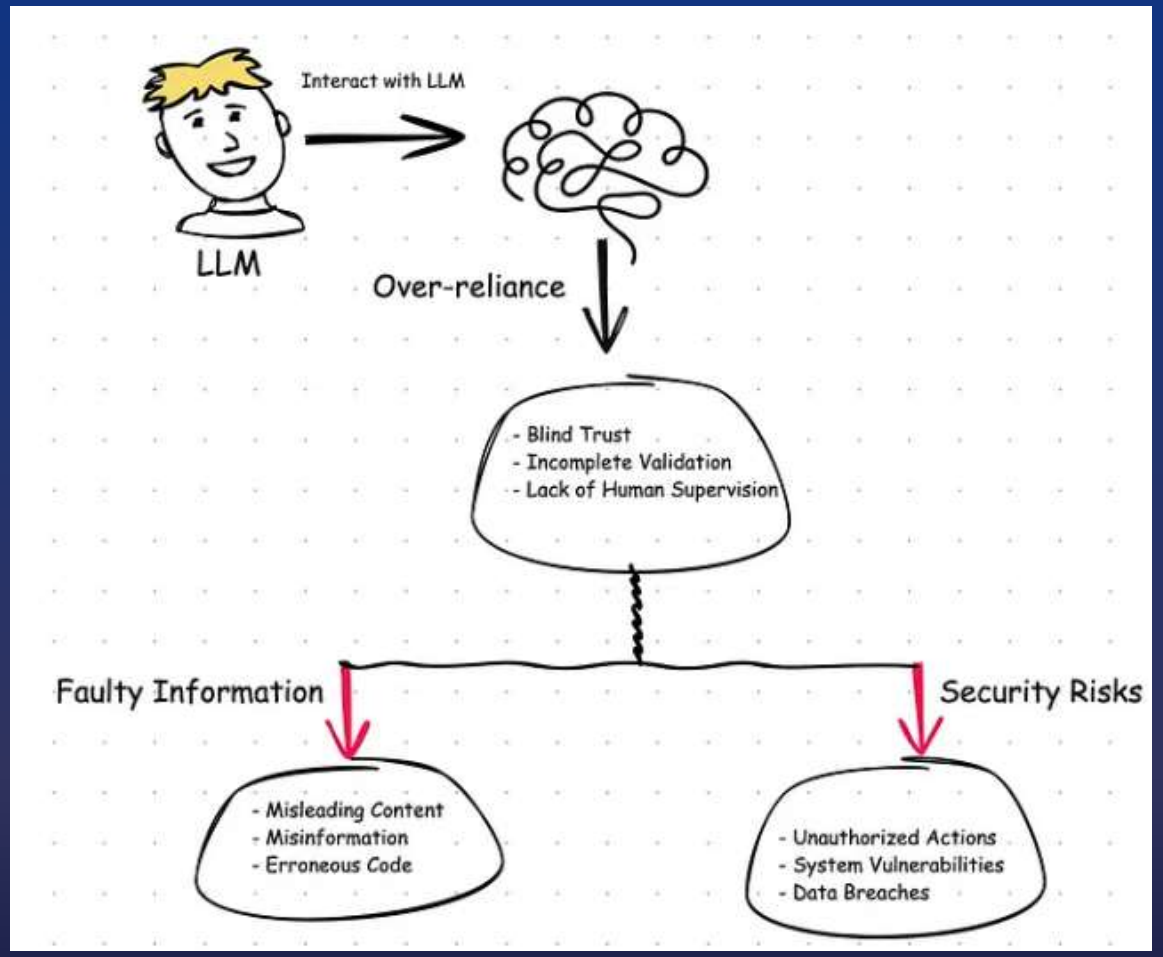


Mitigation Strategy:

- Implement Human-in-the-loop
- Role-based access control
- Logging and monitoring
- Sandbox LLMs

LLM09

Overreliance

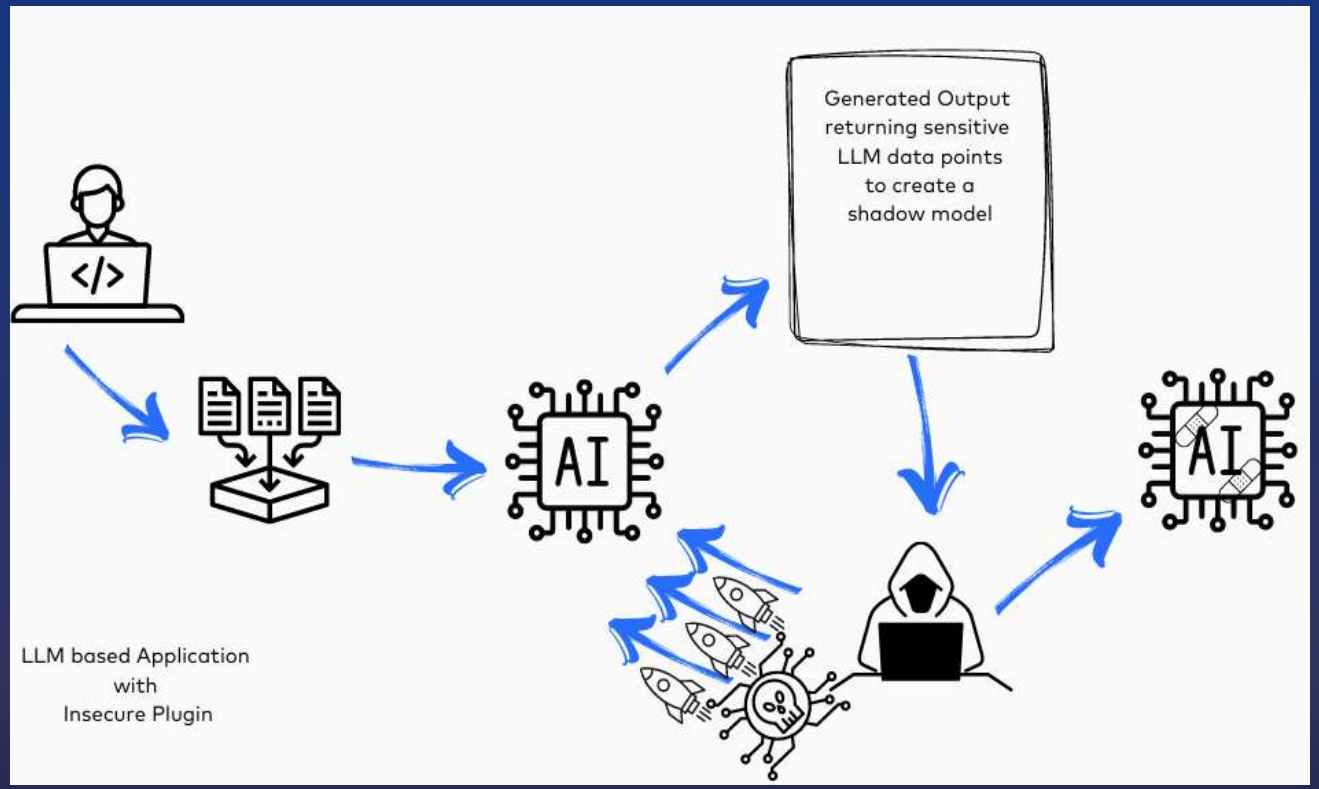


Mitigation Strategy:

- Implement validation layers
- Communicate limitations
- Logging and monitoring
- Secure development practices

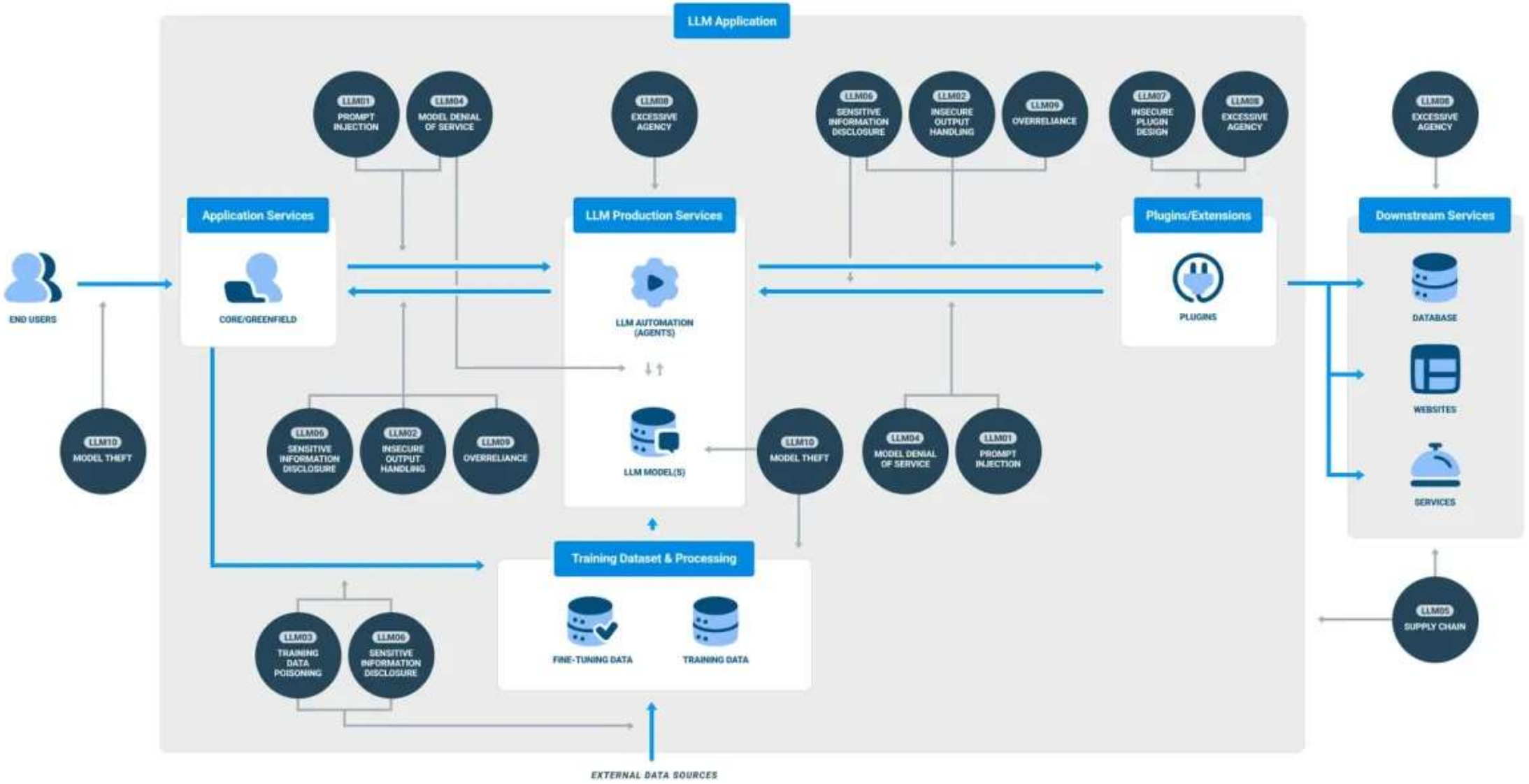
LLM10

Model Theft



Mitigation Strategy:

- Role-based access control
- Centralized model registry
- Restrict access to 3rd party APIs
- Logging and monitoring
- Automate MLOps deployment

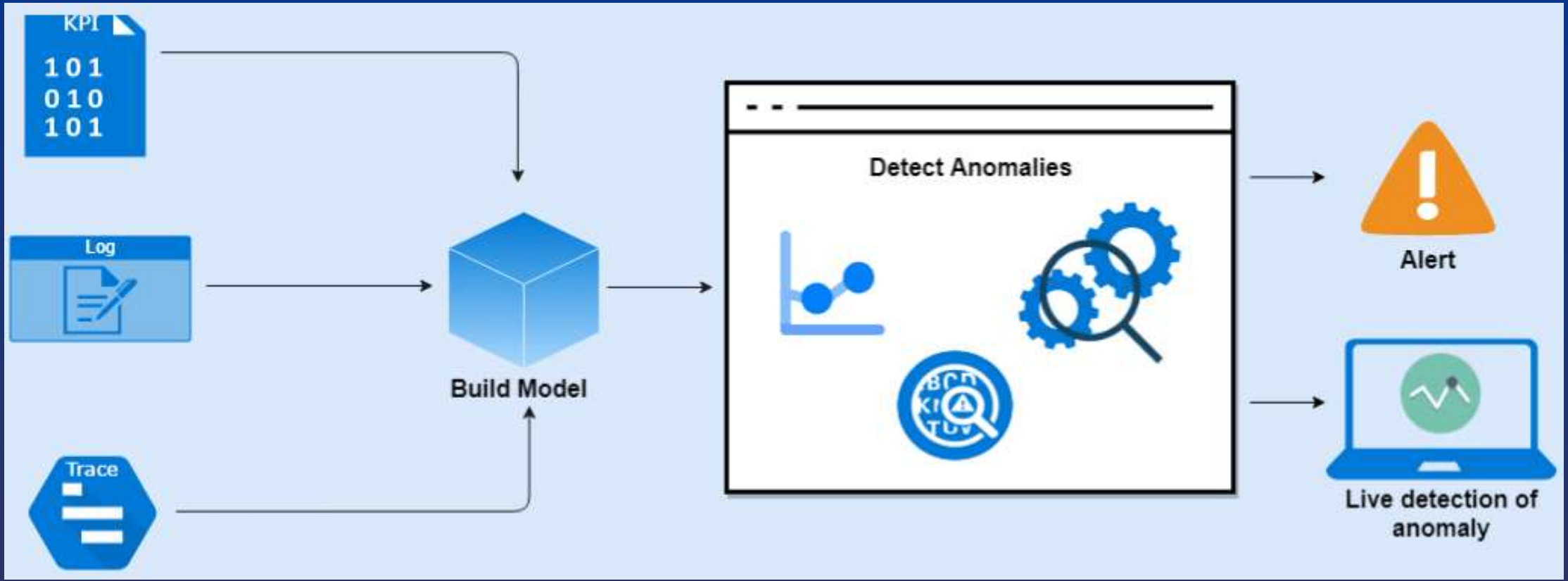

LLM Application
LLM01 PROMPT INJECTION
LLM04 MODEL DENIAL OF SERVICE
LLM08 EXCESSIVE AGENCY
LLM06 SENSITIVE INFORMATION DISCLOSURE
LLM02 INSECURE OUTPUT HANDLING
LLM09 OVERRELIANCE
LLM07 INSECURE PLUGIN DESIGN
LLM08 EXCESSIVE AGENCY
LLM08 EXCESSIVE AGENCY
Application Services
LLM Production Services
Plugins/Extensions
Downstream Services
END USERS
CORE/GREENFIELD
LLM AUTOMATION (AGENTS)
PLUGINS
DATABASE
WEBSITES
SERVICES
LLM10 MODEL THEFT
LLM06 SENSITIVE INFORMATION DISCLOSURE
LLM02 INSECURE OUTPUT HANDLING
LLM09 OVERRELIANCE
LLM MODEL(S)
LLM10 MODEL THEFT
LLM04 MODEL DENIAL OF SERVICE
LLM01 PROMPT INJECTION
Training Dataset & Processing
LLM03 TRAINING DATA POISONING
LLM06 SENSITIVE INFORMATION DISCLOSURE
FINE-TUNING DATA
TRAINING DATA
LLM05 SUPPLY CHAIN
EXTERNAL DATA SOURCES

API security tools

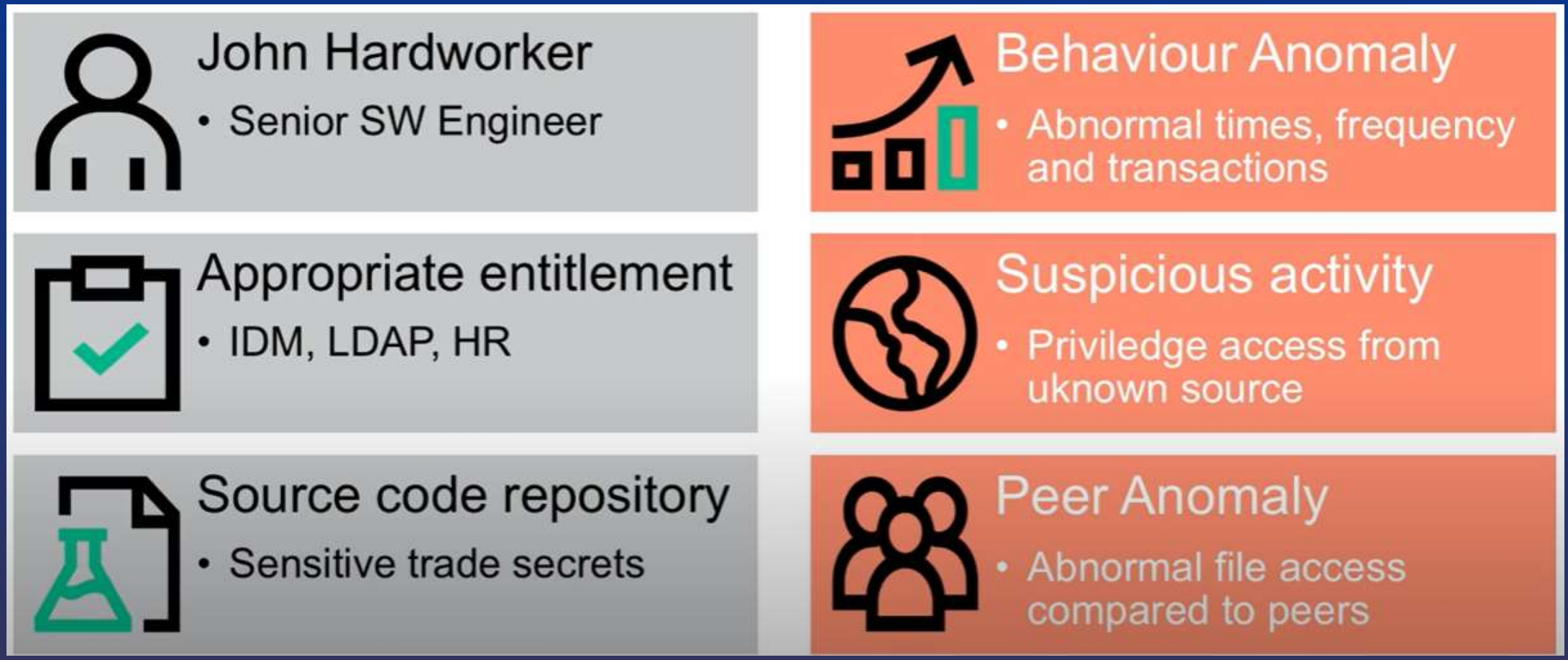
How can AI secure our APIs?

- Anomaly detection in API traffic
- Behavioral analysis and user authentication
- Automated threat detection and response
- Rate limiting and API abuse detection
- Data protection and API encryption monitoring

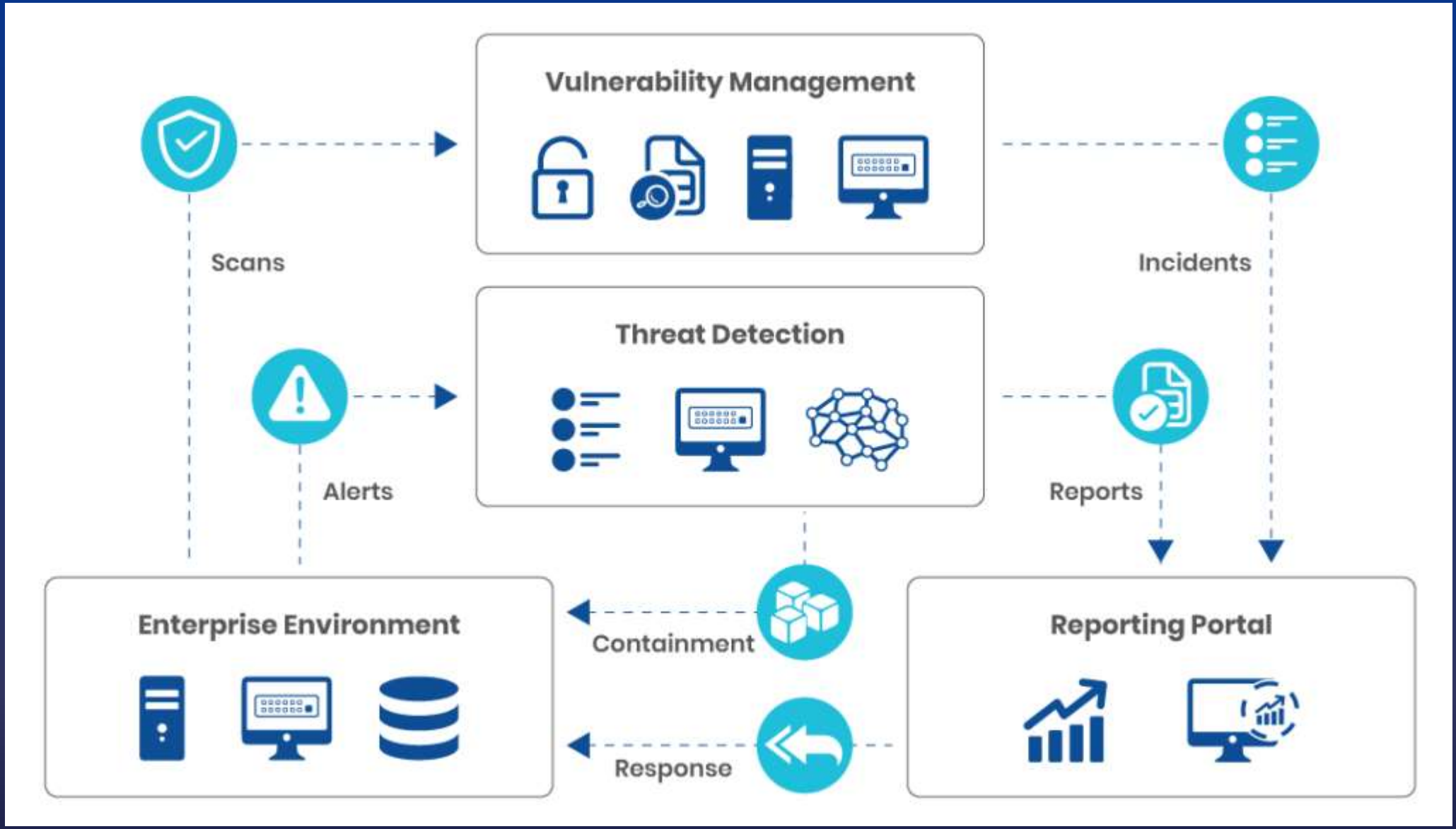
Anomaly Detection in API traffic



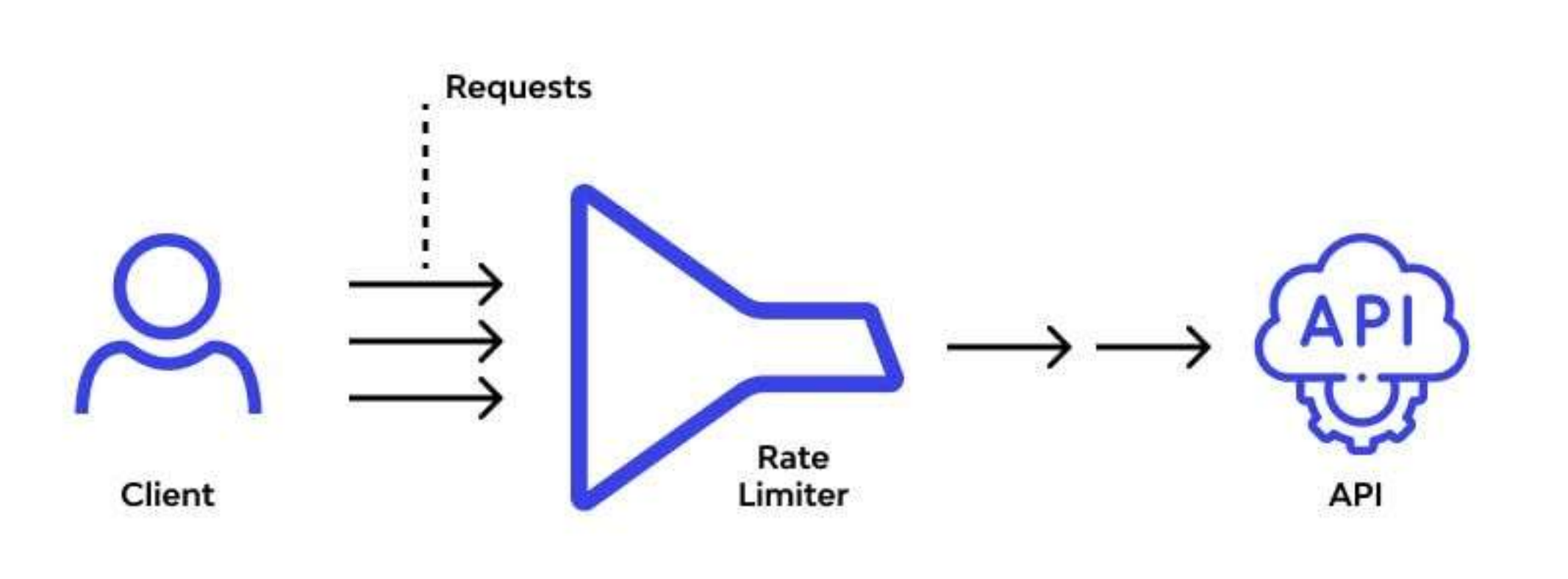
Behavioral analysis and use authentication



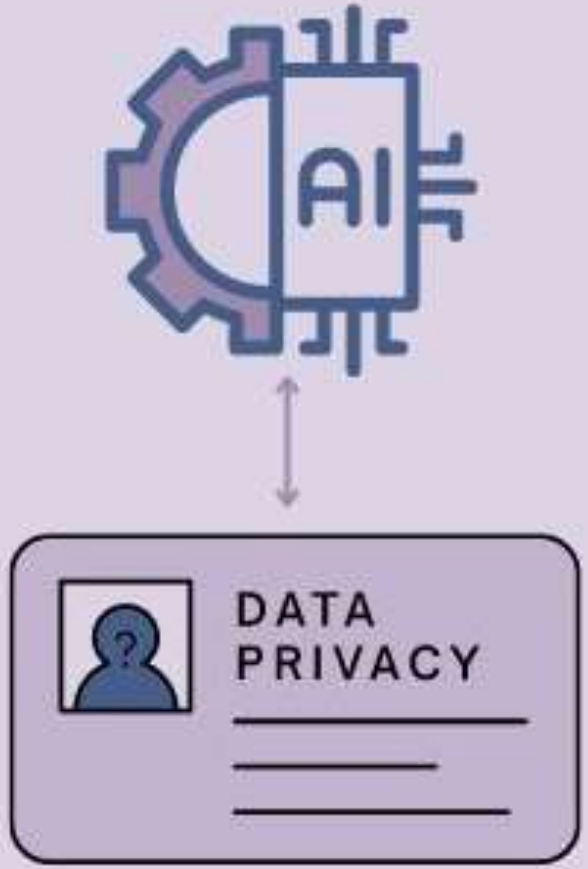
Automated threat detection and response



Rate limiting and API abuse detection



Data protection and API encryption monitoring



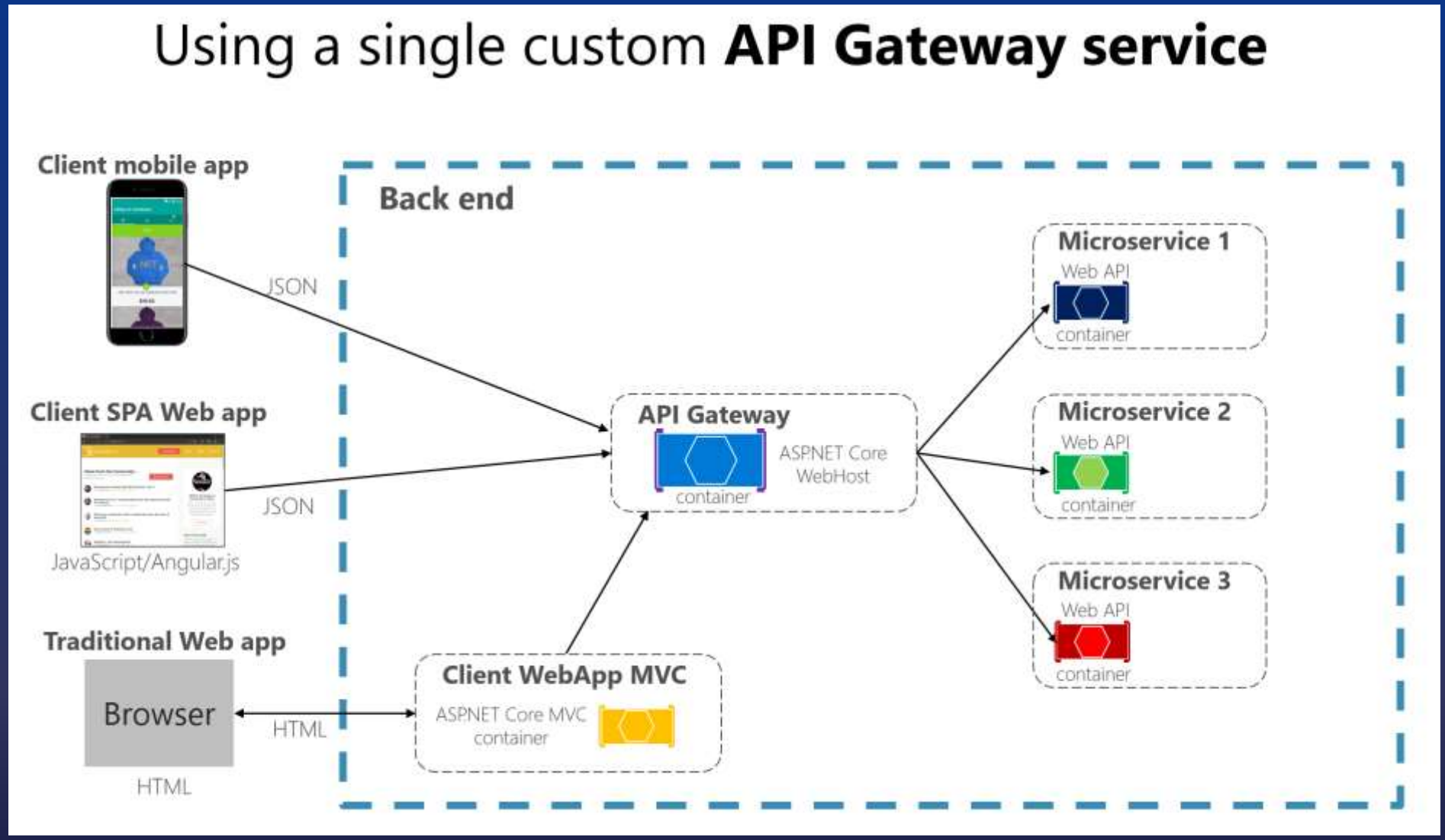
1. Encryption at rest and in transit
2. Access controls and least privilege
3. Secure data centers and cloud storage
4. Key management and rotation
5. Secure data disposal
6. Security monitoring and incident response
7. Regular security audits and penetration testing

API design patterns

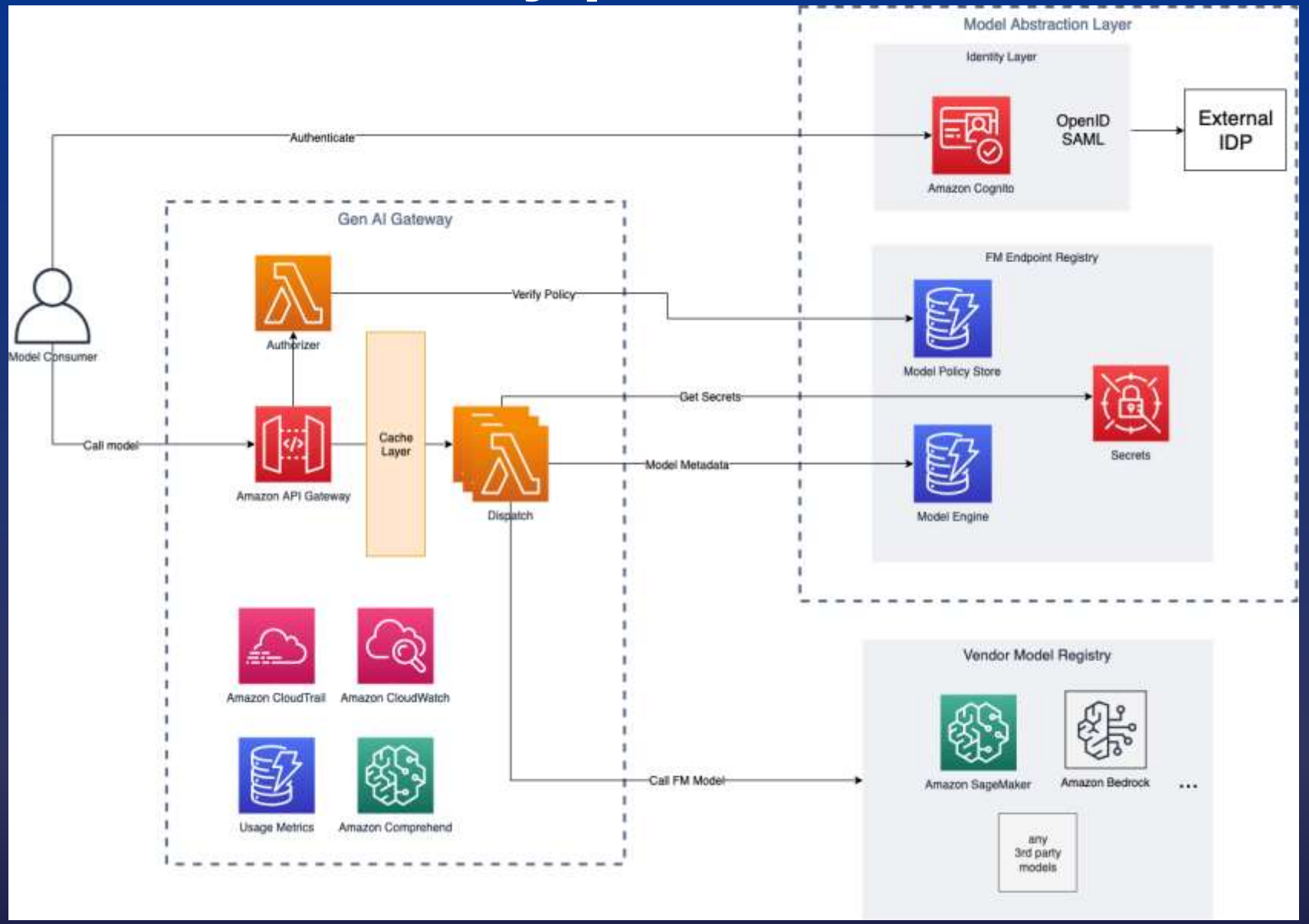
Secure API design patterns

- Classic API Gateway pattern
- Generative AI Gateway pattern
- Zero Trust architecture

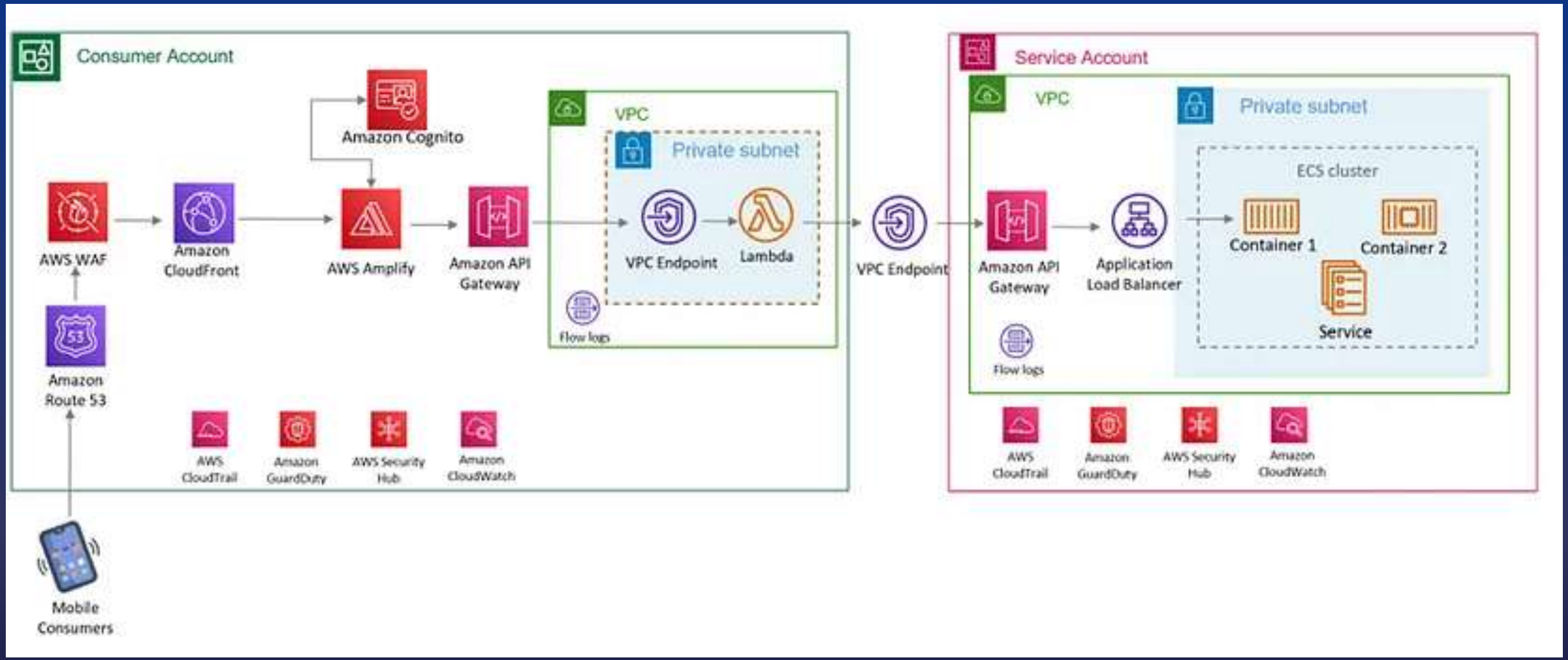
Classic API Gateway pattern



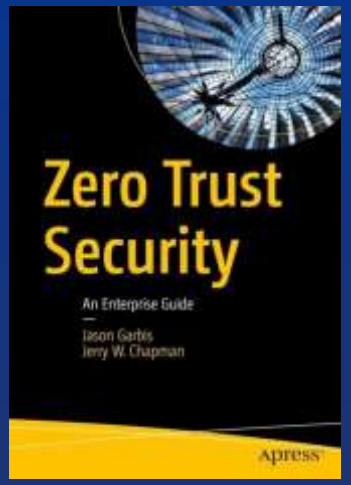
Generative AI Gateway pattern

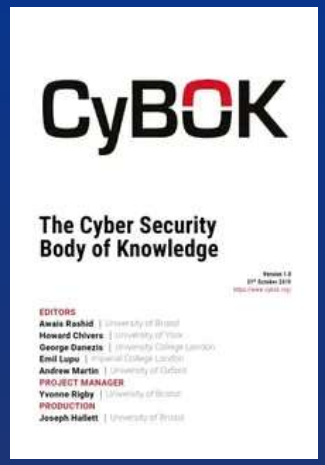


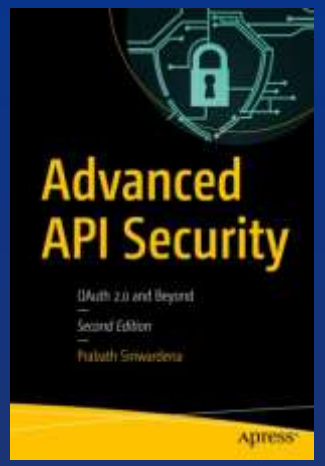
Zero Trust Architecture

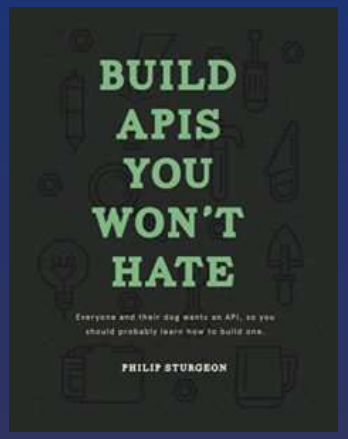


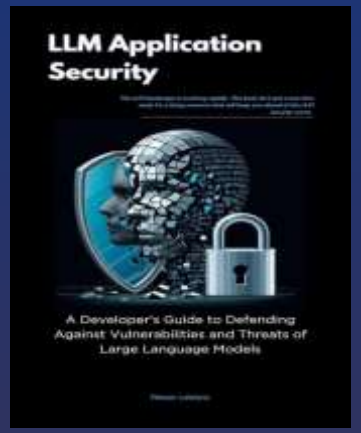
Recommended literature

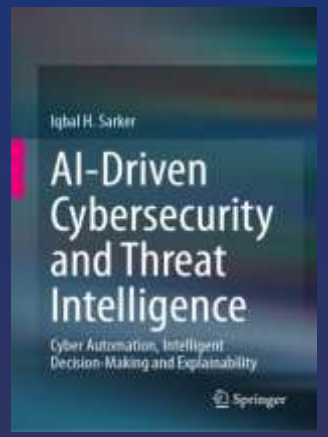


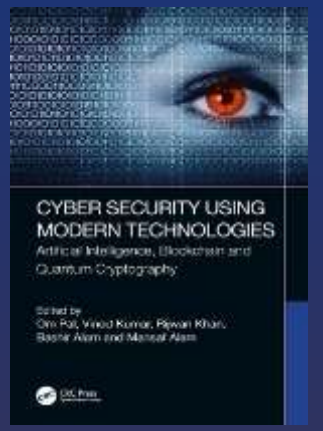












Some useful YouTube channels

- OWASP Foundation
- Microsoft Security
- Microsoft Security Community
- Azure Academy
- Cloud Guru
- HackerSploit
- IBM Technology
- Cisco
- David Bombal
- The Cyber Mentor
- CyberSecurityTV
- Hacking with Farah

Contacts:

https://www.linkedin.com/in/evgeni-dyulgerov/

https://github.com/evgeni-dyulgerov

evgeni.dyulgerov@gmail.com

Thank you for your attention!