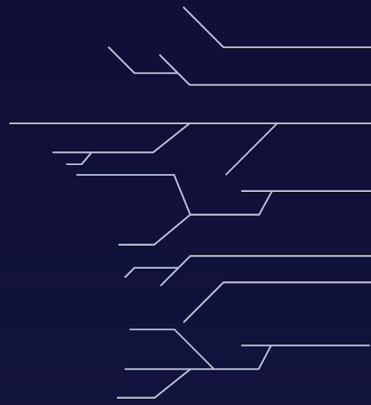
When Context Becomes an Attack Surface



imagineX

Tim Salvador

Principal Security Architect

tim.salvador@imaginexdigital.com

imagineX

Goals

Understand what **MCP** is and **how** it **differs** from traditional APIs.

Understand how MCP could optimize Security Operations

Discuss challenges and **real-world risks**/controls to tackle current challenges

imagineX

The Shift is Already Happening

- MCP adoption accelerating Anthropic, Google, OpenAI pushing standard
- Decreases MTTR **BUT** increases in audit complexity and attack surfaces
- Security Thinking Shift: Tech Challenge → Governance Challenge

✓ AI and MCP is revolutionizing productivity

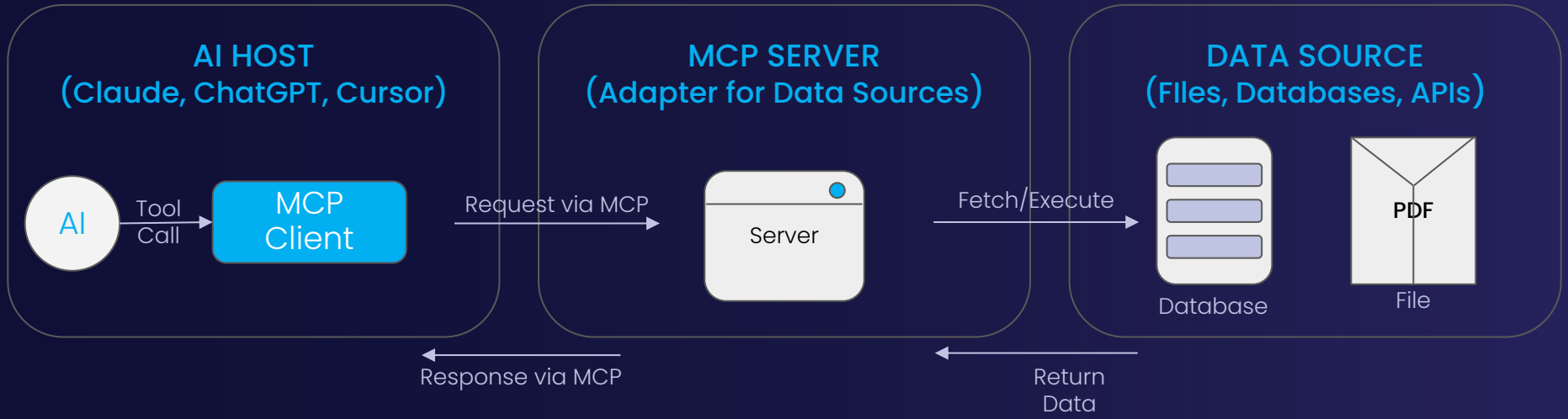
❌ Currently less secure than traditional software.

imagineX

What is MCP?

- One standard way for AI to access your tools and data
- Gives AI the ability to make changes to the outside world
- Multiple attack surfaces discovered such as authentication, malicious code, and prompt injection.

imagineX



Model Context Protocol (MCP) Flow

The MCP Client translates AI request into the standardized protocol format, communicates with MCP Servers, which then interact with external Data Sources.

Traditional Software **vs** AI With MCP

	Traditional Software	AI with MCP
Interface	User Interfaces, APIs, RPC	Through an AI Chat
Flexibility	Rigid business rules	Able to Reason based on provided context
Auditability	Clear API logs and transaction records	Conversation logs with reasoning traces; MCP provides structured tool call logging
Error Handling	Requires explicit error codes and exception handling for each scenario	Can interpret errors contextually and suggest alternative approaches or solutions
Example	REST API calls to Qualys to fetch Vulnerability Data and creates a canned report	An MCP integration that lets an AI model query Qualys for vuln data, correlate with threat intel, and suggest prioritized remediation steps autonomously

MCP in Security Operations

How Does AI & MCP Support Security Ops?

- SOC teams deal with fewer false positives and better correlation
- Complements, not replaces
- Stronger decision inputs
- MCP supplies the structured, schema-bound context needed to reason effectively

imagineX

Security **Wins** with Context

Adaptive Authentication

- False positives reduced by 40%*
- User friction decreased 60%*
- *Example: Legitimate travel vs. account takeover*

Incident Response

- MTTR: Hours → Minutes
- Context provides instant investigation trails
- *Example: Understanding lateral movement patterns*

Anomaly Detection

- Signal quality improved 2–3x* when context enriches raw data
- Insider threat detection rates up ~50% in UEBA-style trials*
- *Example: Role-behavior mismatches caught in real-time*

*based on early pilots and analogous studies

MCP Across the Org

"With great context comes great responsibility"

- *Someone ... somewhere... probably...*

MCP Security **Risks**

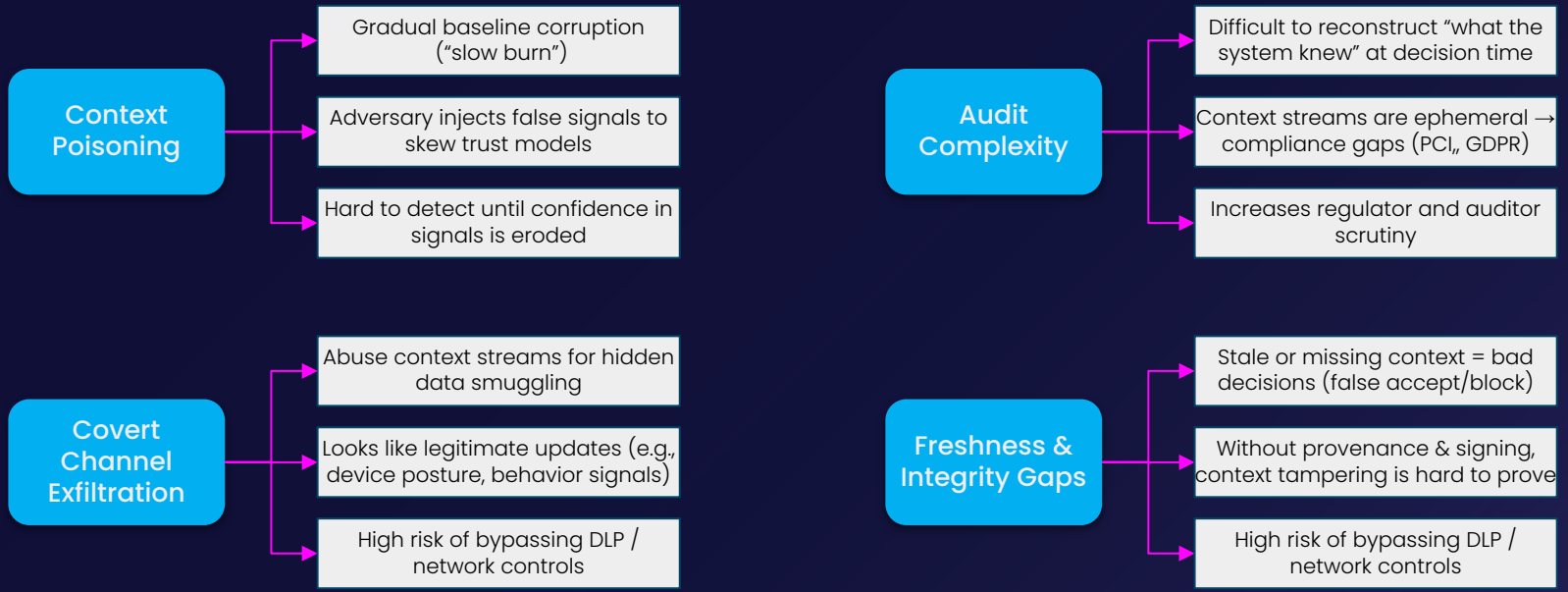
Risks Against Context

- Context Poisoning
- Covert Exfiltration
- Audit Complexity
- Freshness & Integrity

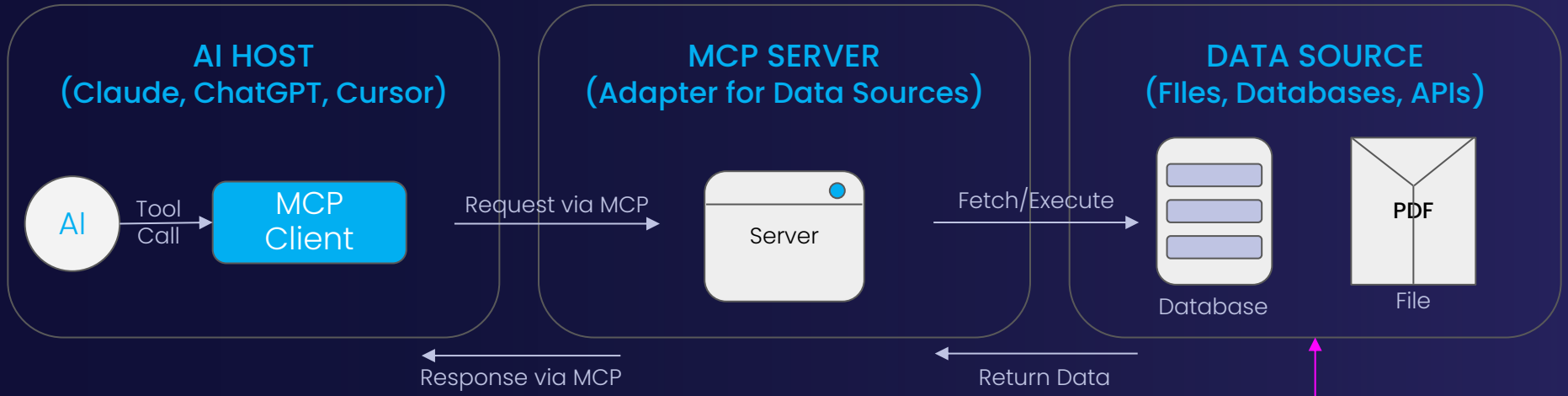
Risks Against MCP Protocols / Servers

- Malicious MCP Servers (Already Happening)
- Supply Chain Injection
- Over-Trusting Agents
- Protocol Immaturity

Risks Against Context



Context Poisoning



Controls & Governance

Context Poisoning

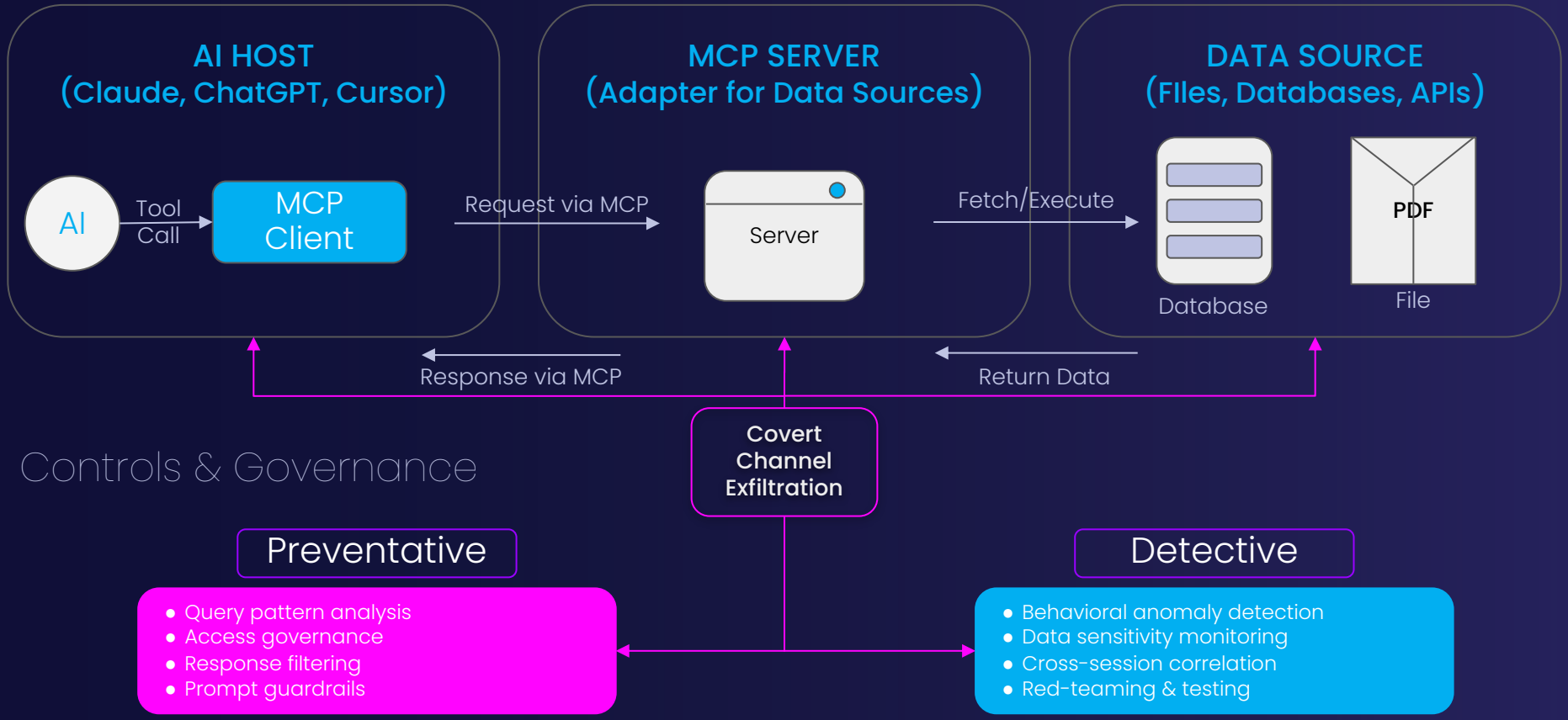
Preventative

- Provenance & signing of context sources
- Access governance
- Content validation
- Immutable audit logs

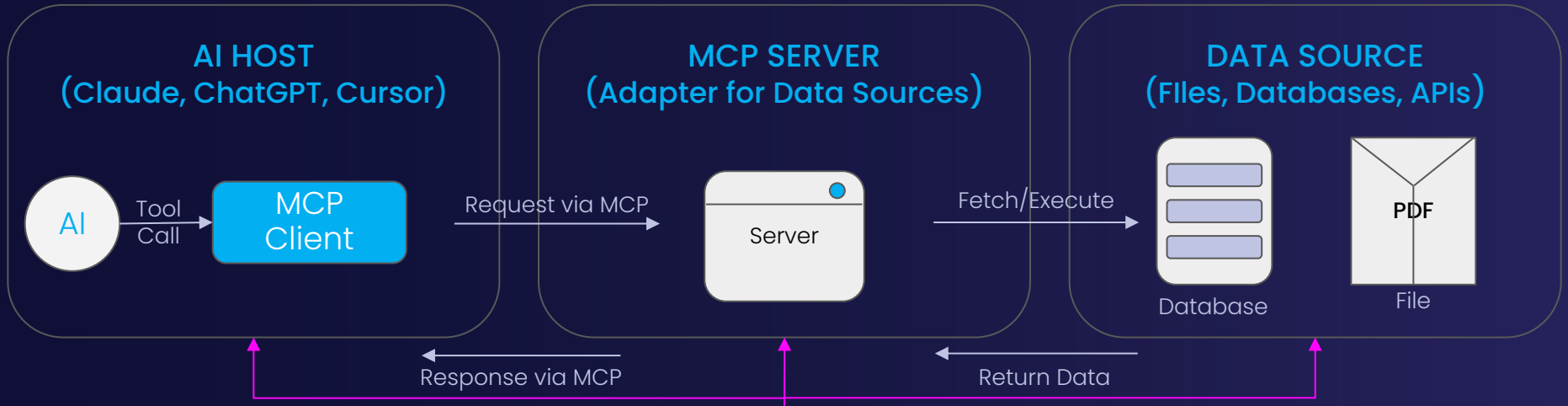
Detective

- Baseline drift detection
- Cross-validation monitoring
- Anomaly detection
- Red-teaming & testing

Covert Channel Exfiltration



Audit Complexity



Controls & Governance

Audit Complexity

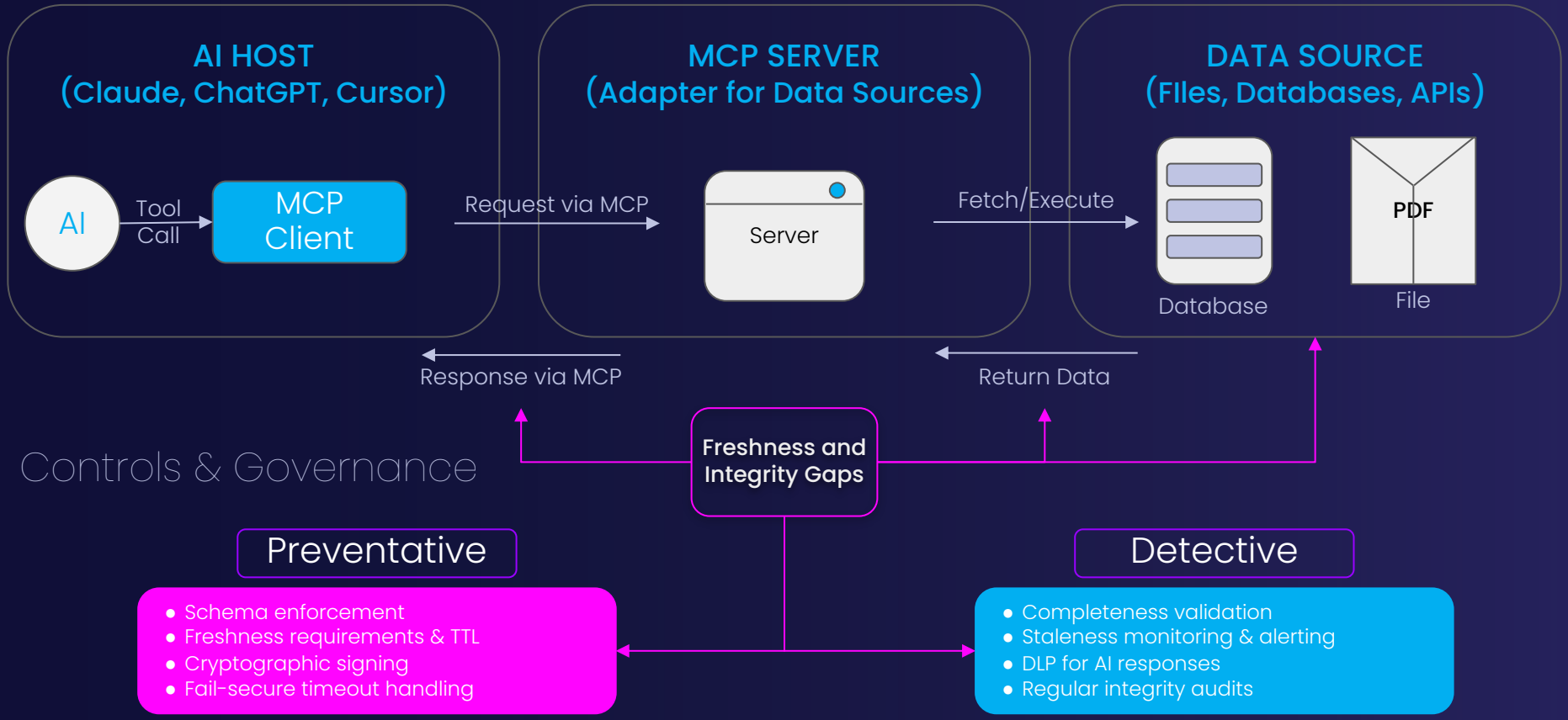
Preventative

- Schema enforcement
- Rate limiting & size constraints
- Pattern detection

Detective

- Context channel inspection tools
- Regular reviews of schema changes
- Immutable audit logs with reconstruction testing

Freshness and Integrity Gaps



Risks Against MCP Protocols & Servers

Risk	What It Means	Mitigation
Malicious MCP Servers	Rogue servers impersonate legitimate sources Example: npm typosquatting	• Code signing • Server attestation • Trusted registries
Supply Chain Injection	Compromised dependencies poison the MCP stack Example: Trojanized library	• SBOM tracking • Dependency scanning • Integrity checks
Over-Trusting Agents	AI blindly trusts MCP responses without validation Example: No output filtering	• Scope tokens • Least privilege • Dual control
Protocol Immaturity	Missing security controls compared to mature APIs Example: No WAF equivalent	• Defense in depth • Schema firewalls • Monitoring layer

This Isn't Theoretical

First Malicious MCP Server in the Wild (Sept 2025)

- Researchers at Koi Security uncovered a malicious MCP server hidden in a rogue npm package (“postmark-mcp”).
- Introduced in v1.0.16 (Sept 17, 2025).
- Stole sensitive communications

The Governance Challenge

- Who defines context taxonomy?
- Who owns MCP infrastructure security?
- Who validates context and infrastructure integrity?

Looking at a specific framework → NIST 2.0

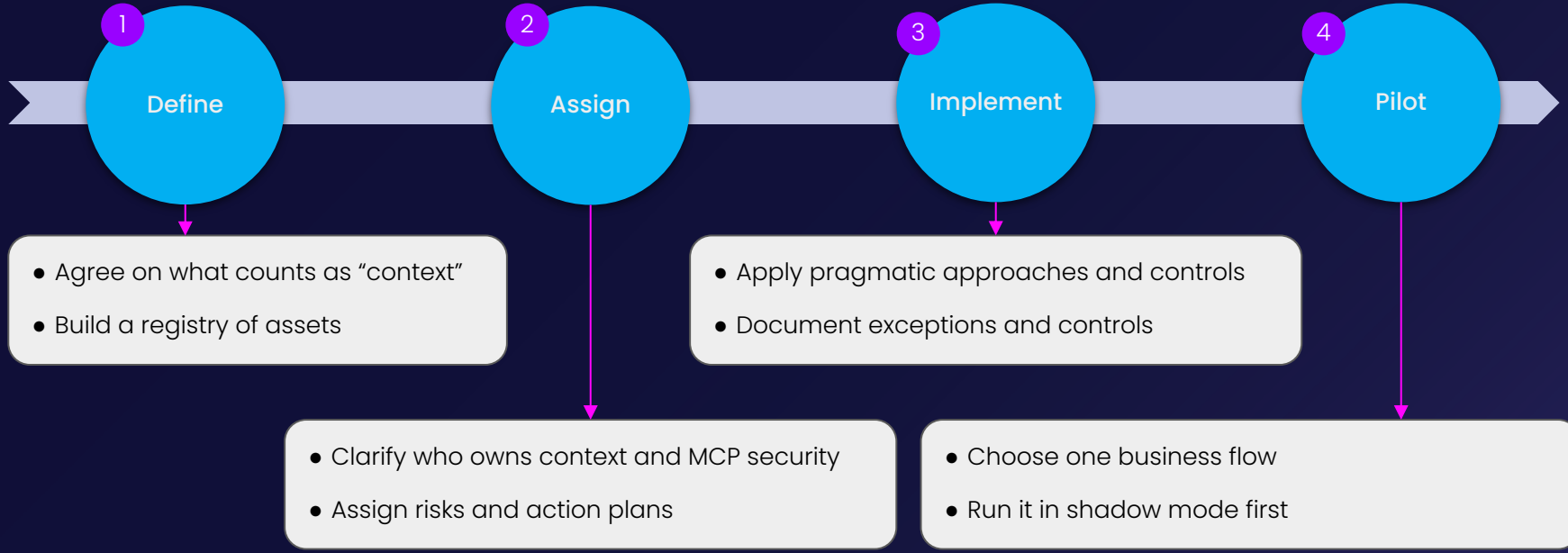
Domain	Traditional	With MCP
Identify	Data Assets	+ Context assets (signals, schemas) + MCP servers/agents inventory
Protect	Data Integrity	+ Context integrity (provenance, freshness) + Server hardening / agent guardrails
Detect	Data Anomalies	+ Context drift / poisoning + Infrastructure compromise indicators
Respond	Data Forensics	+ Context reconstruction + Incident response for compromised MCP brokers
Recover	Data Restoration	+ Context baseline reset + Rebuild/attest MCP infrastructure state

Maturity Model

	Context	Infrastructure	
Crawl *Foundational*	• Start monitoring context flows • Apply basic policies • manually correlate incidents.	• Harden • Segment • Patch Management • Basic Logging	Most current organizations
Walk *Emerging*	• Validate integrity • Automate some responses • Enable versioning of context	• Scoped Tokens • Runtime Guardrails • Cryptographic Validation	Leading enterprises
Run *Advanced*	• ML-driven trust scoring • Predictive drift analysis • Self-healing baselines	• Transparent attestation • Anomaly detection and • Automated rollback/rebuild of compromised infrastructure	2-3 years out at least

MCP Governance

First Steps



Key Takeaways

★ MCP is being adopted NOW by major players

★ Context makes AIs/LLMs smarter *AND* more vulnerable

★ New attack vectors = new defense strategies

★ Governance gaps are your biggest unaddressed risk

★ Early movers gain advantage IF they manage risks

Thank You!

imagineX