SNIA | Cloud Storage Technologies

Agentic AI
Use Cases, Benefits, Risks

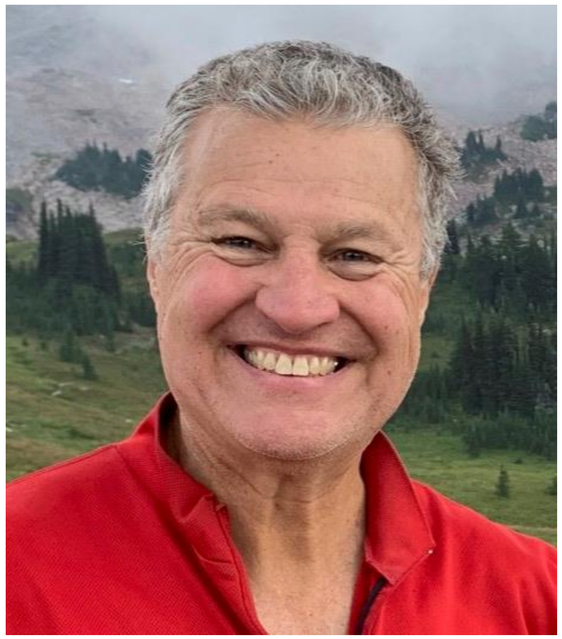
Live Webinar
June 26, 2025
10:00 am PT / 1:00 pm ET

Erin Farr
IBM

Michael Hoard
SNIA CST

Today's Presenters

Erin Farr
Presenter
Senior Technical Staff Member
Storage CTO Office
IBM



Michael Hoard
Moderator
SNIA
Cloud Storage Technologies Chair

SNIA | CST A SNIA Community

SNIA | Cloud Storage Technologies

snia.org/groups/cst

Committed to the adoption, growth and standardization of **intelligent data storage usage** in cloud infrastructures

- New simplified SNIA Membership Model for FY'25
- Pay one fee for SNIA Membership – Participate in ANY group!
- Join SNIA groups at https://www.snia.org/member_com

The SNIA Community

200
industry leading organizations

2,000
active contributing members

50,000
IT end users & storage pros worldwide

SNIA | CST A SNIA Community

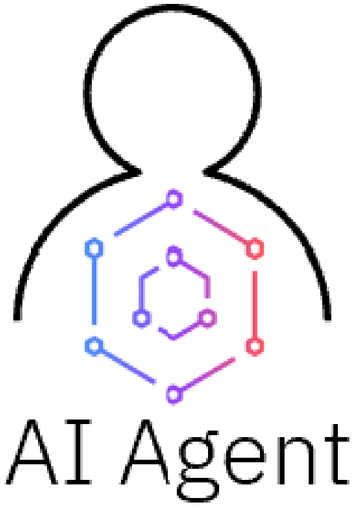
SNIA Legal Notice

- The material contained in this presentation is copyrighted by SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Agenda

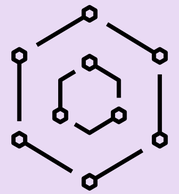
- Market Trends
- Latest Evolution of AI
- What is Agentic AI?
- Use Case Considerations
- Sample use case and demo
- Model Context Protocol
- Considerations when assessing agentic AI solutions for Enterprises



AI is changing at a rapid pace

AI that can create for you

AI that can do for you



Agents era

Market Trends

In 2025, **25%** of companies that use Gen AI will have launched **agentic AI pilots**, growing to **50%** by 2027 [1]

Most tech pros anticipate **agents will become core to operations over the next 12 months**, powering more than 25% of processes by year-end 3

By 2028, at least **15% of day-to-day work decisions** will be made **autonomously through agentic AI**, up from zero percent in 2024 [2]

Investors have poured over **$2 billion into agentic AI startups** in the past **two years**, focusing their investment on companies that target the **enterprise market** [4]

Image by Brigitte Werner from Pixabay

[1] Deloitte, Autonomous generative AI agents: Under development
]2] Gartner, **Top Strategic Technology Trends for 2025: Agentic AI**, G00818765, Oct 2024
[3] State of AI Agent Development Strategies in the Enterprise" survey of over 1,000 enterprise technology leaders and practitioners, Tray.ai, Dec 2024
[4] CB Insights. Gen AI Investment Database, Aug 21, 2024

Technical evolution: from **standalone models** to **agentic systems**

~2022

Large Language Models (LLMs)

- Models that predict the next word
- Pitfalls:
  - Limited by what they were trained on
  - Hallucinations (weather)
  - Bad at certain things (e.g. math)

Instead of trying to put all the knowledge inside the model, design systems on top of the model

~2023

Compound AI Systems (fixed flows, e.g. RAG)

- Way to infuse new knowledge without retraining the model
- Uses multiple interacting components - calls to models, retrievers, or external tools (e.g. guard rails)
- Reduces risk of hallucinations
- Pitfalls:
  - Needs set up at build time ("fixed flow") which can limit use cases

What if, instead of responding to a question, AI can accomplish a task or goal?

~2024+

Agentic Systems

- A program whose **execution logic is controlled by an LLM**
- Instead of "generating content" to return to the user, it performs actions ("has **agency**") on behalf of the user
- Can define how to solve, then reflect, and update the plan (vs. fixed systems that are set up at build time)

Generates content

Takes action

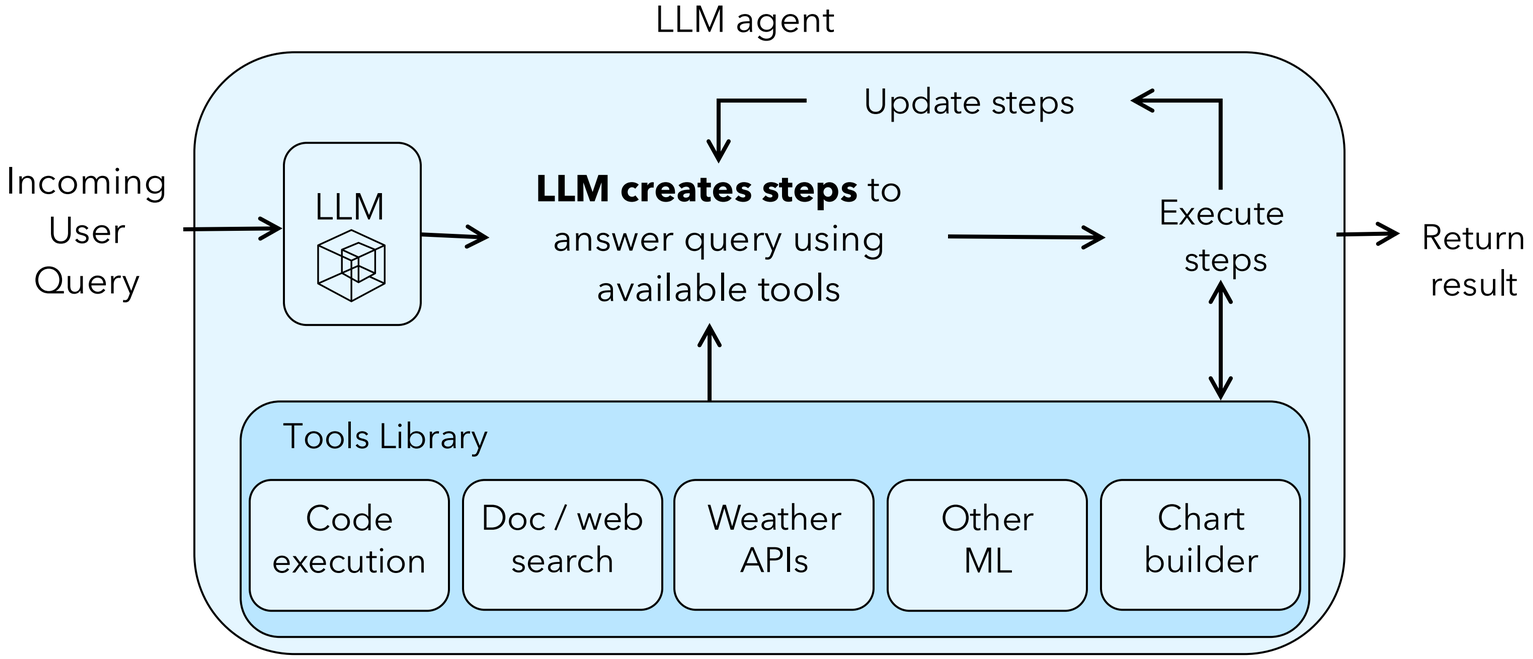
More reliable
Less flexible

Less reliable
More flexible

Expect to see combinations of fixed flows and autonomous agents

AI Agent

An LLM agent is a system that can (more or less autonomously) interact with a designated environment to achieve a defined goal.



Key Features

- **Performs tasks** (Goal oriented) defined by incoming user query, using "skills" - tools you give your model access to, such as:
  - APIs, python function you write, RAG, another model (e.g. translation function)
- **Feedback loop**

Some say: "With agentic AI, today's front-end systems will become back-end systems."

- Replacing UI completely depends on use case (e.g. Alexa doesn't need a UI, Storage management tools probably would)

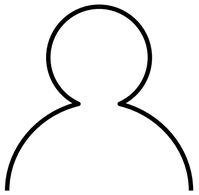
Advantages:

- Adaptability to support dynamic environments
- Can improve upon past actions
- Interactivity with other systems
- Autonomy

Disadvantages:

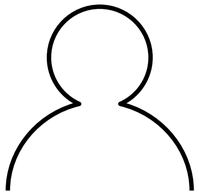
- Less reliable
  - Hallucinations
  - Brittleness
  - Stuck in feedback loops
- Resource intensive
- Security risks amplified
- Debugging complexity

AI – Value Evolution



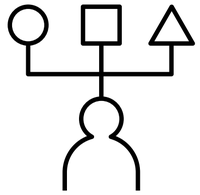
AI Assistant

Information retrieval



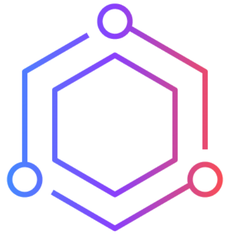
AI Agent

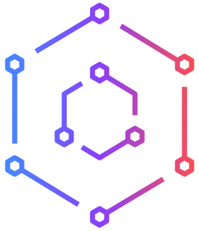
Perform tasks with oversight

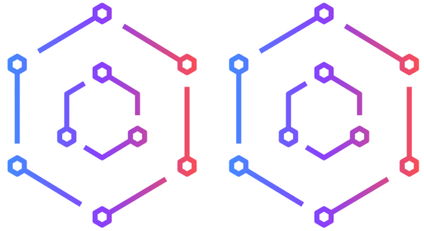


Multi-agent Platform

Autonomous action-taking







Reactive	Proactive
Generates content	Makes operational decisions
Prescriptive tasks	Goal-oriented
Single-step processes	Multi-step processes
Human interaction required	Less or no human interaction required
Fixed flow at build time	Reflective and Self-correcting

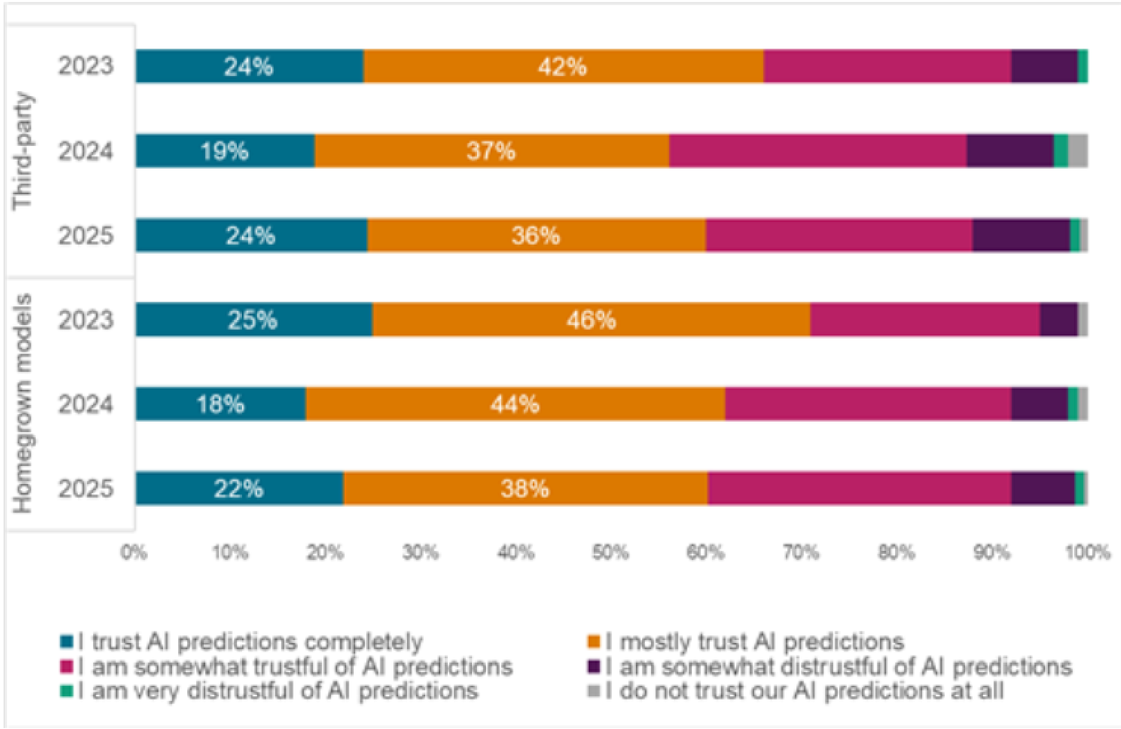
Acceptance Challenges

Confidence in the accuracy of AI models remains low

- Only about 1/5 of AI decision-makers fully trust their models
- For many industries, particularly those with a low tolerance for inaccuracy, this level of trust is insufficient

- 451 Research, How might rapid adoption of agentic AI technologies impact the future of SaaS?

Figure 3: Confidence in model accuracy is low



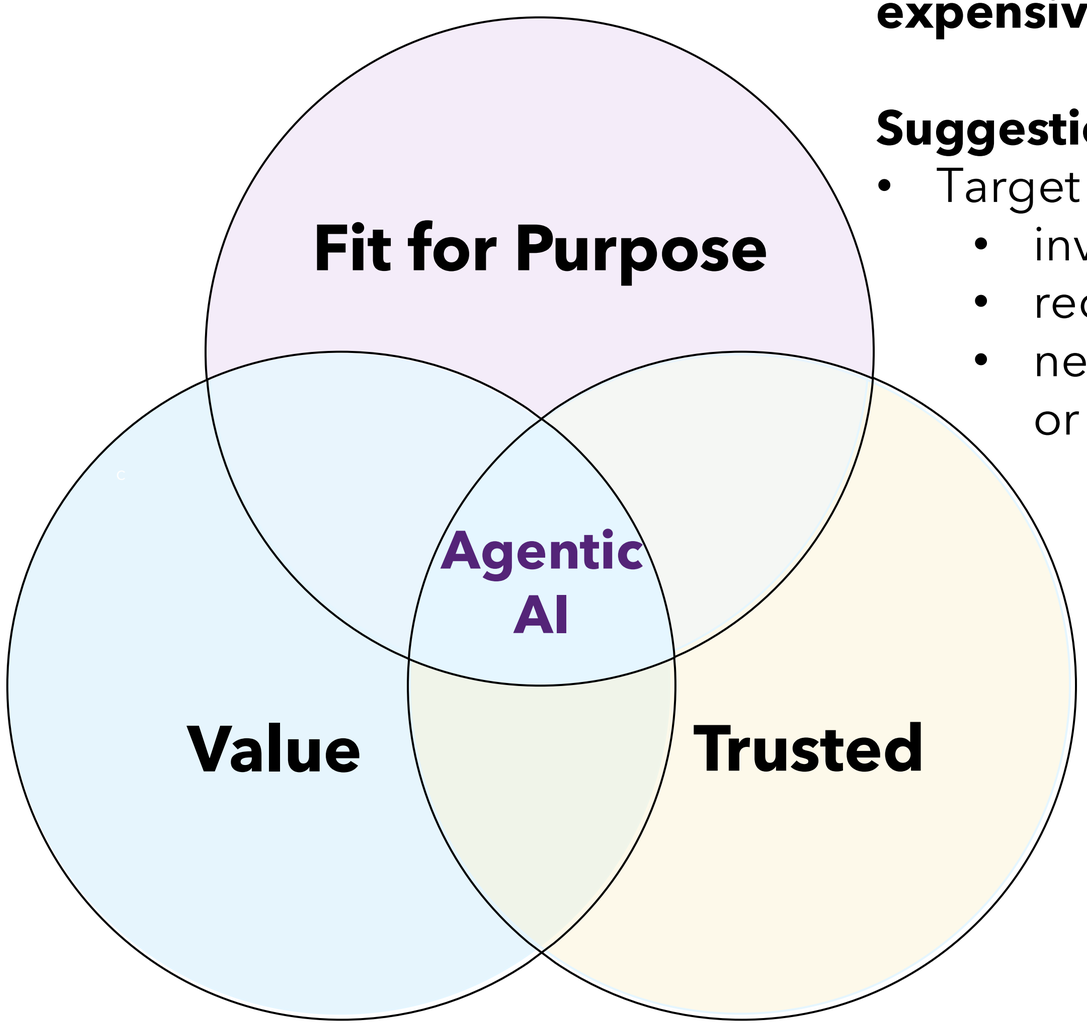
Source: 451 Research's Voice of the Enterprise: AI & Machine Learning, Use Cases 2025 (n=538), 2024 (n=579), 2023 (n=996).
Q. How trusting are you of predictions made by your organization's AI applications/algorithms?
Base: All respondents, abbreviated fielding
Q. And how trusting are you of predictions made by AI applications/algorithms from organizations other than your own?
Base: All respondents, abbreviated fielding.

SNIA | CST A SNIA Community

Use Case Considerations
for Agentic AI

The **complexity** of agentic architectures may create disillusionment as enterprises try to move to production

Suggestion:
- Focus on enterprise qualities of service with fit-for-purpose use cases
- Target work an "intern" may do, or currently intractable use cases that provide *new* value



Agentic is **less reliable** and **more expensive** than imperative programming

Suggestion:
- Target use cases that may:
  - involve uncertainty
  - require decision-making
  - need to adapt to new information or changing environments

Agentic AI amplifies **security and operational risk**

Suggestion:
- Start with lower-risk actions
- Ensure Security, Privacy, Governance and Transparency are built-in

Use Case - Cyber Resiliency

"How do I *really* know when I recover my data that it's a good copy and not corrupted?"

- Large Financial Services company

PROVABILITY

- **Lack of confidence** that running a single tool truly ensures data integrity
- **Manual verification** is often required
- Organizations are **unaware of how recoverable** their mission critical data is

DYNAMIC + COMPLEX ENVIRONMENTS

- **Microservices** (especially those with persistent storage) **complicate the boundaries and scope** of an application
- Updating of data integrity checks as **applications evolve**
- Ensuring **compatibility** of recovery environment with application/production

SCALE OF RECOVERY

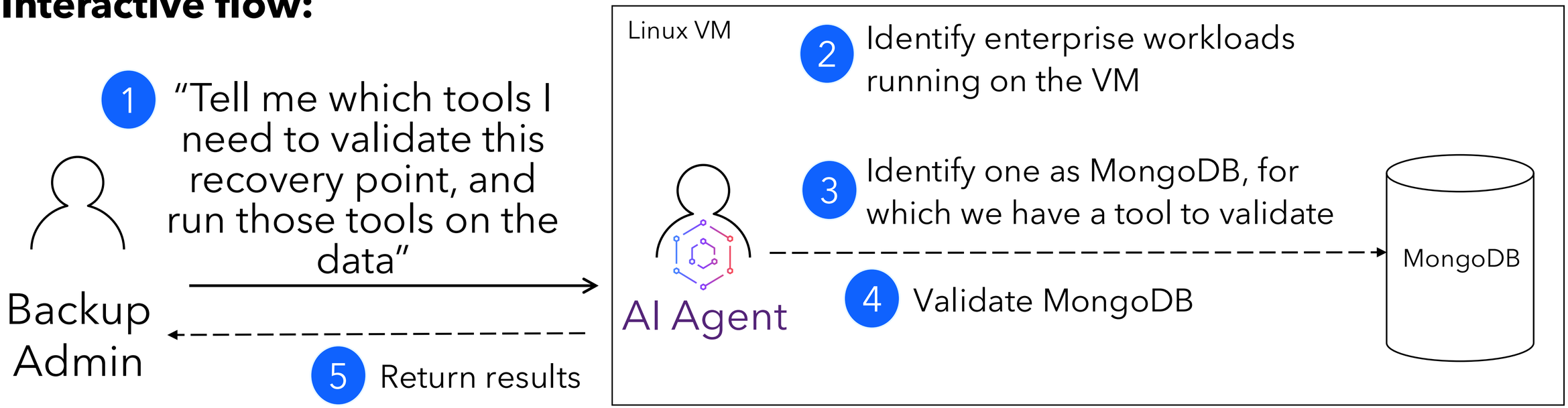
- Disconnect between **infrastructure** and **application** teams that need to collaborate on data recovery planning
- Constraints on **staffing and expertise**
- **Regulations** (e.g. DORA) require reports on provability of recovery
- Cyber Resiliency **recovery testing is not performed** (time, cost)

All of this is expensive at scale, making it hard to prove to stakeholders that your data and applications are recoverable.

SNIA | CST A SNIA Community

Agentic Use Case – Cyber Recovery Data Validation

Interactive flow:

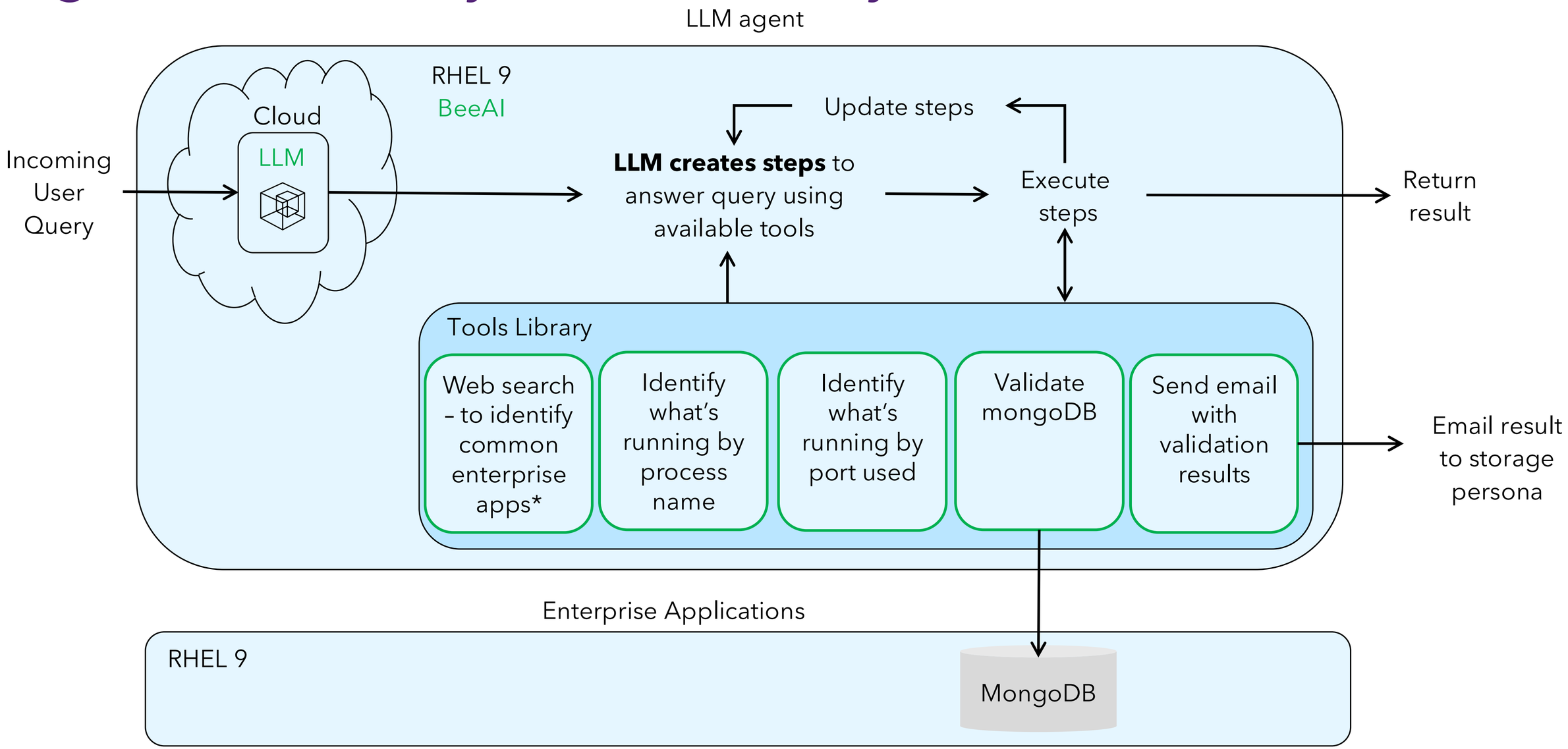


Autonomous flow:

Schedule runs of the AI agent with preloaded prompts and email the Admin with summarized results

SNIA | CST A SNIA Community

Agentic AI for Cyber Resiliency - Architecture



* Agent sometimes just uses data from LLM rather than doing web search

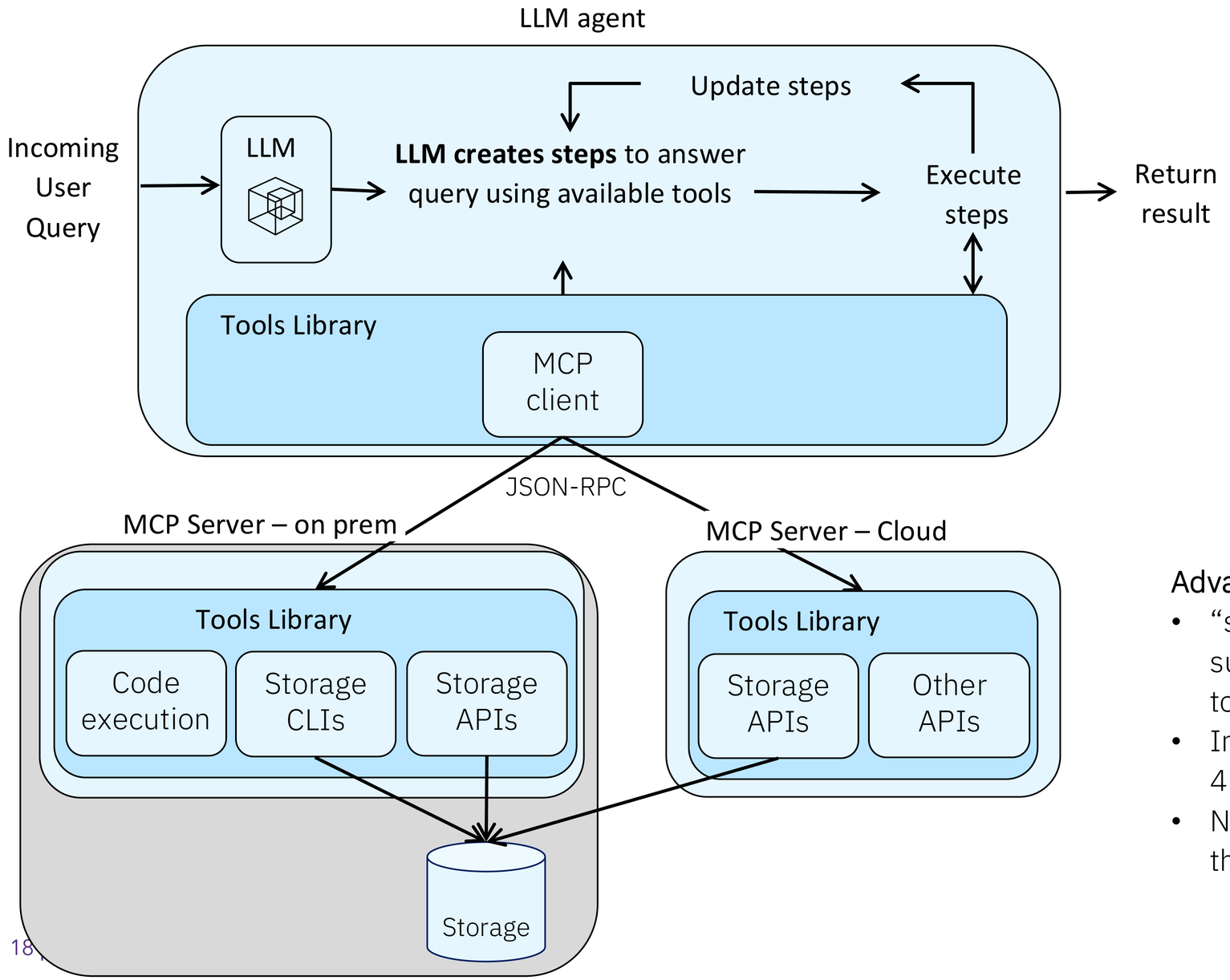
Demo

Link to Code:
https://github.com/IBM/agentic-ai-cyberres

Code | Blame | 152 lines (132 loc) · 4.84 KB | Raw

```
1   //
2   // Copyright contributors to the agentic-ai-cyberres project
3   //
4   import { DynamicTool, StringToolOutput } from "beeai-framework/tools/base";
5   import { z } from "zod";
6   import { ObjectId } from "mongodb";
7   import * as mongoDB from "mongodb";
8   import { execSync } from 'child_process';
9   import { getEnv, parseEnv } from "bee-agent-framework/internals/env";
10
11  /*
12   * Tool to look at running processs to determine what applications may be running
13   */
14  export const FindRunningProcessesTool = new DynamicTool({
15    name: "FindRunningProcesses",
16    description: "Determine what applications are running on the system by looking at running pr
17    inputSchema: z.object({
18      min: z.number().int().min(0),
19    }),
20    async handler(input) {
21
22          var returnString = new String;
23          var stdout = new String;
24
25          // do shell escape to run ps to see what processes are running.  Exclude kernel proces
26          try {
27                  stdout = execSync('ps  --ppid 2 -p 2 --deselect').toString();
28                  console.log(`stdout: ${stdout}`);
29                  returnString = stdout;
30          } catch (error: any) {
31                  console.error(`Error: ${error.message}`);
32                  if (error.stderr) {
33                          console.error(`stderr: ${error.stderr.toString()}`);
34                  }
35                  returnString = "Validation Failed.  Details:\n" + error.stderr;
36          }
37
38          return new StringToolOutput(returnString);
39
40    },
41  });
```


AI Agent with Model Context Protocol (MCP)



- **Open protocol** that standardizes how applications provide context to LLMs
- "USB-C" of AI apps
- **Decouples the tools from the agent so any agent can use any tool that's provided by MCP servers**

Advantages of MCP:

- "standardized" way to surface existing APIs to AI agents
- Industry traction (over 4K servers)
- NOTE: connects more than just tools

What to watch:

- Open questions and potential blockers to production still exist
- Tailored to work with Claude desktop
- Fitting right "tools" into limited model context

Why not put all APIs into a MCP server?

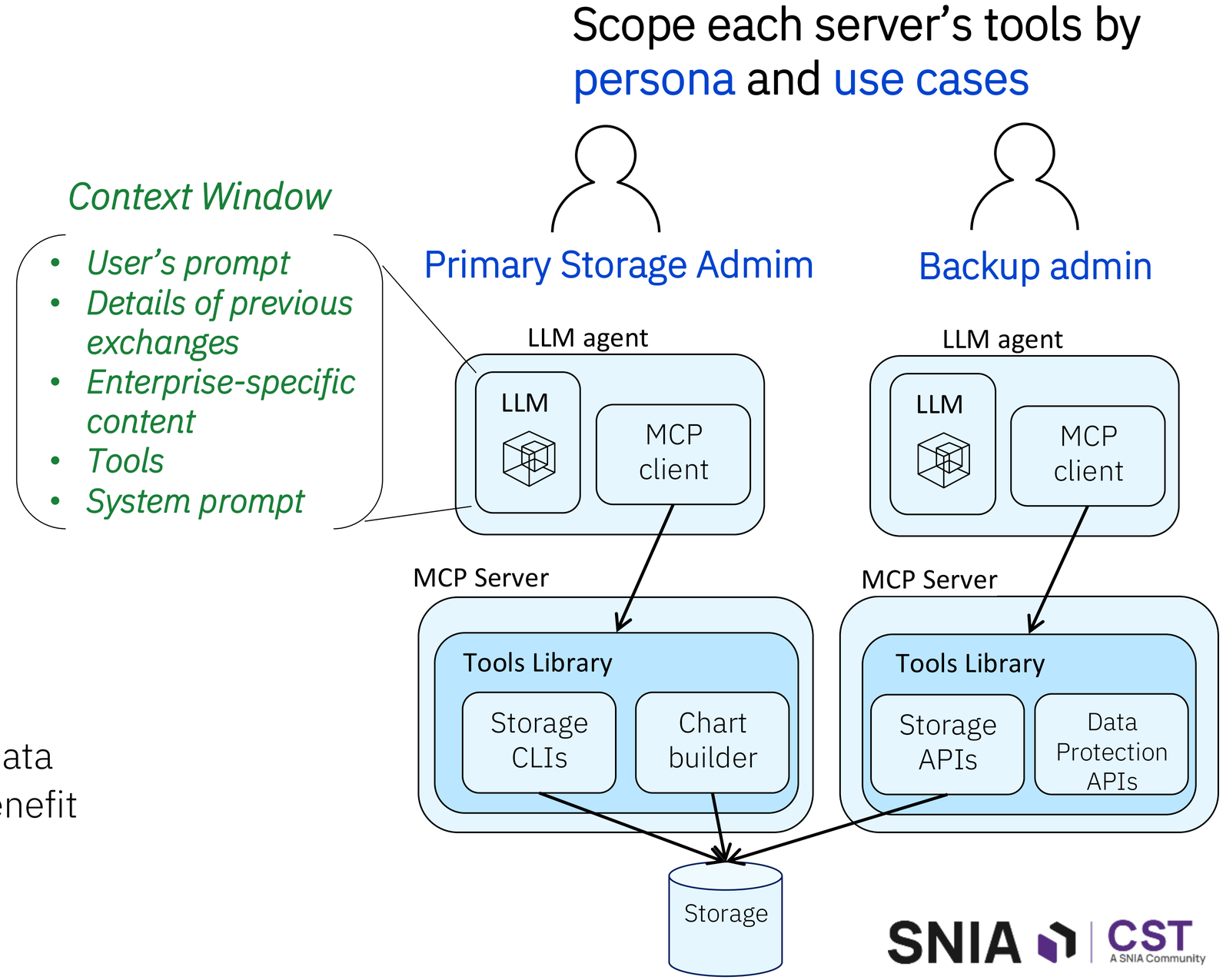
Current industry experience:

Don't scope MCP server too broadly

- Config of > 30-50 MCP tools can exceed the context window of LLMs
- Requires the user to have knowledge of which tools are necessary to perform the work – ***defeating the purpose of using AI***

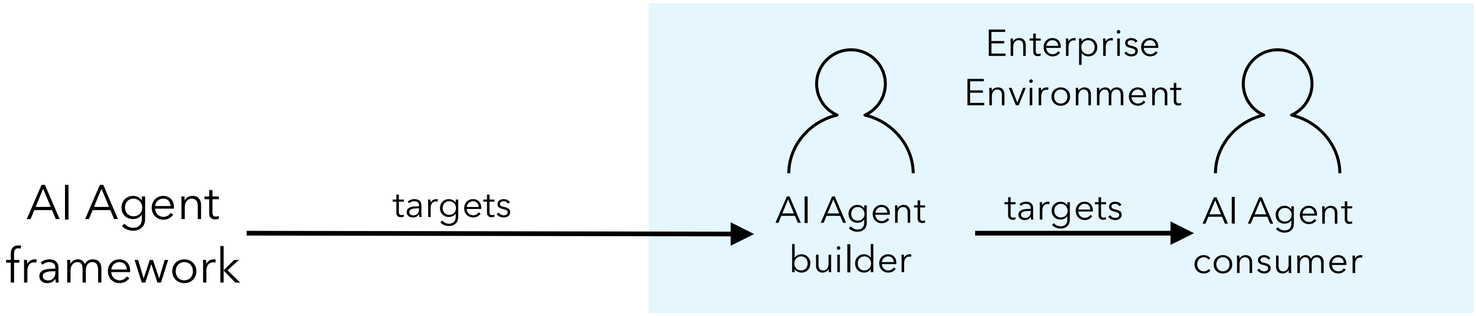
Defining optimal scope is ongoing work

- MCP "rules"
  - Associates keywords with tool metadata
  - Only clients that support rules can benefit

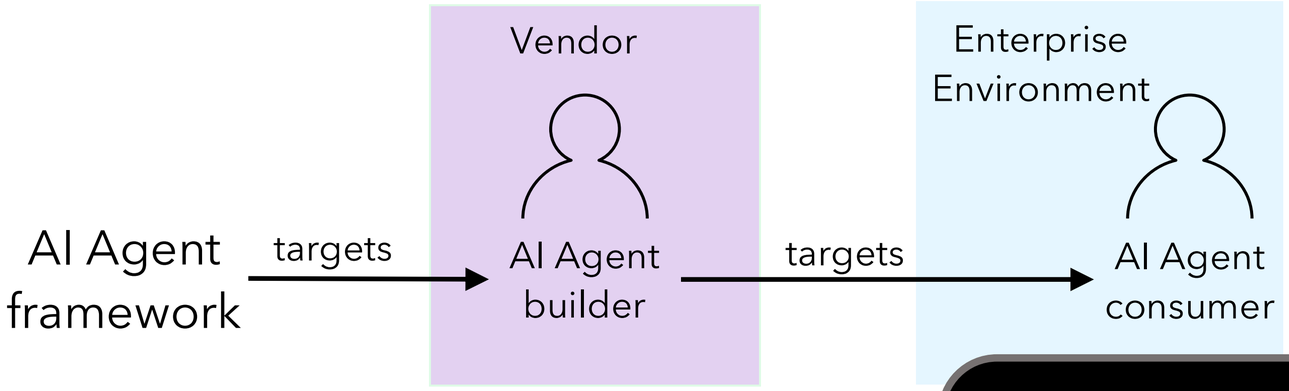


SNIA | CST A SNIA Community

Agentic AI - Serviceability and Evaluation of Success



Frameworks seem to target enterprises building their own agents for their business use cases



What will a Storage persona accept?

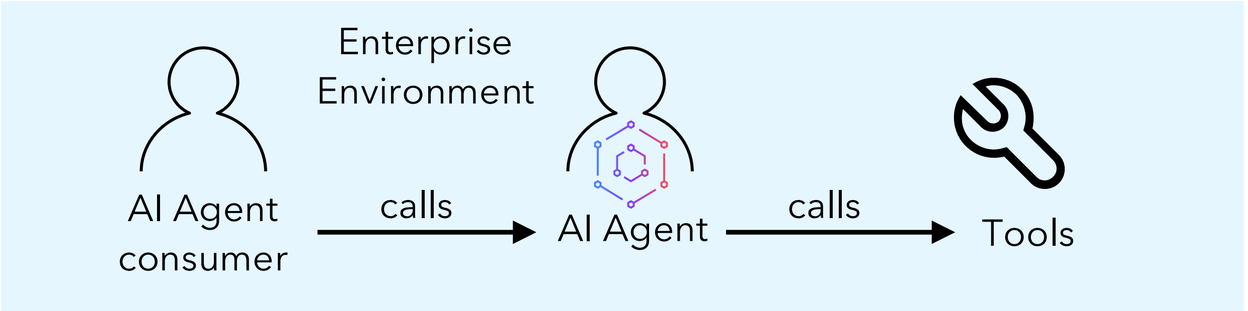
The value provided needs to outweigh the work created

What happens when a vendor is shipping an agent to run elsewhere?

How to evaluate?
How to debug?

Agent 🤖 (thought) : To answer this question, I need to identify enterprise applications that run on Linux, check if they are currently running on the system, and validate their data if they are. I will use the 'FindRunningProcesses' tool to check for running applications, and the 'MongoDBDataValidator' tool to validate the data if the applications are found.

Agentic AI - Security Considerations



What authority do the agent and tools run under?
That of the calling user?
A machine identity?

Under what identity and access control is the agent running?

And the tools? (actions the agent can take)

What tools does the agent have?

What does a mistake look like?

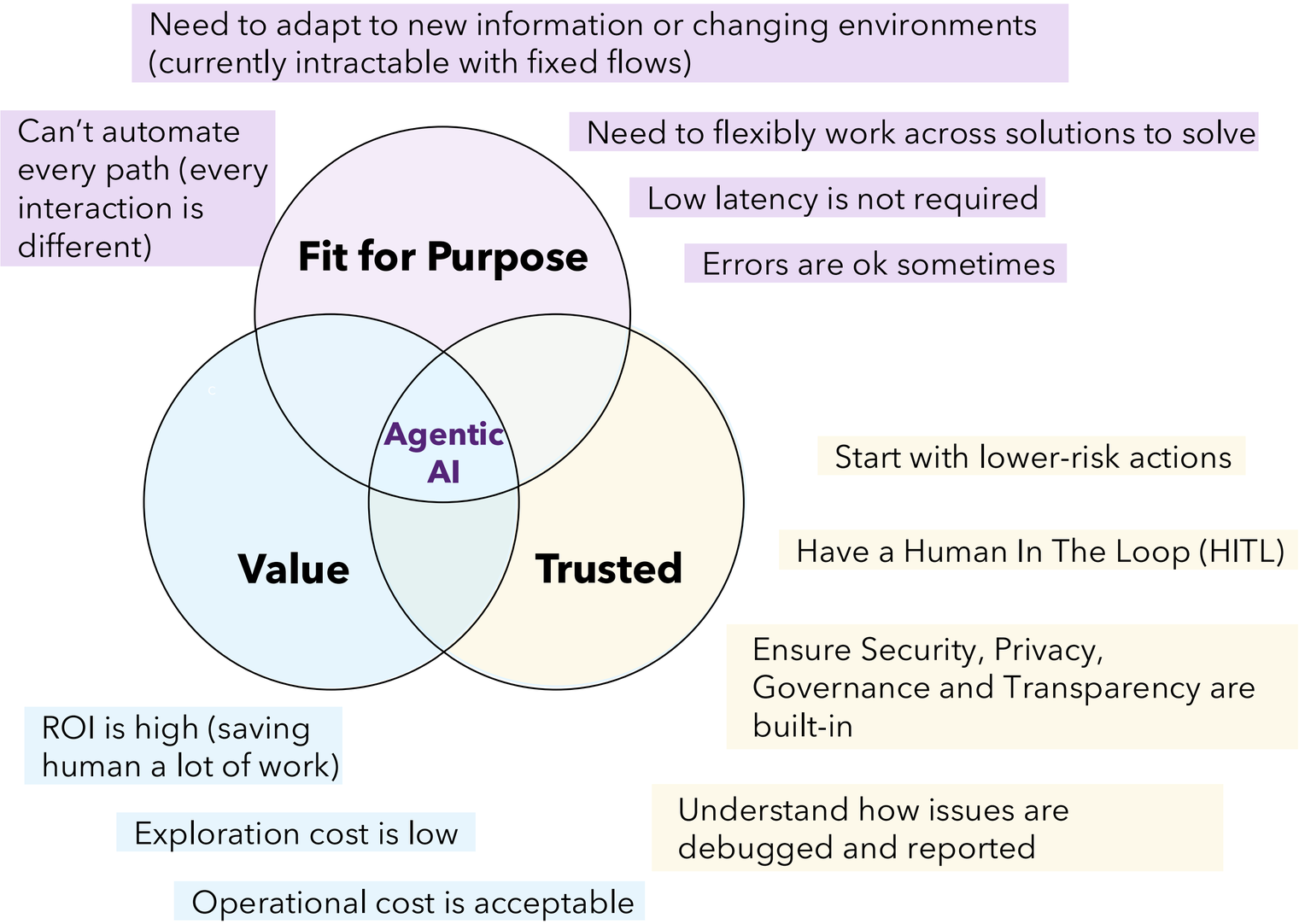
Suggestion:

Start with small use cases that do no harm, with a Human In The Loop (HITL)

Summary

When considering Agentic AI solutions for Enterprises:

- Check if the use case is appropriate for agentic AI and provides value
- Verify the security characteristics meet your requirements
- Understand how serviceability will be provided



Resources

- Link to Code:
  https://github.com/IBM/agentic-ai-cyberres
- BeeAI agentic AI framework
  https://github.com/i-am-bee
- Model Context Protocol (MCP)
  https://modelcontextprotocol.io/introduction

Questions?

Thanks for Viewing this Webinar

- Please rate this webinar and provide us with feedback
- This webinar and a copy of the slides are available at the SNIA Educational Library
- A Q&A blog from this webinar will be posted at snia.org/blog
- Follow us on LinkedIn and X for future webinar dates

Thank You

