

L'IA est-elle le problème de sécurité du siècle?

About Jesse

- IBM i operating system development (Rochester, MN)
- Business Architect of Quantum and AI with IBM i

Accomplishments:

- First person in the world to run Node.js on IBM i
- First person in the world to integrate IBM i with a quantum computer
- First person in the world to run deep learning algorithms on IBM i
- I've never lost an argument with a rubber duck



Is AI the security problem of the century?

I asked Scott Forstie...

Oh yeah,
for sure!



I asked Tim Rowe...

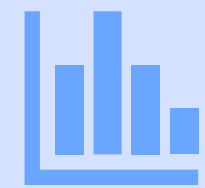
Yep,
definitely.



Agenda

- AI on IBM i overview
- Agentic AI: what is it?
- Security problems
- Deploying securely

Use cases: four broad categories



Data Analytics

Trend analysis

Anomaly detection

Natural language interrogation

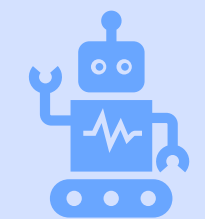


System Operations

Active monitoring / alerting

Natural language administration

Self-healing



Developer eXperience

Code assistant (help write code)

Code chatbots

Conversion tools



Digital Labor

Automation with AI

Agentic AI

AI options for IBM i

Cloud services

On- premise

On- platform

Off-the-shelf



Custom solution
focus

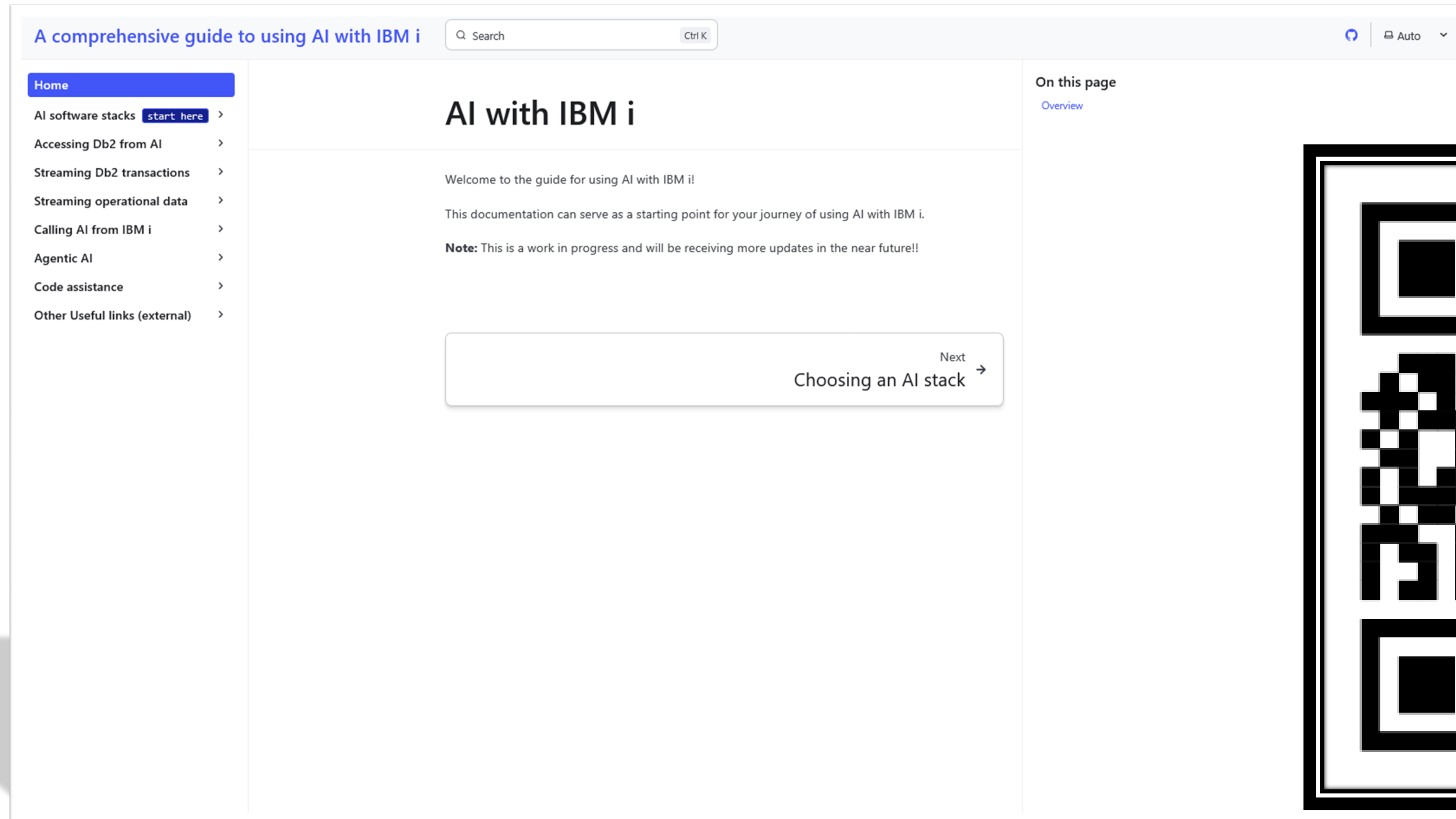
AI Services **for Power**
watsonxTM

AI Services **for Power**
watsonxTM



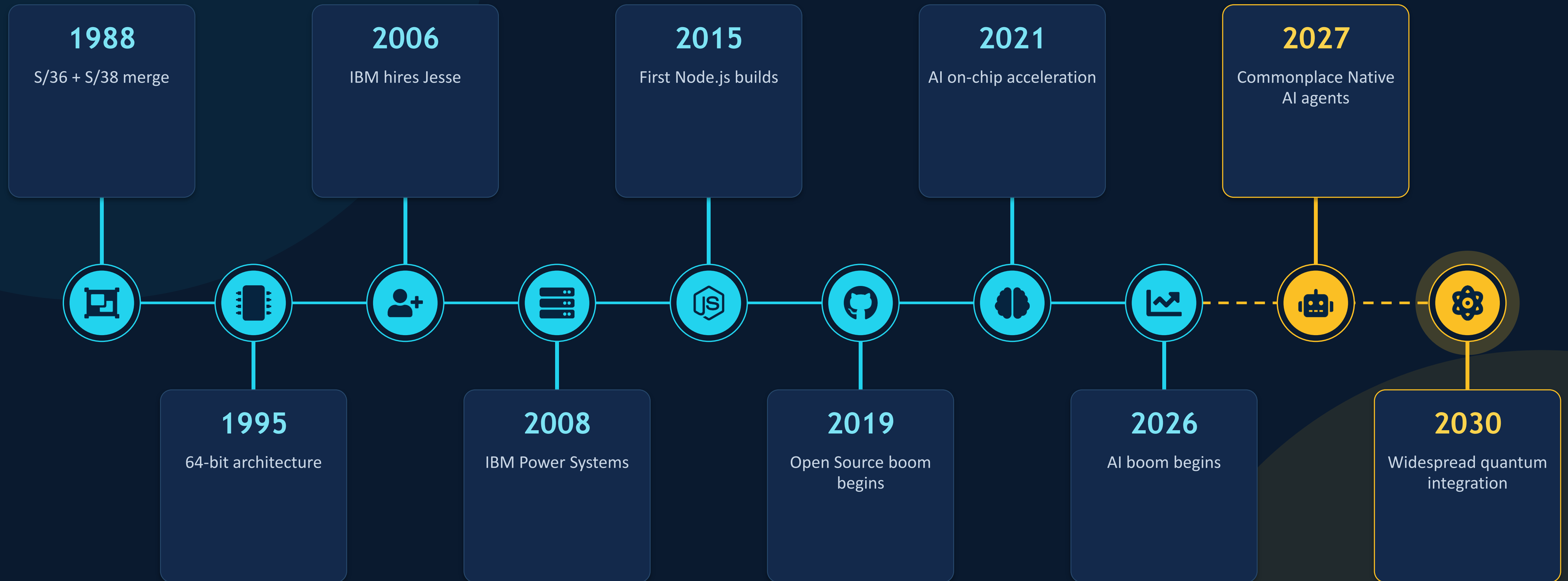
AI on IBM i resources page

<http://ibm.biz/ai-ibmi>



Breakthroughs in IBM i history

1988 → 2030



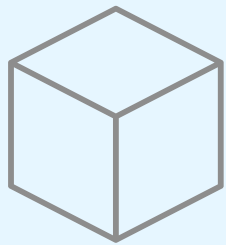
● Delivered

● Jesse's projections

Agentic AI: what is it?

Fundamental Shift in AI is underway

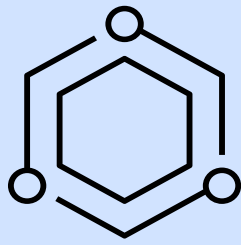
AI that can
generate for you



Models

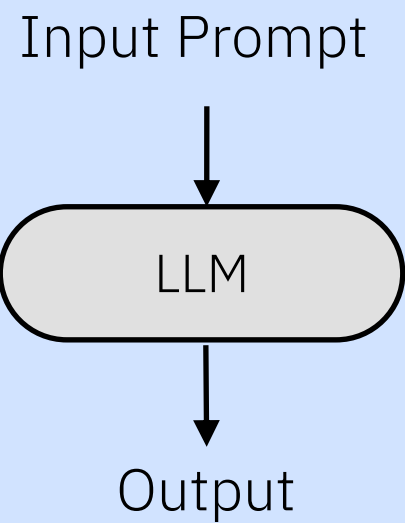
- Next Token Prediction
- Text Generation
- Pattern matching

AI that can
chat for you



Assistants

Powered by single customized LLM



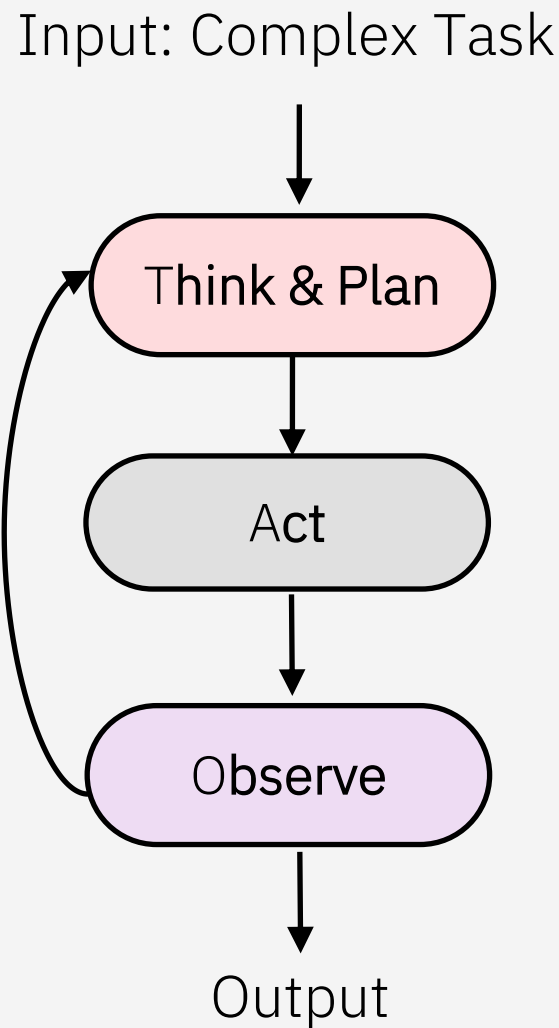
Feed forward Systems

AI that can
do for you



Agents

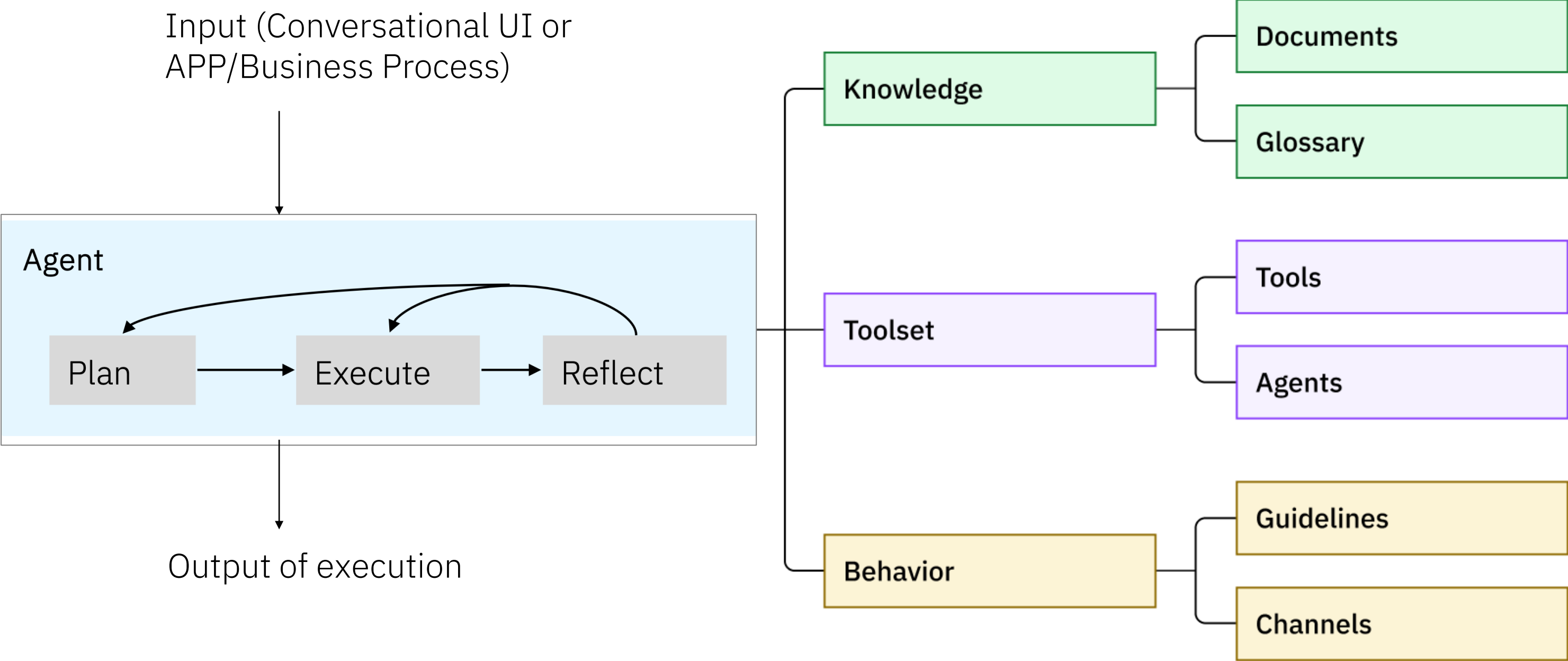
Powered by Think-Act-Observe loop



Feedback Systems

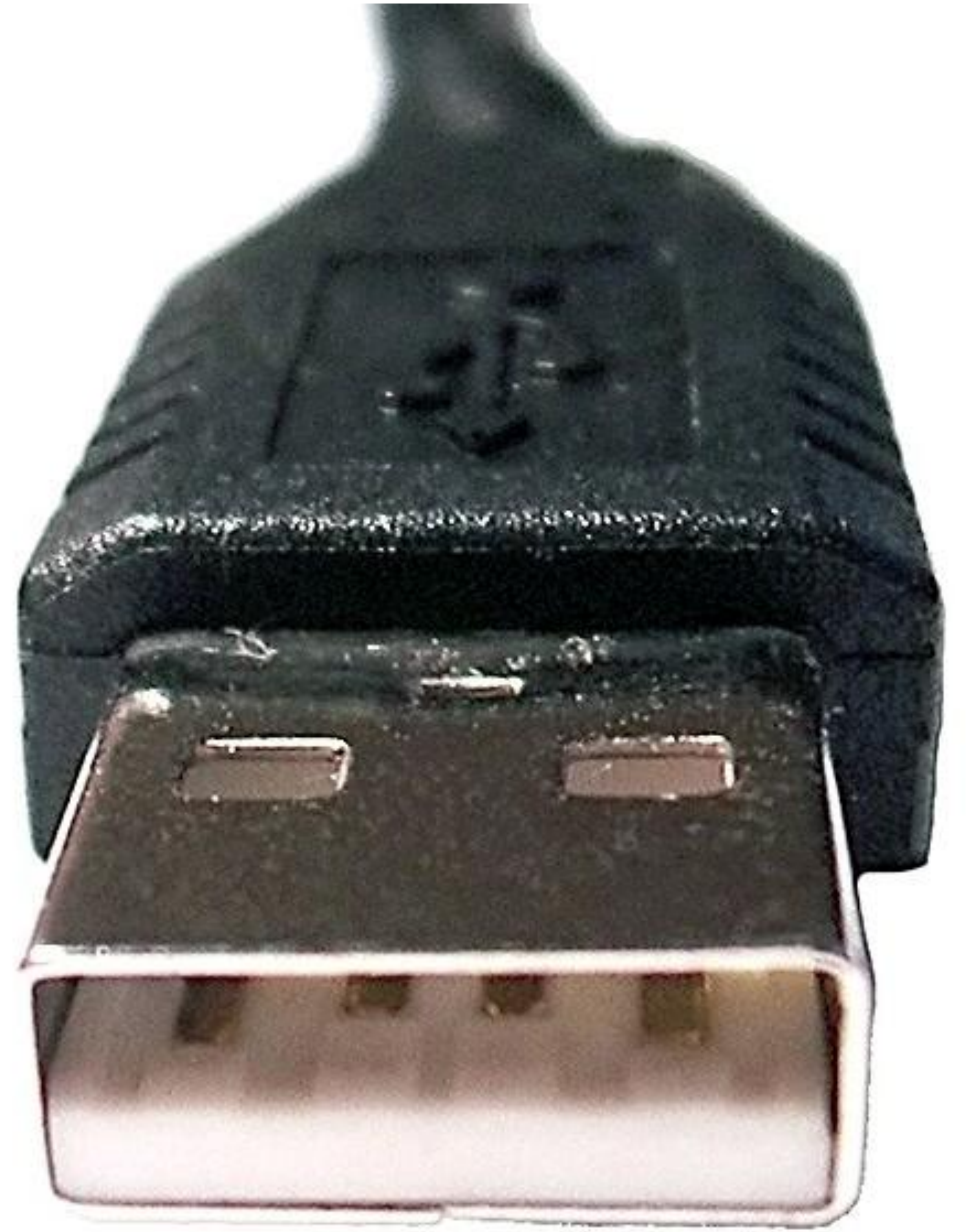
AI Agents

An AI agent is an **autonomous** system that can use **tools** and collaborate with other agents to **plan** and **act** on tasks. After it acts, the agent **reflects** on the results of its actions, learning iteratively and refining its approach to better align with its defined objectives.



The Orchestrator Agent is a multi-agent, multi-tool supervisor, router, and planner, facilitating complex task execution

Model Context Protocol (MCP)



Deterministic vs Non-Deterministic Computing

Two fundamental paradigms for how a system responds to inputs

DETERMINISTIC

Same input → same output, always

Behavior

Fully predictable

Examples

Sorting algo, calculators, SQL

Debuggability

Easy — repeatable bugs

Creativity

None — follows fixed rules

Best For

Safety-critical, auditable systems

VS

NON-DETERMINISTIC

Same input → potentially different outputs

Behavior

Probabilistic / stochastic

Examples

LLMs, neural nets, quantum

Debuggability

Hard — bugs may not reproduce

Creativity

High — emergent, novel outputs

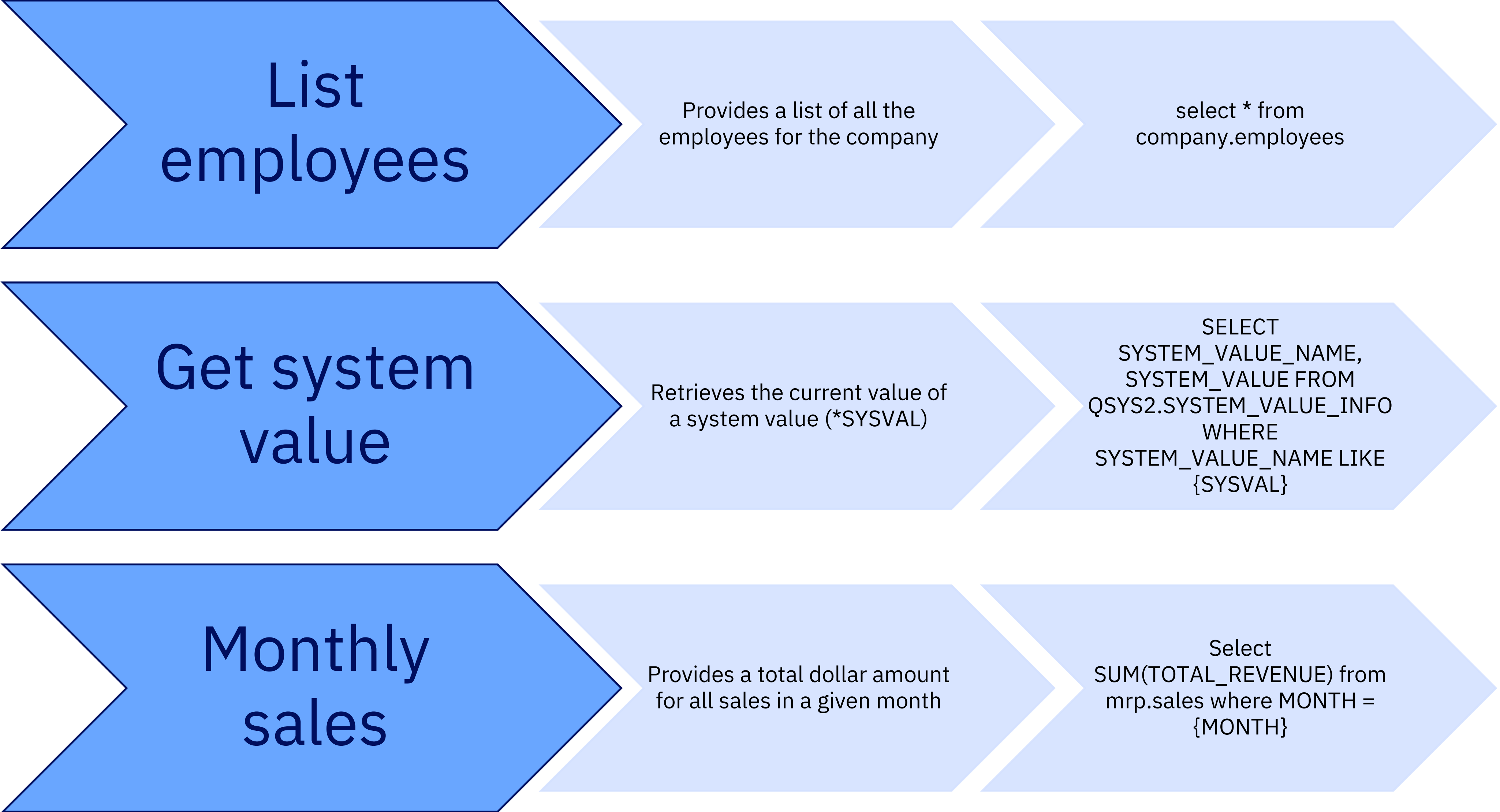
Best For

Open-ended, creative tasks

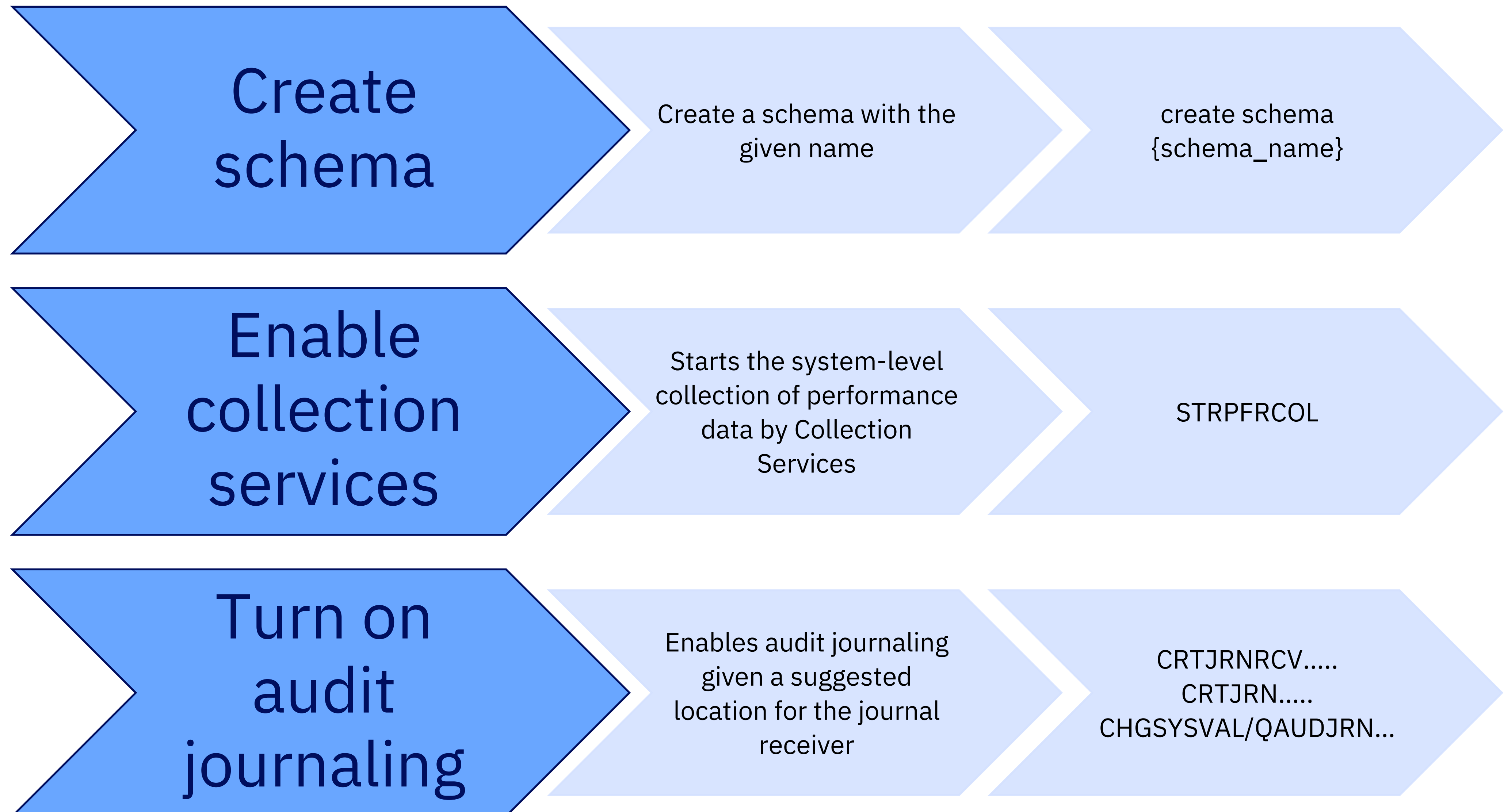
💡 Key Insight

Agentic AI blends both: deterministic infrastructure (pipelines, retrieval) + non-deterministic reasoning (LLM inference)

Mapping natural language to discrete data



Mapping natural language to discrete actions



Universal MCP Server for i

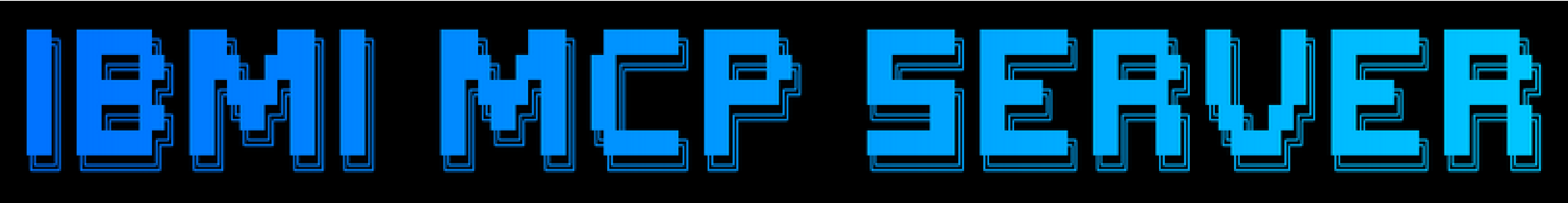
<https://ibm.biz/ibmi-mcp>

 **ibmi-mcp-server** Public

MCP server for IBM i

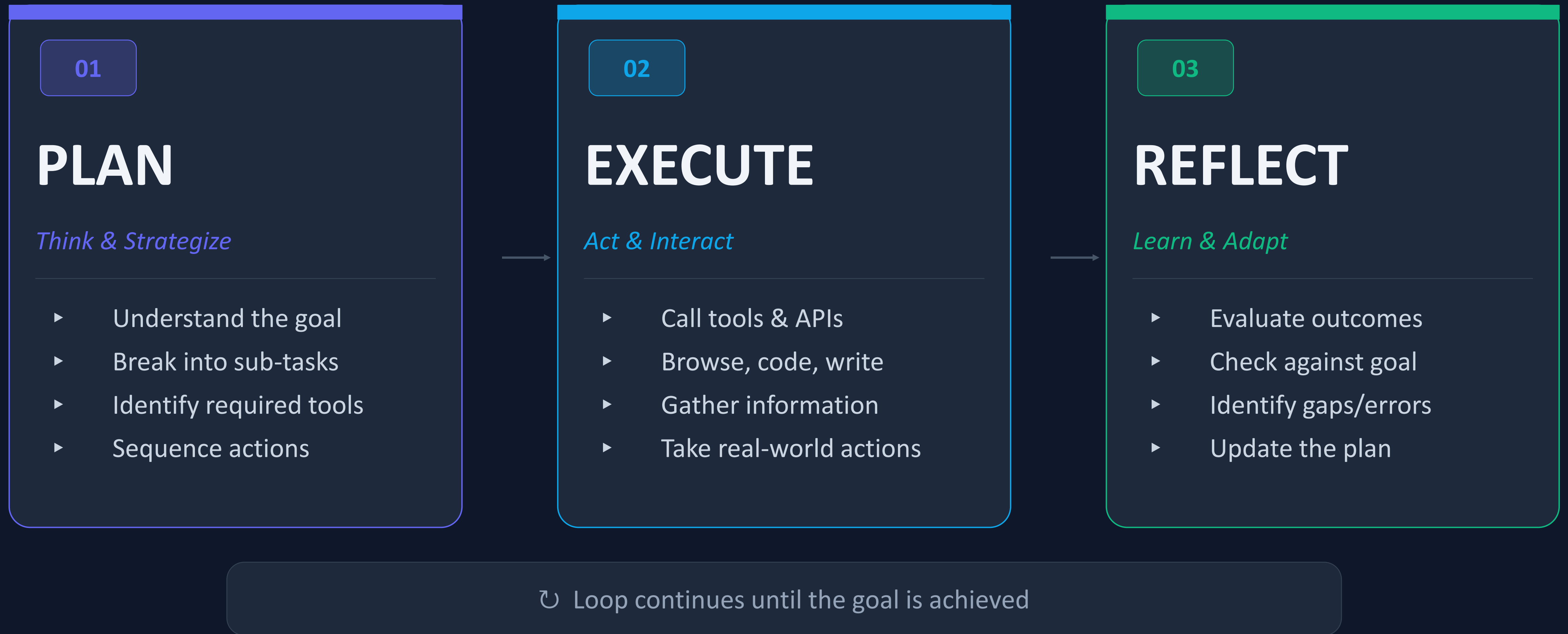
TypeScript	^5.8.3	MCP SDK	^1.17.1	MCP Spec	2025-06-18	Version	1.9.1	Coverage	64.67%	License	Apache 2.0
Status	Stable	 Ask DeepWiki									

 [Documentation](#) |  Docs are under active development



The Agentic AI Loop

How AI agents autonomously reason, act, and improve



Agent skills

Encapsulation of how to perform a task

Consists of:

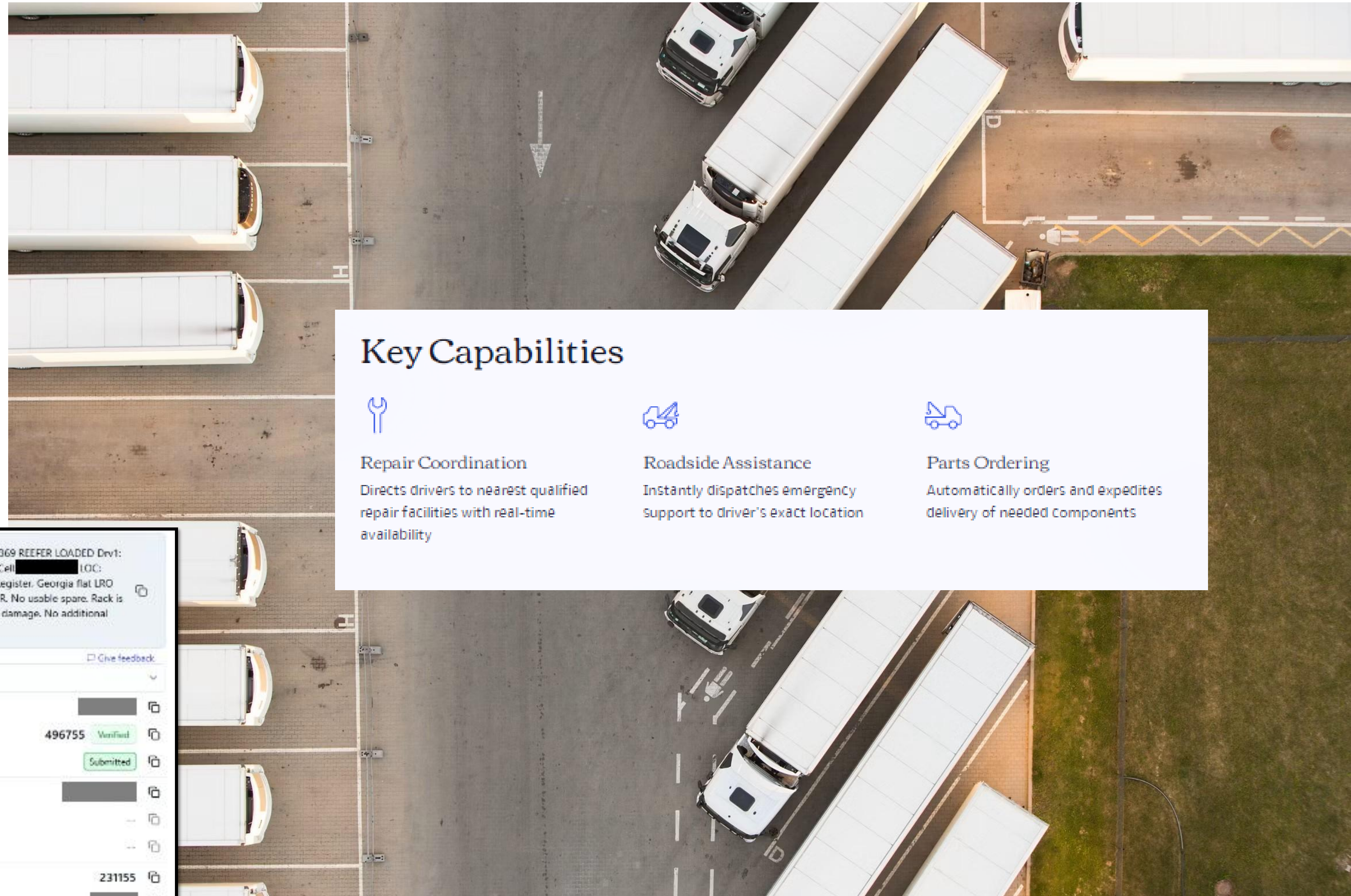
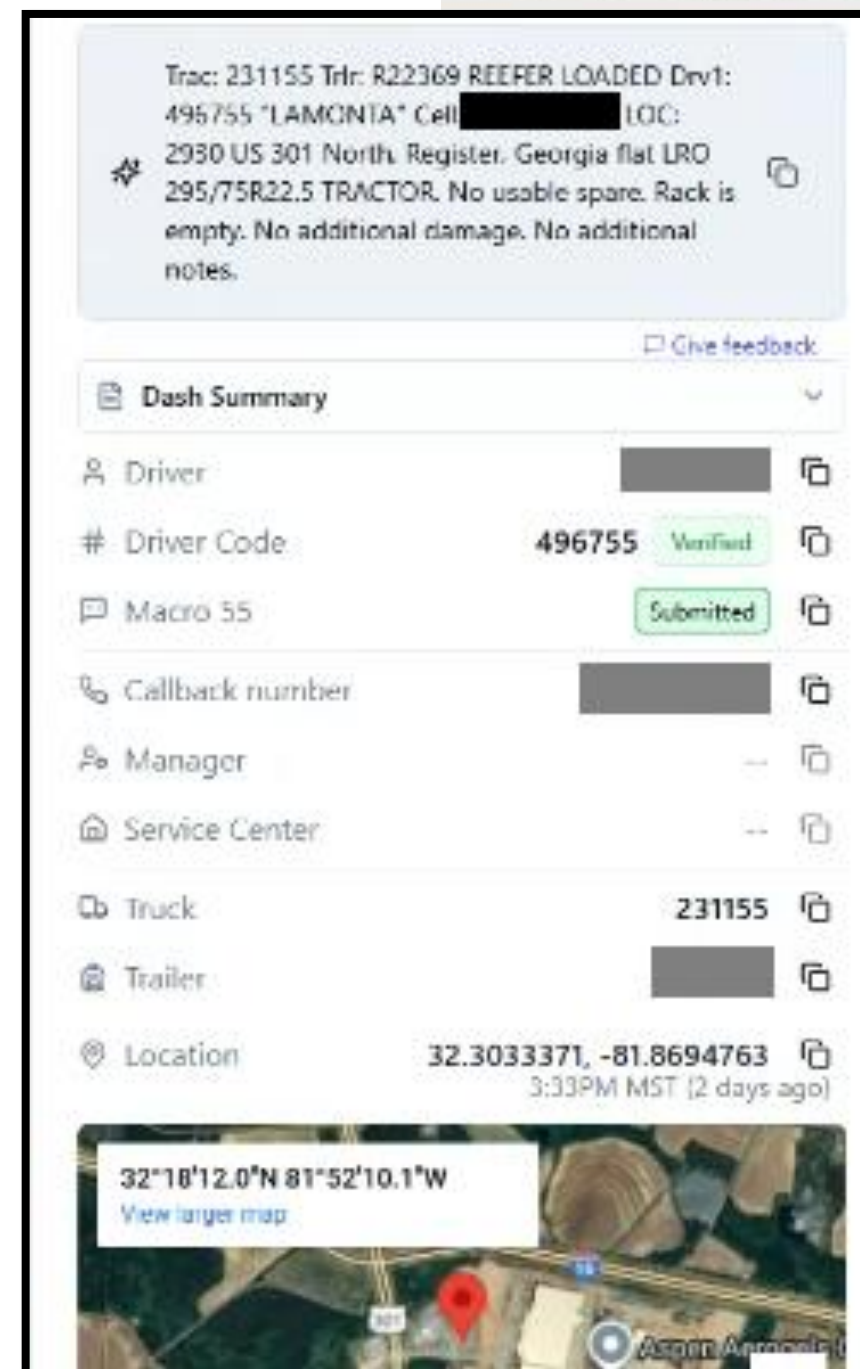
- Instructions
- Metadata
- Scripts or other resources if needed

```
1  ---
2  name: attack-vectors
3  description: >
4    Procedures for investigating IBM i privilege escalation attacks. Load when:
5    (1) investigating trigger attacks, rename attacks, or adopted authority
6    exploitation, (2) correlating authority exposure with exploitable attack paths,
7    (3) assessing compound attack chains that escalate individual findings.
8  ---
9
10 # Attack Vectors & Privilege Escalation
11
12 Procedures for investigating how authority weaknesses chain into active
13 privilege escalation. These investigations answer: "How can an attacker
14 use these exposures to gain elevated access?"
15
16 ## Quick Dispatch
17
18 | Vector | MCP Tool | Severity |
19 |-----|-----|-----|
20 | Trigger Attack | `list_db_files_exposed_to_trigger_attack` | CRITICAL |
21 | Rename Attack | `list_files_exposed_to_rename_attack` | CRITICAL |
22 | Adopted Authority | `list_adopted_authority_programs_with_public_access` | HIGH-CRITICAL |
23 | Command Authority | `list_public_authority_on_attack_vector_commands` | HIGH |
24
25 Load `attack-procedures.md` for full investigation steps per vector.
26
27 ## Critical Gotchas
28
29 - NAMING(*SQL) defaults USRPRF(*OWNER) for BOTH static AND dynamic SQL – many
30   programs adopt authority unknowingly. Always check DYNUSRPRF settings.
31 - Trigger programs execute under the adopted authority of the *accessing* program,
32   NOT the trigger creator – this is the core escalation mechanism.
33 - Rename attack creates a view that looks identical to the original table. Even
34   read-only access triggers malicious code in the WHERE clause.
35 - Heritage Navigator is known to be exposed to trigger attacks.
36
37 ## Compound Risk Patterns
38
39 Watch for combinations that escalate severity. See `compound-risk-patterns.md`
40 for the 6 named chains (Identity Takeover, Library List + Trigger, Trigger +
41 Adopted Authority, Rename + View Substitution, SQL Injection + Command Execution,
42 Service Account Exposure).
43
44 When you detect a compound pattern, call it out explicitly and escalate severity.
45
```

Real-life example:
Revolutionize driver support with Agentic AI

A large trucking company, whose core business runs on IBM i, creates an agentic voice-driven support assistant to...

- fully automate many support tasks, **such as coordinating repairs, providing roadside assistance, and ordering parts.**
- increase driver satisfaction with **24/7 availability and zero wait time.**
- improve fleet efficiency and bottom-line performance by **resolving issues 60% faster.**



Security problems

Security Concerns Around AI Adoption



What security and IT leaders should be watching for as AI tools spread across the organization



Shadow AI

Employees adopting AI tools outside IT-approved channels creates blind spots — unmanaged apps process company data with no visibility, logging, or oversight.



Data Governance with Third-Party Services

Sending sensitive or regulated data to external AI vendors raises questions about storage location, retention, model training use, and contractual protections.



Prompt Injection & Model Manipulation

Malicious inputs embedded in documents, emails, or web content can hijack an AI system's behavior or exfiltrate data through indirect instructions.



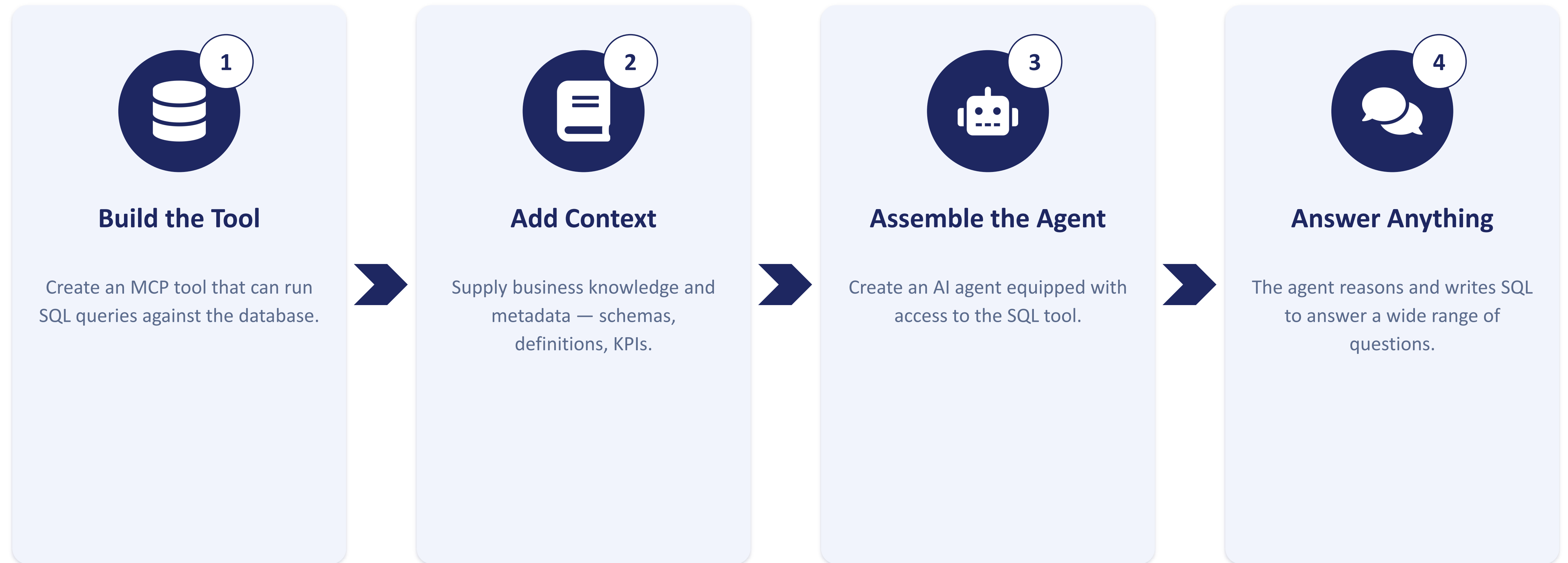
Compliance & IP Exposure

Outputs may inadvertently reproduce protected content or violate industry regulations, creating legal and reputational risk without clear accountability.

Treat AI security as a governance discipline: visibility into usage, clear data-handling policies, and vendor due diligence.

Example customer journey: From Raw Data to Answers

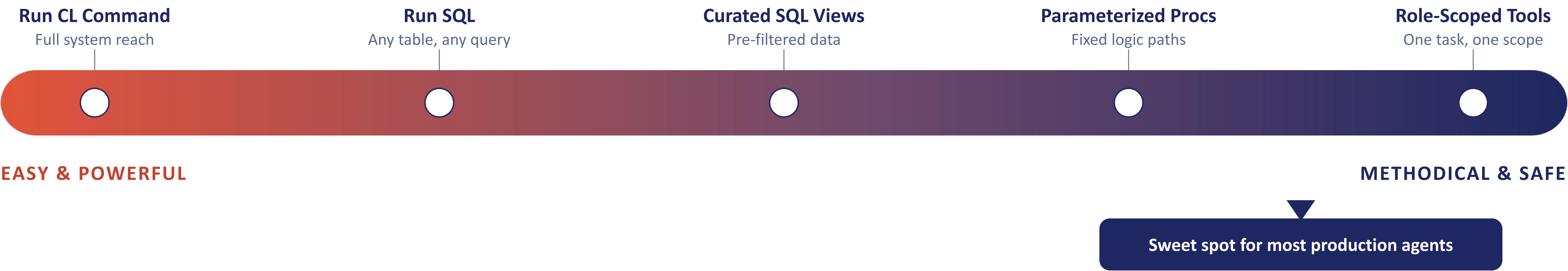
How an MCP tool and business context turn into a self-service SQL agent



Result: the agent grounds every question in real schema and business context, so it can answer a multitude of questions with correct, trustworthy SQL — no analyst required.

Easy & Powerful ↔ Methodical & Safe

Every tool choice sits somewhere on this spectrum — the right spot depends on the blast radius you're willing to accept.



Easy & Powerful

- Fastest to prototype and ship
- Broad, often unscoped access
- Large blast radius if misused



Methodical & Safe

- More upfront design effort
- Least-privilege, purpose-built access
- Small, auditable blast radius

Agents with Excessive Authority

The most common pitfall: provisioning an agent like a superuser instead of like the narrow task it actually performs.



Over-
Provisioned



Read-Only May Be Sufficient

Most agent tasks are lookups and reporting, not writes — default to SELECT-only access and add write authority only once it's proven necessary.



“Just in Case” Authority Creeps In

Broad or *ALLOBJ-style access is often granted for convenience during development and never scoped back down before production.



Bigger Authority, Bigger Blast Radius

A prompt injection, bug, or hallucinated action can only do as much damage as the profile behind it allows — excess authority removes that ceiling.

Lack of Proper Guardrails

Powerful access without checks, review, or visibility turns a small mistake into an unrecoverable one.



No Safety Net



No Human Approval on Sensitive Actions

Writes, deletes, and CL commands execute immediately, with no human-in-the-loop checkpoint before anything irreversible happens.



No Validation Before Execution

Generated SQL or commands run as-is — nothing checks that the agent's output actually matches the user's intent first.



No Monitoring or Kill Switch

Teams can't see what an agent is doing in real time and have no fast way to halt it mid-task if it goes off the rails.

Execution of Hallucinated SQL

Language models sometimes generate confident, plausible-looking SQL that is factually wrong — and without a check, that statement runs anyway.



**Confidently
Wrong**



Confident, Not Correct

A hallucinated query can reference tables, columns, or joins that don't exist, or misread the request entirely — while still reading as valid, well-formed SQL.



No Gap Between Generation and Execution

Without a review or dry-run step, the model's first guess becomes the query that actually touches the database.



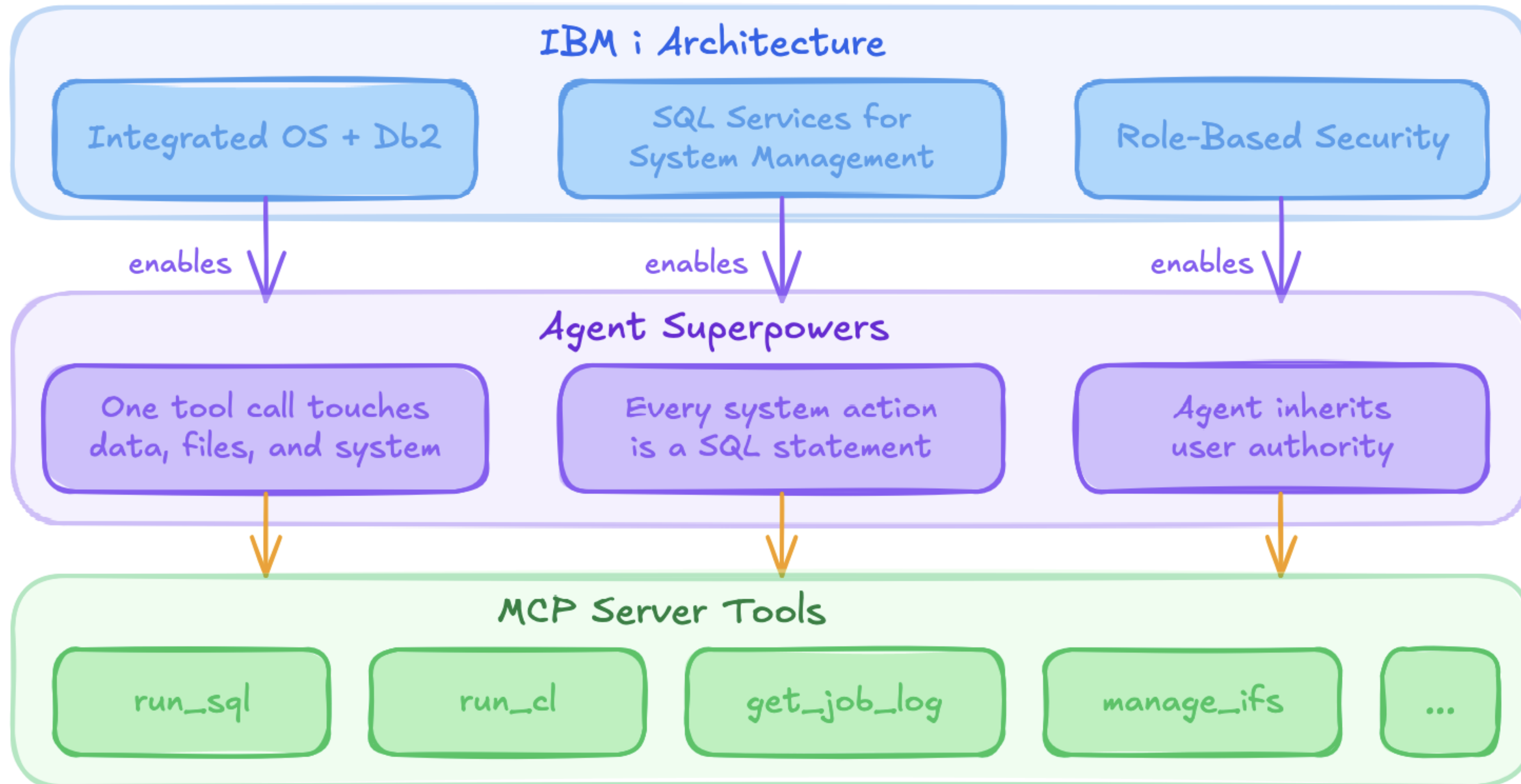
Damage Can Go Unnoticed

A hallucinated UPDATE or DELETE can silently alter or destroy data long before anyone realizes the SQL was wrong.

Deploying securely

IBM i: Built for Agents

Architectural decisions decades ago = AI-ready today



5 Requirements of an Enterprise AI OS

1. Governed Data Access

Access business data without crossing system boundaries

2. Object-Level Enforcement

Enforce what an agent can do at the object level

3. Intrinsic Audit Trails

Every action traceable to a decision, identity, and policy

4. Hardware Abstraction

Absorb new hardware without rewriting core applications

5. Bounded Autonomy

Agents operate within autonomy the system enforces

IBM i answers YES to all five -- by design.

Be Explicit About Design

Before writing a single tool or prompt, answer these three questions clearly.

QUESTION 1



What problem are you trying to solve?

Define the specific business question or workflow the agent should address — not just “AI for X.”

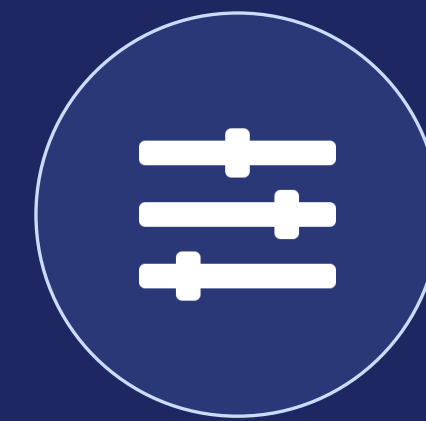
QUESTION 2



What IBM i data do you need access to?

Identify the exact files, tables, or views the agent should be able to query — and nothing more.

QUESTION 3



How should your agent behave in your environment (scope of actions)?

Decide what's in bounds — read-only, specific writes, or command execution — before you build.

Leverage What IBM i Already Gives You

Native platform controls provide agent guardrails before you write a single line of new code.



**Built-in
Guardrails**



Object-Level Authorities

The agent's *USRPRF is still bound by the IBM i security model — every object access is checked, no matter what the agent requests.



SQL Query Governor

Cap runaway or expensive queries an agent generates so a bad prompt can't degrade production performance.



Workload Capping Groups

Limit CPU and resource consumption per subsystem so one agent can't monopolize the partition.

Maintain Deny-by-Default Methodologies

Start from zero access and grant only what the agent's task truly requires.



**Zero Trust
by Default**

Minimal-Access *USRPRF



Run every agent under a dedicated profile scoped to only the objects and commands its task requires — nothing implicit, nothing inherited.

Audit with Authority Collection



Use Authority Collection and related tooling to capture real usage, right-size permissions, and catch privilege drift over time.

Apply Role-Based Access Control (use an orchestration tool)

Match agent exposure to business need instead of granting universal reach.



Scope
by Role

Scope Agents to Roles



Assign each agent to the roles that actually need it, mirroring how you would provision a human user.

Don't Expose Every Agent Everywhere



Limit which apps and users can invoke a given agent — availability should match business need, not default to global.

Avoid Broad-Spectrum Tools

Wide-open execution tools inherit the full reach of the agent's profile — narrow the blast radius by design.



**Narrow the
Blast Radius**

“Run SQL” Tools



An open execute-any-SQL tool can read or write anything the underlying profile can touch — prefer purpose-built, parameterized queries.

“Run CL Command” Tools



An unrestricted CL execution tool is effectively full system access — replace it with a small set of task-specific tools instead.

Avoiding broad-spectrum tools

BE DILIGENT!!!

Carefully map out exactly which data to expose to an AI agent.

Provide MCP tools to provide that data, and only that data

Use deterministic logic gates to protect against malicious agentic queries

Seriously.... Well-defined tools is **crucial**



Specialized tool example

```
find_lumber_parts_by_composition:
  source: ibmi-lumber
  description: Find lumber parts by wood composition type
  statement: |
    SELECT
      PART_ID,
      DESCRIPTION,
      LENGTH_INCHES,
      COMPOSITION
    FROM LUMBER.LUMBER_PARTS
    WHERE COMPOSITION = :composition
    ORDER BY DESCRIPTION, LENGTH_INCHES
  parameters:
    - name: composition
      type: string
      description: "Wood composition type"
      required: true
      enum: ["Pine", "Cedar", "Oak"]
```

Other Security Risks Associated with AI Deployment

AI agents introduce risk beyond authority and guardrails alone — keep an eye on these as well.



Prompt Injection

Malicious instructions hidden in documents, web content, or tool output can hijack the agent and override its original task.



Sensitive Data Leakage

Responses, logs, or context windows can expose confidential data across users, sessions, or downstream systems.



Insecure Tool & Plugin Integrations

Third-party tools and MCP servers bring their own vulnerabilities, widening the attack surface beyond the agent itself.



Model & Supply Chain Risk

Untrusted models, weights, or dependencies can introduce hidden backdoors or tampered behavior.



Inadequate Logging & Traceability

Without full audit trails, teams can't reconstruct what an agent did or who is accountable after an incident.



Resource Abuse & Denial of Service

Malicious or runaway prompts can consume excessive compute, driving up cost or degrading service for others.

One continuous trajectory

Each revolution didn't erase the last — it built on it, accelerating how we make, move, and connect.



01 · 1760 – 1840

Mechanization

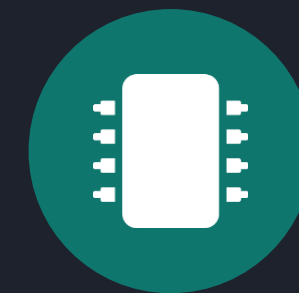
Steam and machines power the first factories.



02 · 1870 – 1914

Mass Production

Electricity and assembly lines scale output.



03 · 1969 – 2000s

Automation

Computers and the internet digitize industry.



04 · 2010s – present

Smart Systems

AI and Quantum computing drive the next era of human advancement.

Klaus Schwab popularized the term “Fourth Industrial Revolution” (World Economic Forum, 2016).

Mot de la fin

Retenez ceci :

- L'IA ne peut pas être ignorée. Elle peut (et va) apporter de la valeur à votre entreprise !
- Les pratiques de sécurité classiques d'IBM i sont plus importantes que jamais
- Ne vous laissez pas séduire par les victoires faciles. Assurez-vous que votre IA est déployée en toute sécurité.
- Pour des informations générales sur l'IA → <http://ibm.biz/ai-ibmi>

Notes personnelles :

- Jesse adore la France
- Jesse adore les Français
- J'espère vous rencontrer en personne prochainement, mais un grand merci aux organisateurs de l'événement. Même en virtuel, c'est un plaisir d'être présent

Thank you

For more
information

IBM Sites:

IBM i Home Page	https://www.ibm.com/it-infrastructure/power/os/ibm-i
IBM Strategy Whitepaper	https://www.ibm.com/it-infrastructure/us-en/resources/power/i-strategy-roadmap/
IBM Client Success	https://www.ibm.com/it-infrastructure/us-en/resources/power/ibm-i-customer-stories/
Support Life Cycle	https://www.ibm.com/support/lifecycle/
License Topics	https://www-01.ibm.com/support/docview.wss?uid=nas8N1022087
IBM i Release Life Cycle	https://www.ibm.com/support/pages/release-life-cycle
IBM i TR Wikis	IBM i Technology Updates

For more
information:

Blogs to follow

- [TechChannel You and i \(Steve Will\)](#)
- [TechChannel: i Can \(Dawn May\)](#)
- [TechChannel: OpenYour i with Jesse Gorzinski](#)
- [IBM Db2 for i \(Kent Milligan\)](#)
- [IBM i Analytics Blog \(Jon Westcott Jr\)](#)
- [iSee with Scott and Tim](#)

More to follow

[@IBMservers](#)
[@COMMONug](#)
[@IBMChampions](#)
[@IBMSystemsISVs](#)
[@LinuxIBMMag](#)
[@OpenPOWERorg](#)
[@ITJungleNews](#)
[@SAPonIBMi](#)
[@SiDforIBMi](#)
[@IBMAIXeSupp](#)
[@IBMAIXdoc](#)

Hashtags to Use

#IBMPower
#IBMi
#IBMAIX
#Power10
#LinuxonPower
#HANAonPower
#ITinfrastructure
#OpenSource
#HybridCloud

Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.

Special notices (cont.)

IBM, the IBM logo, ibm.com AIX, AIX (logo), AIX 5L, AIX 6 (logo), AS/400, BladeCenter, Blue Gene, ClusterProven, Db2, ESCON, i5/OS, i5/OS (logo), IBM Business Partner (logo), IntelliStation, LoadLeveler, Lotus, Lotus Notes, Notes, Operating System/400, OS/400, PartnerLink, PartnerWorld, PowerPC, pSeries, Rational, RISC System/6000, RS/6000, THINK, Tivoli, Tivoli (logo), Tivoli Management Environment, WebSphere, xSeries, z/OS, zSeries, Active Memory, Balanced Warehouse, CacheFlow, Cool Blue, IBM Systems Director VMControl, pureScale, TurboCore, Chiphopper, Cloudscape, Db2 Universal Database, DS4000, DS6000, DS8000, EnergyScale, Enterprise Workload Manager, General Parallel File System, , GPFS, HACMP, HACMP/6000, HASM, IBM Systems Director Active Energy Manager, iSeries, Micro-Partitioning, POWER, PowerExecutive, PowerVM, PowerVM (logo), PowerHA, Power Architecture, Power Everywhere, Power Family, POWER Hypervisor, Power Systems, Power Systems (logo), Power Systems Software, Power Systems Software (logo), POWER2, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, POWER7, System i, System p, System p5, System Storage, System z, TME 10, Workload Partitions Manager and X-Architecture are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries.

A full list of U.S. trademarks owned by IBM may be found at: <http://www.ibm.com/legal/copytrade.shtml>. Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Altivec is a trademark of Freescale Semiconductor, Inc.

AMD Opteron is a trademark of Advanced Micro Devices, Inc.

InfiniBand, InfiniBand Trade Association and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

Microsoft, Windows and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries or both.

NetBench is a registered trademark of Ziff Davis Media in the United States, other countries or both.

SPECint, SPECfp, SPECjbb, SPECweb, SPECjAppServer, SPEC OMP, SPECviewperf, SPECapc, SPEChpc, SPECjvm, SPECmail, SPECimap and SPECsfs are trademarks of the Standard Performance Evaluation Corp (SPEC).

The Power Architecture and Power.org wordmarks and the Power and Power.org logos and related marks are trademarks and service marks licensed by Power.org.

TPC-C and TPC-H are trademarks of the Transaction Performance Processing Council (TPPC).

UNIX is a registered trademark of The Open Group in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.

IBMi