

Agentic AI in Industry

Tools, workflows and controlled autonomy

Ayush Kumar Shah, Ph.D.

18 September 2026



My Background

2015 – 2019

Kathmandu University, Nepal — BE in Computer Engineering



2019 – 2020

Fusemachines, Nepal — ML Engineer, my first job



2020 – 2025

Rochester Institute of Technology, NY, USA — PhD in Computer Science

2022 (14 weeks)

Amazon Alexa, Sunnyvale, CA — Applied Scientist Intern, speaker recognition



2025 – Present

Meta, Menlo Park, CA — Research Scientist, video enhancement and Applied AI



Some of my Ph.D. Research Projects

The screenshot shows the MathDeck search interface. The search bar contains the query "language model $P(X|x)$ ". Below the search bar, there are two search results. The first result is titled "(1996) An Overview of Corpus-Based Statistics-Oriented ((CBSO)) Techniques for Natural Language Processing" by Su, Keh-Yih and Chiang, Tung-Hui and Chang, Jing-Shin. It includes a BibTeX entry and a link to the paper. The second result is titled "(1999) Unsupervised Models for Named Entity Classification" by Collins, Michael and Singer, Yoram. It also includes a BibTeX entry and a link to the paper. The interface includes a "MathDeck" logo, a search bar, and a results section with a "By rank" filter.

MathDeck

Formula-and-text search

The screenshot shows the ReactionMiner search interface. The search bar contains the query "Burke Group". Below the search bar, there are search results. The first result is titled "Reaction 3" and shows a chemical reaction scheme. The second result is titled "Molecule 3" and shows a chemical structure. The interface includes a "ReactionMiner" logo, a search bar, and a results section with a "By rank" filter.

MultiModal Chemical Search

Chemical structures and reaction descriptions

Agentic AI in Industry

Tools, workflows and controlled autonomy

01 Why agents

02 Context, skills and MCP

03 Everyday workflows

04 A second brain

05 Personal tools, voice and remote

06 Agent teams

07 Controlled autonomy

08 Production and research

At Meta, small failure rates reach many people

3.60B

daily active people

Facebook, Instagram, Messenger, WhatsApp
June 2026 reported average

0.1% of that population

3.60 million

people in a hypothetical
impact scenario

AI agent and agentic AI

Term	Working definition	Example
AI agent	One system choosing actions toward a goal	An assistant investigates a failing test
Agentic AI	The broader approach of goal-directed action	A briefing workflow that retrieves, checks and revises
Multiple agents	One possible architecture	A scout, reviewer and writer cooperate

The loop behind a research assistant



Learning may update memory or procedure

Context, skills and MCP

We have a working picture of an agent.

What context and capabilities make it useful?

01 Why agents

02 Context, skills and MCP

03 Everyday workflows

04 A second brain

05 Personal tools, voice and remote

06 Agent teams

07 Controlled autonomy

08 Production and research

A tool for each kind of delegation

Your task	A useful starting point	First experiment
Work inside a repository	Claude Code + project instructions	One small change with a check
Recurring personal workflow	Cron or OpenClaw or Muse	Draft a morning briefing
Understand a source collection	NotebookLM	Audio discussion of two documents
Repeated team procedure	A skill + connected tools	Meeting notes from a transcript

An example CLAUDE.md

Project: research briefing

Purpose: produce a cited daily draft.

Inputs: sources/ and preferences.md

Outputs: drafts/YYYY-MM-DD.md

/init

/context

Keep source URLs beside each claim.

Say when no relevant papers are found.

Do not invent results or citations.

Record unresolved questions in state.md.

A task prompt with a checkable finish

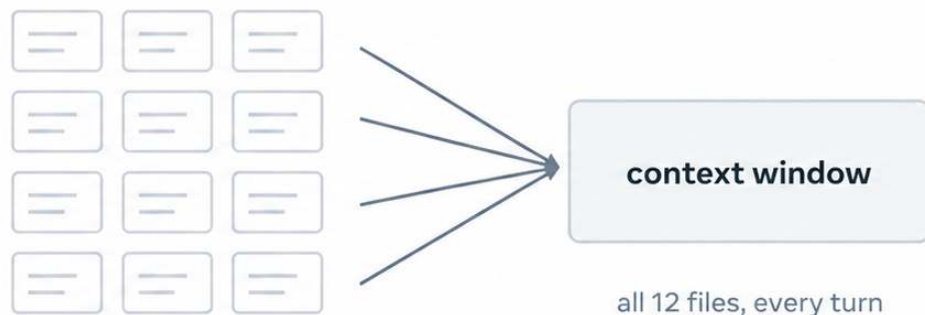
Read ~~CLAUDE.md~~ and sources/demo-paper.md.
Create drafts/example-brief.md.
Include the finding, limitation and URL.
Label this paper as a fictional example.

Before finishing:
Check every factual sentence against the source.
List any unsupported claims you removed.

Tip: Use AI to generate a good prompt that you want to use for the AI

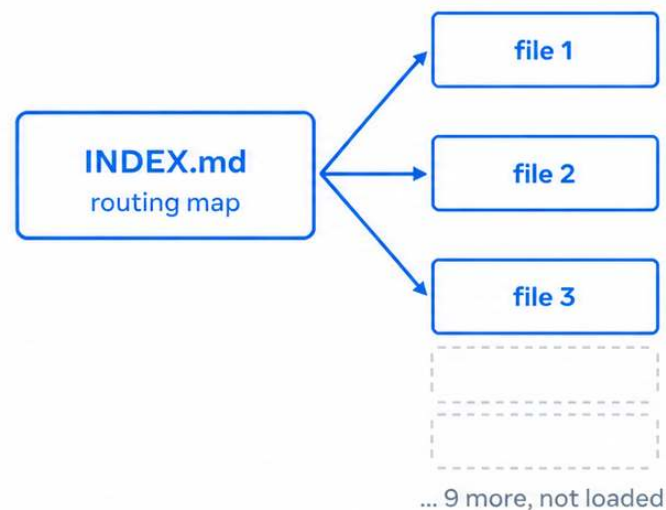
An index selects the relevant evidence

DUMP EVERYTHING



Extra files consume context.
Load the evidence relevant to the question.

SHIP AN INDEX



This example selects 1-3 relevant files.

At Meta, agents need a map of the codebase

59 context files

Navigation for 4,100+ code files

Commands and key files

Hidden compatibility rules

Cross-module dependencies

Periodic checks for stale context

```
# Module: event ingestion
```

```
Test: run compatibility tests
```

```
Key files: schema/, readers/
```

```
Never reuse serialized IDs.
```

```
Old readers still depend on them.
```

```
See also: dependency-index.md
```

A meeting-notes skill

```
name: meeting-notes
description: Draft notes from a transcript.
disable-model-invocation: true
```

Extract decisions and explicit action owners.
Quote the supporting transcript lines.
Mark missing owners or dates as unresolved.
Write a draft. Never send it automatically.

Superpowers brainstorming before coding

Start

“Help me plan a daily research briefing.”

Clarify

Audience? Sources? Delivery? Failure behavior?

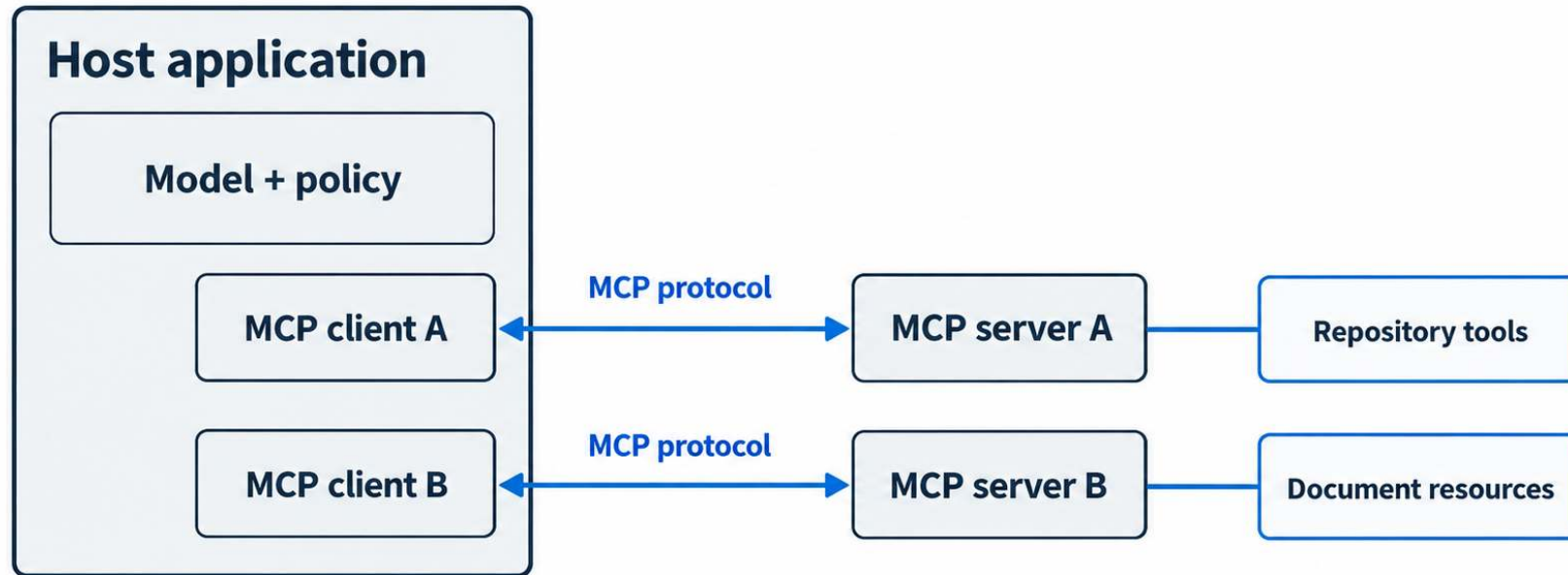
Compare

A scheduled script or an adaptive agent?

Approve

Review the design before implementation

MCP separates the host from integrations



Host policy and server authorization still apply

Connecting a service with MCP

```
claude mcp add --transport http notion \  
https://mcp.notion.com/mcp
```

```
# Inside Claude Code  
/mcp
```

Authenticate the account you intend to use.
Inspect tools and access before a real task.

Script, skill, MCP or agent?

You need...

Reach for...

The same known sequence every time

A script or fixed workflow

A repeatable judgment procedure

A skill

A standard connection to a service

An MCP integration

Adaptive choices based on results

An agent loop

Agentic AI in Industry

NEXT · 03

Everyday workflows

The agent now has instructions, skills and tools.

**What can we delegate in an ordinary
workday?**

01 Why agents

02 Context, skills and MCP

03 Everyday workflows

04 A second brain

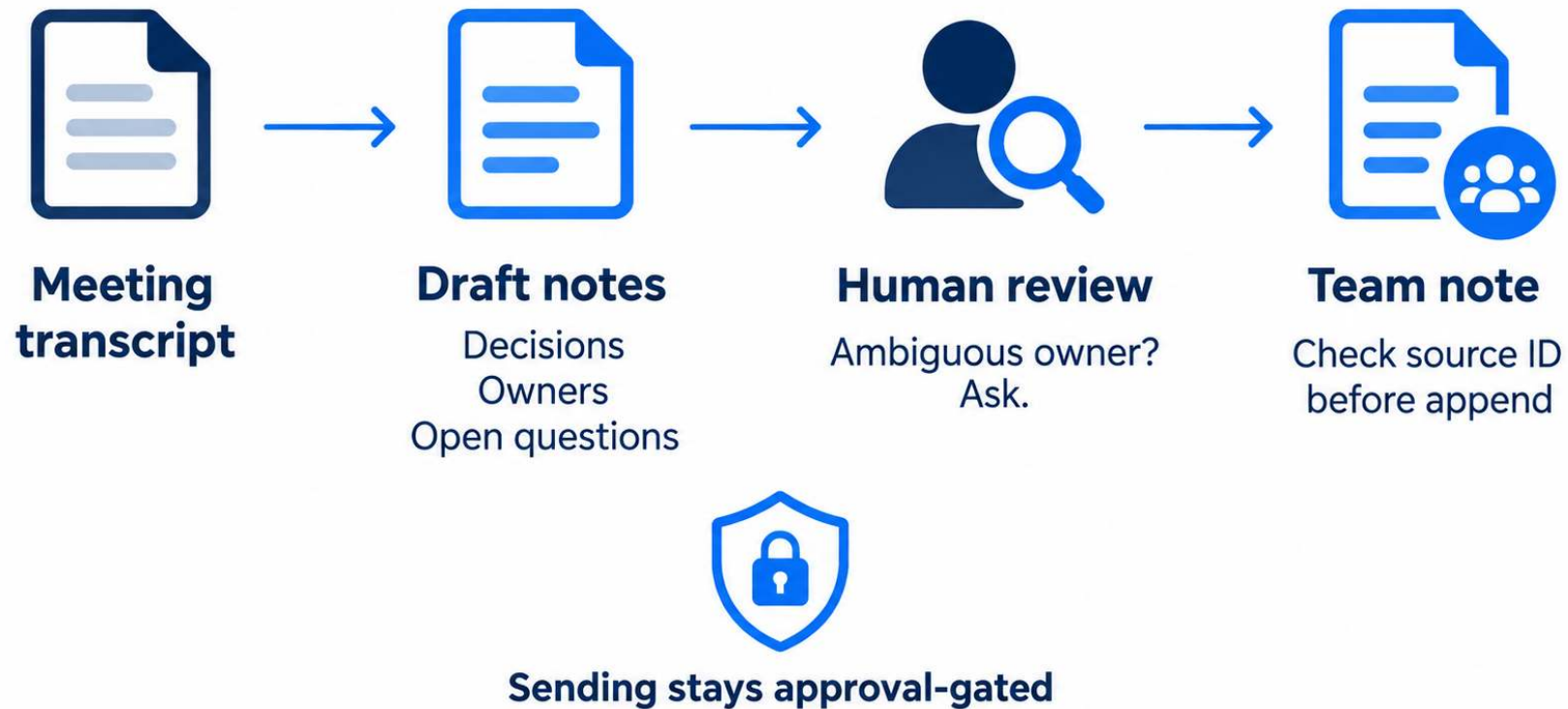
05 Personal tools, voice and remote

06 Agent teams

07 Controlled autonomy

08 Production and research

Meeting notes with an approval gate



A daily paper briefing



— Empty results are a valid result —

OpenClaw/Claude gives the workflow a schedule

Trigger

A scheduled job starts a bounded task

Workspace

Instructions, sources and durable state

Delivery

A draft or an explicitly authorized channel

Control

Run history, pause, retry and review

Agentic AI in Industry

NEXT · 04

A second brain

Daily workflows produce useful notes and evidence.

How do those outputs become reusable knowledge?

01 Why agents

02 Context, skills and MCP

03 Everyday workflows

04 A second brain

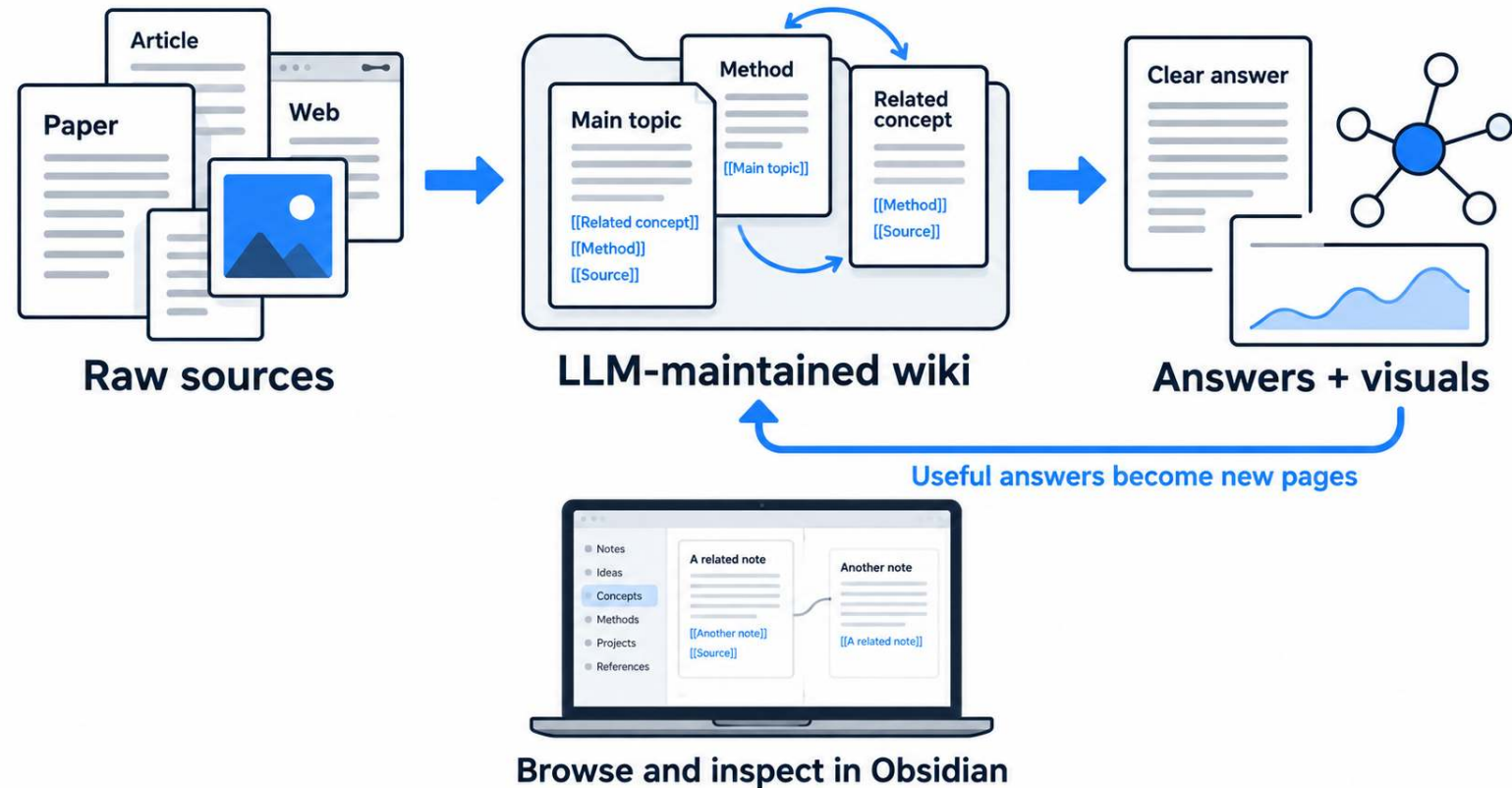
05 Personal tools, voice and remote

06 Agent teams

07 Controlled autonomy

08 Production and research

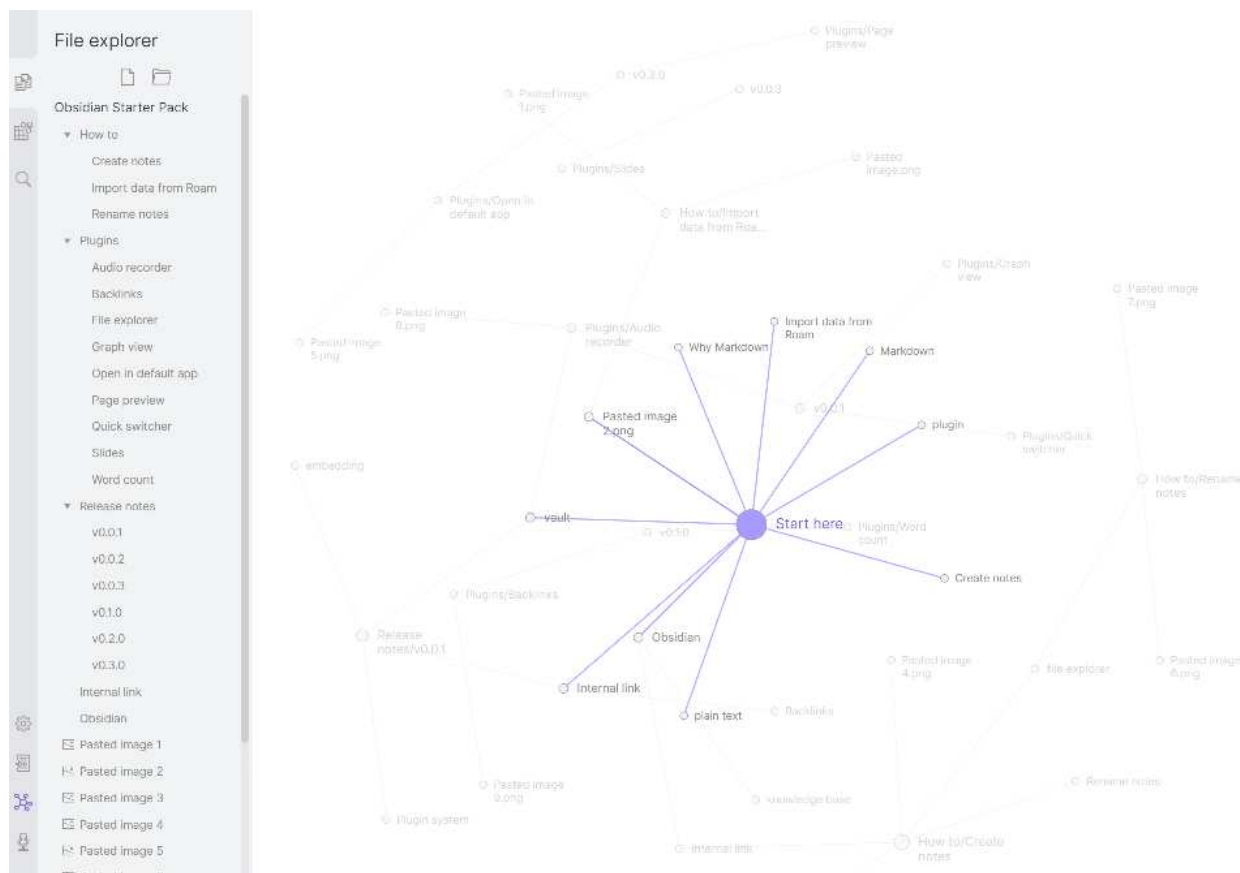
Karpathy's LLM-maintained personal wiki



PARA organizes knowledge around action

Folder	What belongs here	Example
Projects	An outcome you are actively pursuing	Prepare the Fuse fellowship session
Areas	An ongoing responsibility	Research practice or team operations
Resources	Reference material and interests	Document AI papers and useful techniques
Archive	Inactive or completed material	Finished project notes and old plans

Obsidian is the shared Markdown workspace



Readable files

The person and agent use the same .md notes

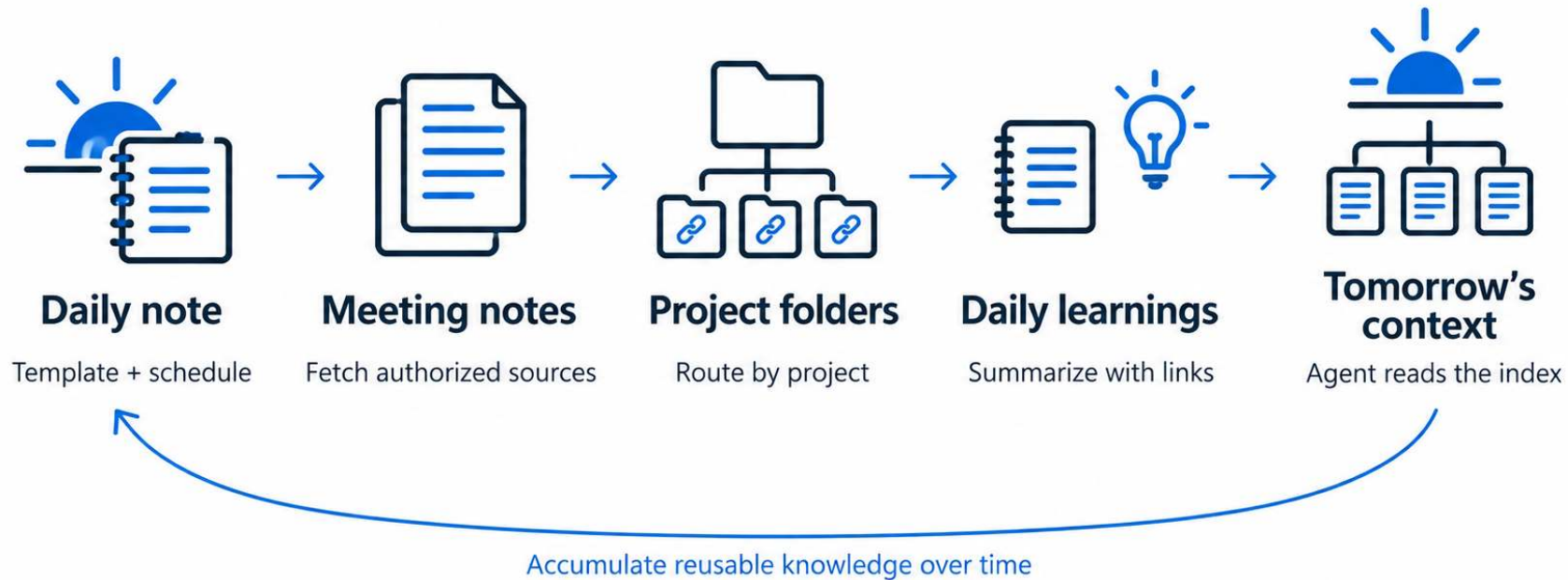
Connected context

Links and graph view expose relationships

Reusable structure

Daily Notes, Templates and Templater

My daily Obsidian automation loop



A daily note template you can adapt

```
# <% tp.date.now("YYYY-MM-DD") %>
```

```
## Schedule
```

```
## Meetings
```

```
## Project progress
```

```
## Learnings
```

```
## Tomorrow
```

```
Project: [[Projects/Fuse-session/Overview]]
```

```
Source links stay beside imported claims.
```

Three jobs keep the day organized

Trigger	Concrete output	Check before finishing
Start of day	Today's note and a schedule draft	Correct local date and time zone
After meetings	Notes linked to the matching project	Source ID, owner and duplicate check
End of day	Learning summary and tomorrow's context	Evidence links and unresolved questions

NEXT · 05

Personal tools, voice and remote

The second brain preserves context between tasks.

How do we use it without staying at the keyboard?

01 Why agents

02 Context, skills and MCP

03 Everyday workflows

04 A second brain

05 Personal tools, voice and remote

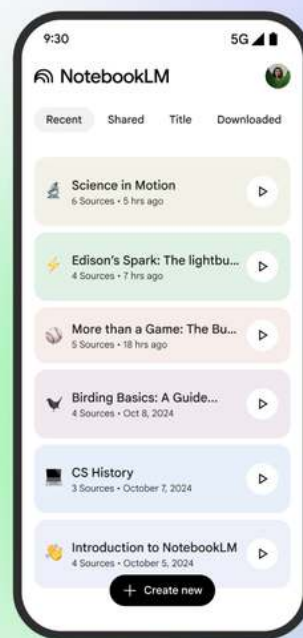
06 Agent teams

07 Controlled autonomy

08 Production and research

NotebookLM turns sources into listening

 NotebookLM



A NotebookLM commute demo

Sources

Add two papers you want to compare

Audio

Generate an Audio Overview

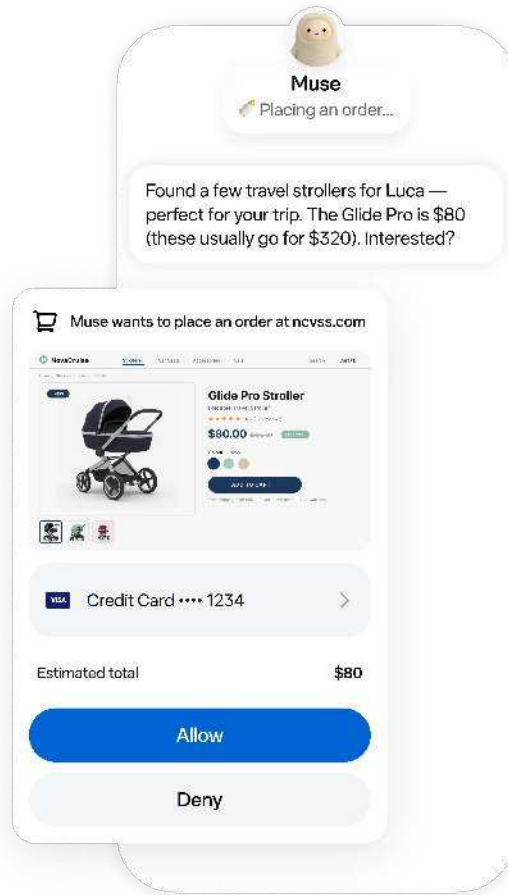
Conversation

Join interactive mode and ask a question

Follow-up

Return to the source for exact evidence

Muse: a personal agent at muse.ai



Try a bounded task

“Compare three laptops within my budget. Show the tradeoffs. Ask before buying.”

Inspect the controls

Review permissions before approving. Check the activity log and goal progress.

<https://muse.ai/>

Voice captures the context while it is fresh



Speak the context

Goal, constraints, examples

Review the transcript

Names, paths, numbers

Send the task

Same permissions apply

Claude /voice, Wispr Flow and Superwhisper

Tool	Practical use	Try it
Claude Code /voice	Dictate directly into a CLI prompt	Enable /voice, hold Space, review, Enter
Wispr Flow	Dictate into other apps too	Try quiet speech near the microphone
Superwhisper	Voice-to-text with configurable modes	Choose a mode suited to the task

Remote Control puts the session on my phone

```
# Start from the project directory  
claude remote-control
```

```
# Or connect an existing session  
/remote-control
```

Open the session link or scan its QR code.
Check progress, send guidance, review approvals.
Execution stays on the local machine.

I sleep. The machine stays awake.



Agentic AI in Industry

NEXT · 06

Agent teams

We can listen, dictate and steer work from a phone.

When should several agents work together?

01 Why agents

02 Context, skills and MCP

03 Everyday workflows

04 A second brain

05 Personal tools, voice and remote

06 Agent teams

07 Controlled autonomy

08 Production and research

A concrete team request

Create a team to prepare a research briefing.

Scout: find relevant papers and source IDs.

Reviewer: challenge claims and limitations.

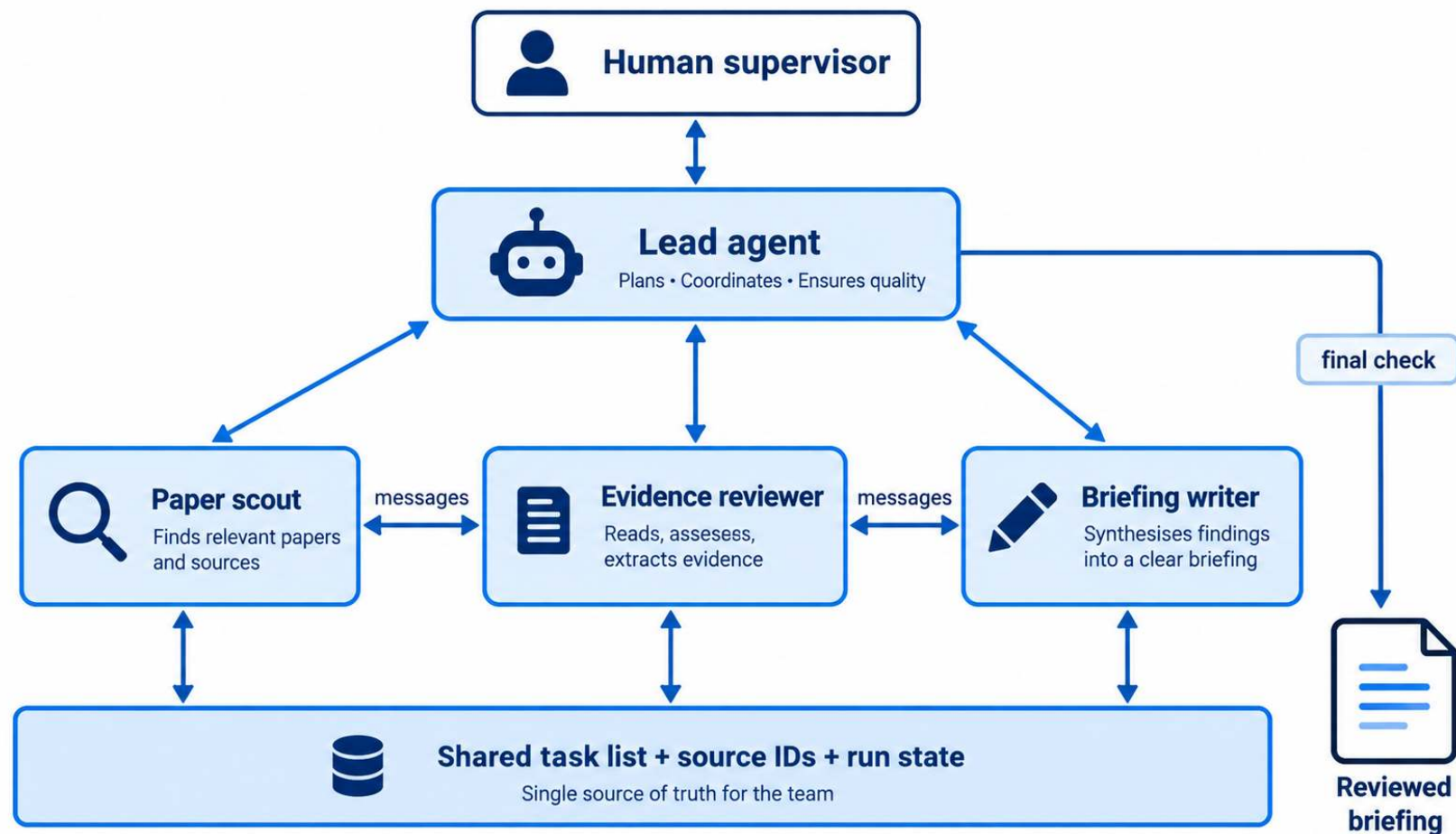
Writer: draft only from reviewed evidence.

Use separate working files.

Lead owns the final brief and resolves conflicts.

Stop after 3 papers or the agreed budget.

A fleet with distinct responsibilities



Agentic AI in Industry

NEXT · 07

Controlled autonomy

Teams can divide work and challenge one another.

What keeps their actions visible and bounded?

01 Why agents

02 Context, skills and MCP

03 Everyday workflows

04 A second brain

05 Personal tools, voice and remote

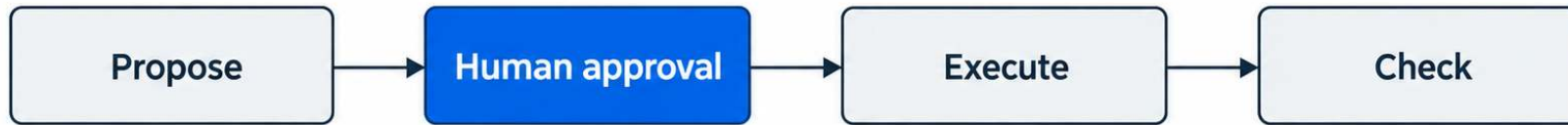
06 Agent teams

07 Controlled autonomy

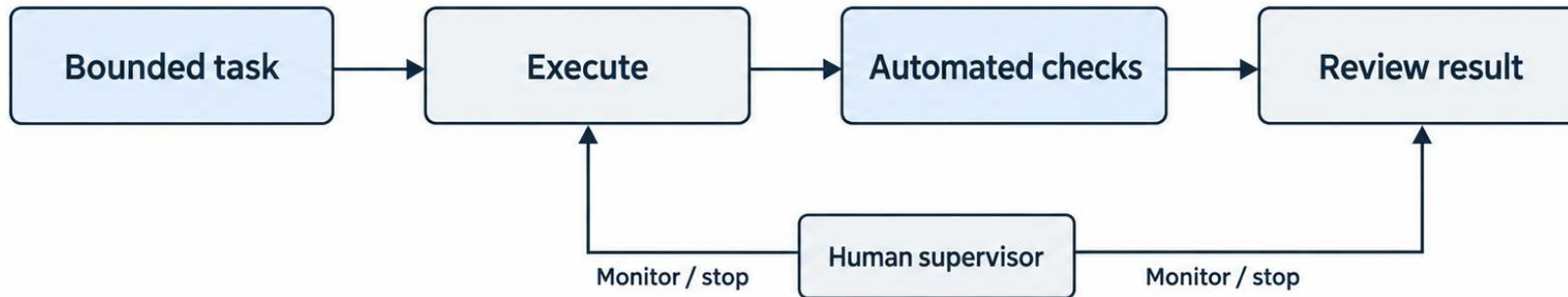
08 Production and research

Human-in-the-loop and human-on-the-loop

Human in the loop



Human on the loop



Permissions live outside CLAUDE.md

```
// .claude/settings.json (remove this comment)
{ "permissions": {
  "allow": ["Bash(npm test)",
  "ask": ["Bash(git push *)"],
  "deny": ["Read(./env)"]
} }
```

/permissions

An audit trail you can follow

Event	Record
Task started	Task ID, goal, allowed scope and budget
Evidence retrieved	Source ID, URL, timestamp and access result
Action proposed	Exact artifact or tool arguments
Approval received	Approver and the approved action
Task finished	Result, checks, unresolved items and state

Self-learning through reviewed feedback



Changed behavior does not imply changed model weights

Agentic AI in Industry

NEXT · 08

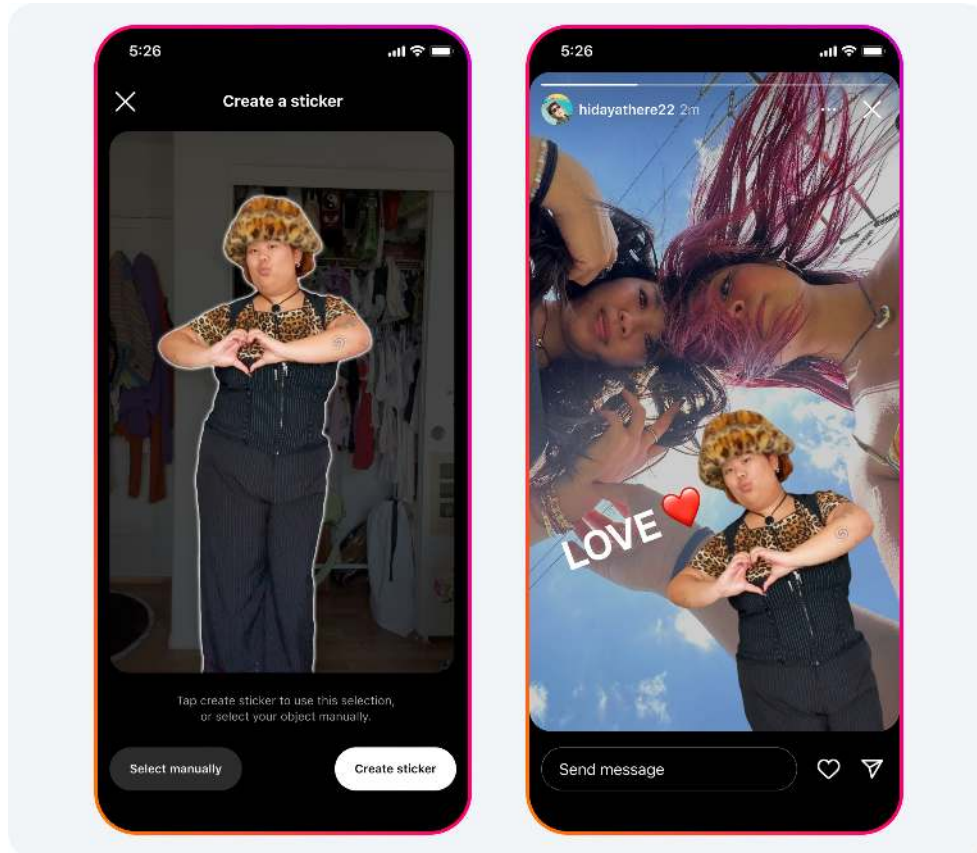
Production and research

Permissions, evidence and state make work reviewable.

What changes when real people depend on it?

- 01 Why agents
- 02 Context, skills and MCP
- 03 Everyday workflows
- 04 A second brain
- 05 Personal tools, voice and remote
- 06 Agent teams
- 07 Controlled autonomy
- 08 Production and research**

At Meta, AI has to fit the product



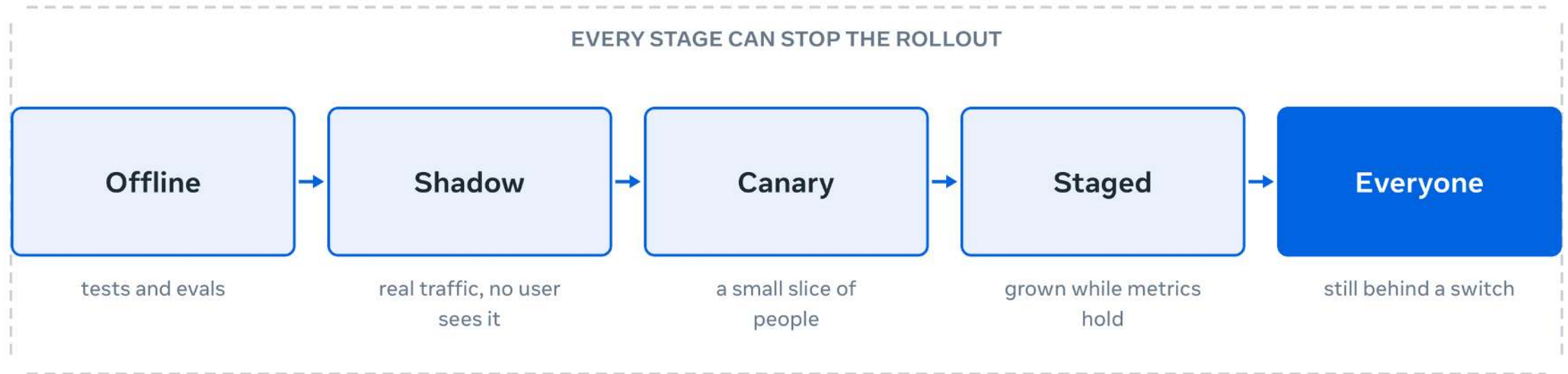
Instagram Cutouts

On-device segmentation with ExecuTorch

WhatsApp calls

Bandwidth estimation balances quality, call smoothness and device limits

A change reaches people in stages



Each gate checks user impact

Compare latency, errors and quality against the stable version.

A regression stops expansion

Restore a known-good version when safe. Alert the owner and verify recovery.

An offline win is only a hypothesis

The metric says

Accuracy went up on the eval set

Quality improved on average

The model is better

The experiment is positive

You still have to show

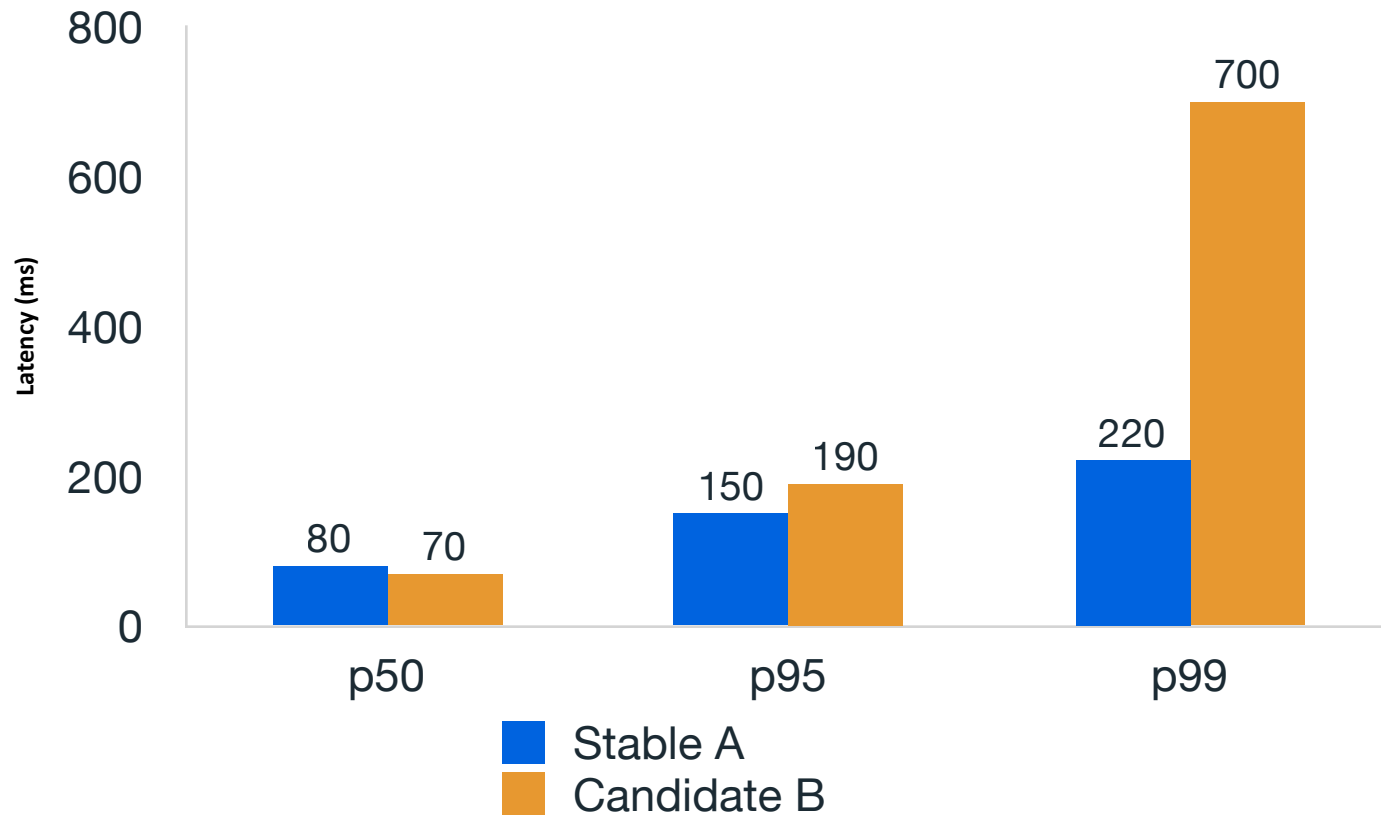
The eval set still looks like today's traffic

Important cohorts stay within quality guardrails

It fits the latency and the memory budget

The effect survives a longer run

A faster median can hide a worse user experience



Ship B?

Median improves.
Tail latency
regresses.

Check the affected
cohort before
expanding.

Illustrative release comparison. All latency values are synthetic.

Production testing includes failure and recovery

What can fail?	Test or signal
Model quality	Regression cases and quality by cohort
Device or traffic limits	Old phones, weak networks, bursts, p95/p99
Data and dependencies	Stale features, timeouts, missing inputs
Agent behavior	Wrong tool, duplicate action, weak evidence
Recovery	Kill switch, fallback and rollback rehearsal

At Meta, AI helps assess a change before release

Code change + metadata

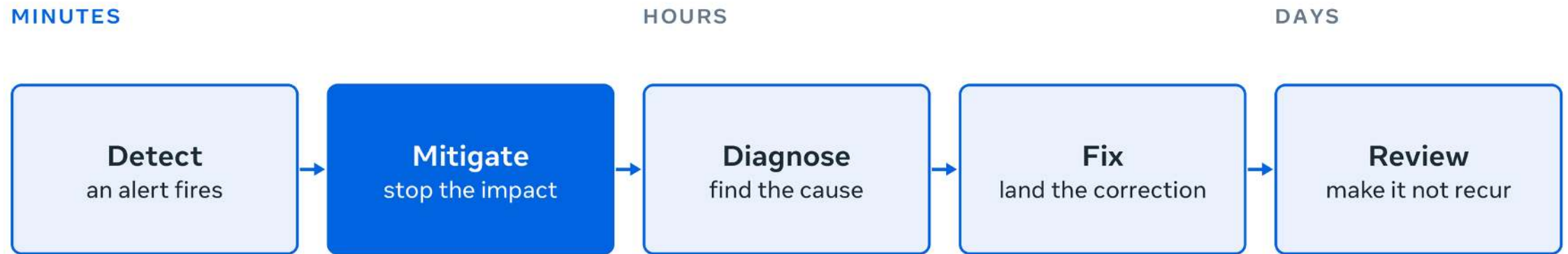
10,000+

Diff Risk Score

changes enabled during
a sensitive event in 2024

Risk estimate and suspect snippets
help focus review and release decisions

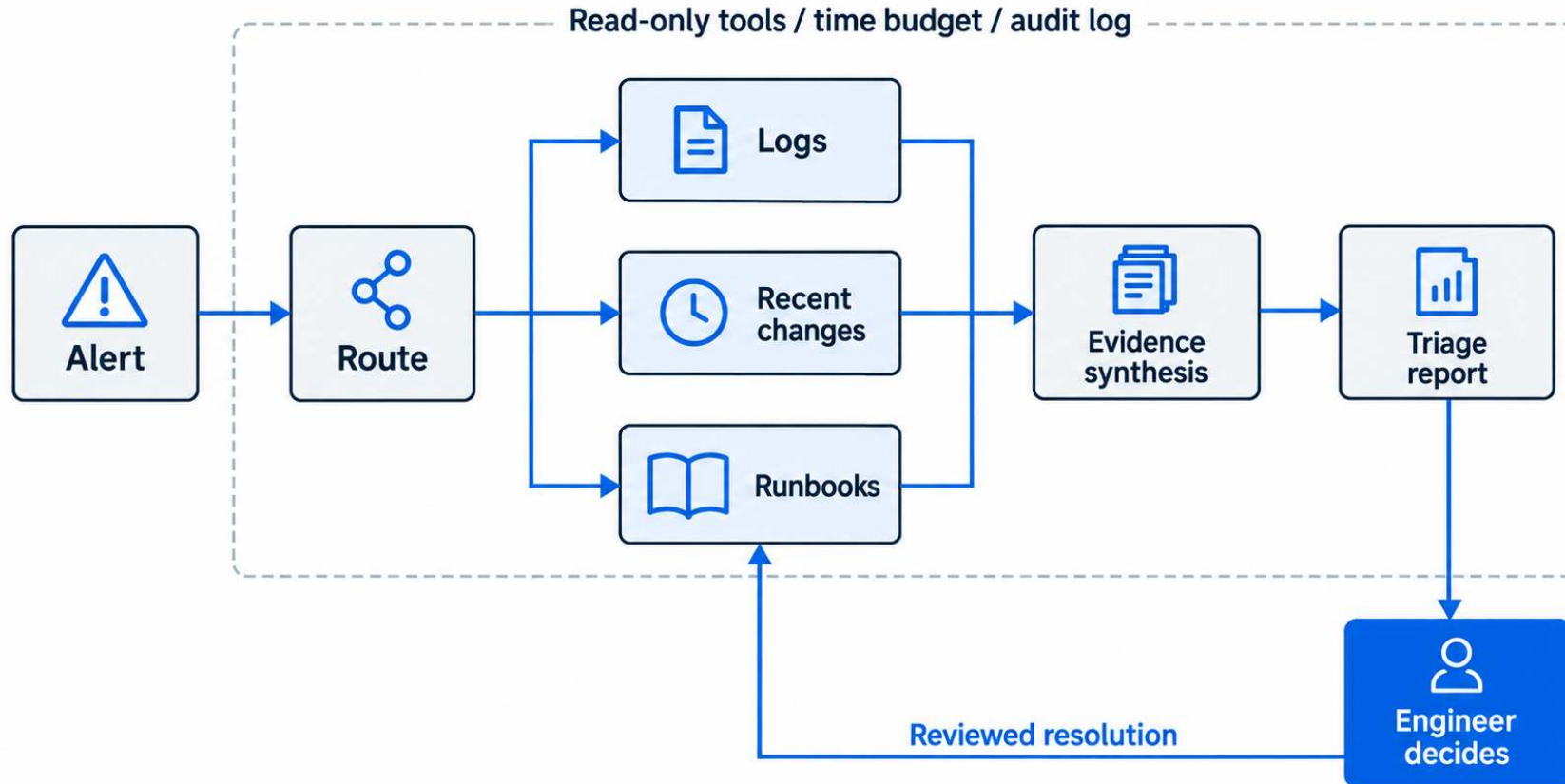
When it breaks, mitigation limits the impact



Restore service where safely possible

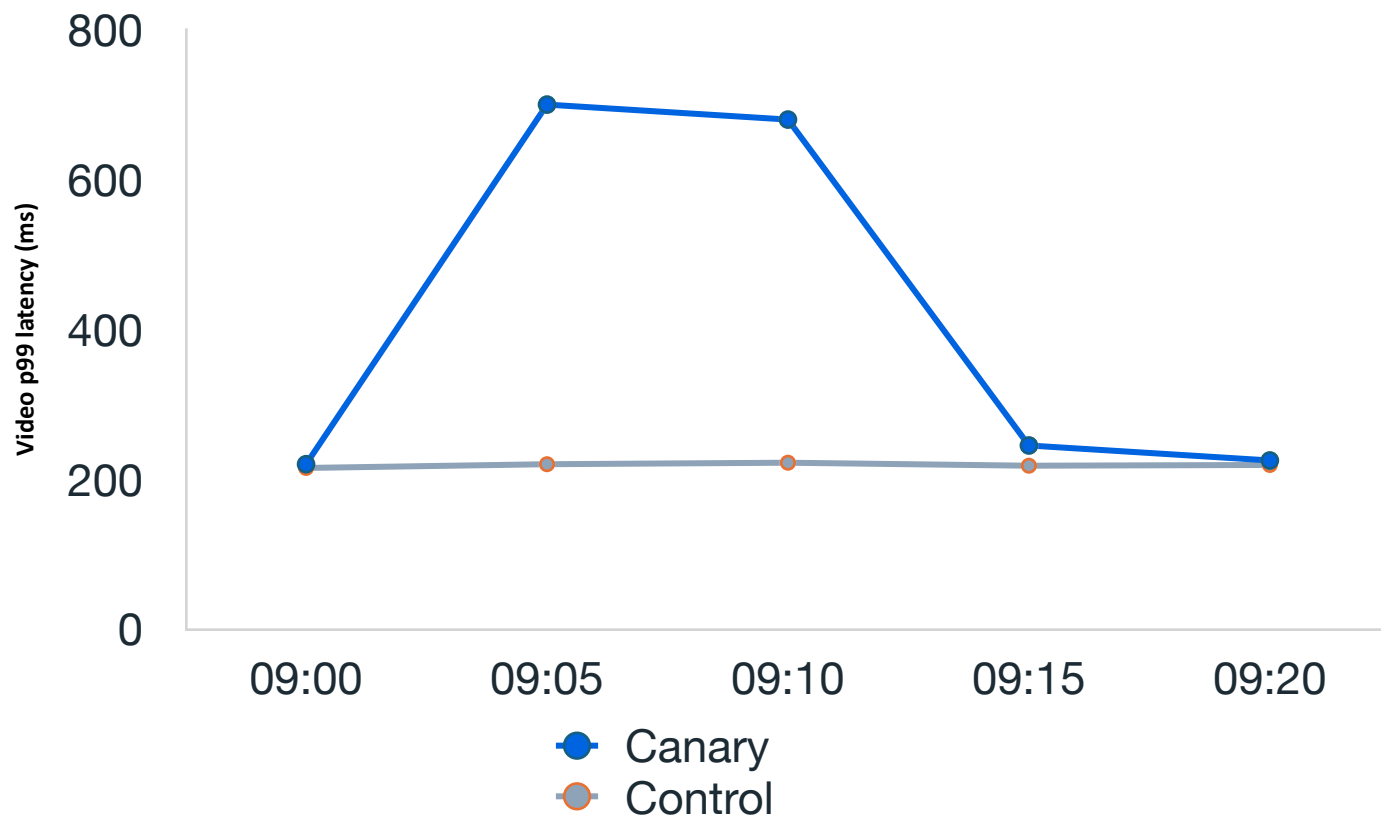
A rollback can limit harm while diagnosis continues.
The review identifies better tests, alerts and recovery steps.

A bounded triage agent gathers evidence



Autonomous investigation ends at a reviewable report

Recovery needs evidence after the action



09:05

Regression detected

09:10

Rollback approved
Then verify recovery

Synthetic incident timeline. Recovery requires more than a successful rollback command.

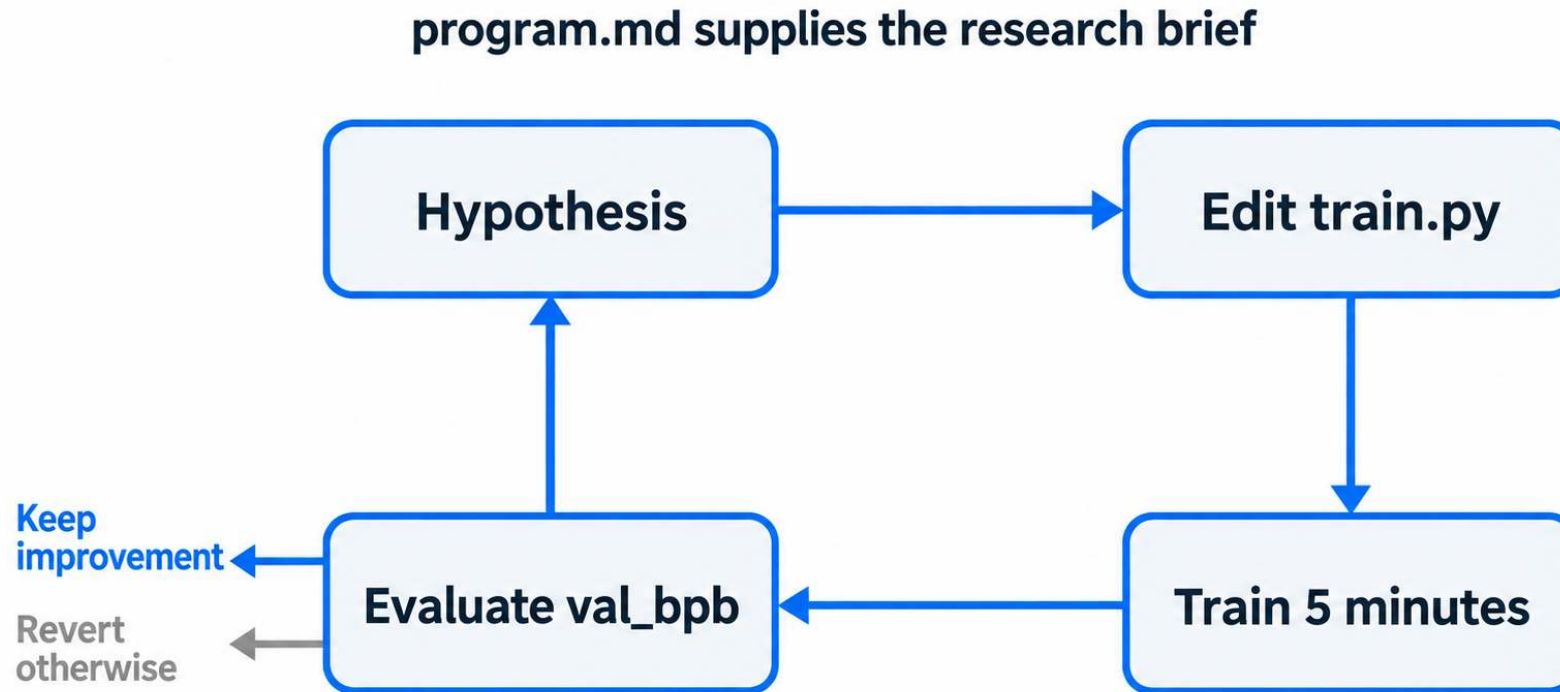
Agentic AI in Industry

08 PRODUCTION AND RESEARCH

A controlled experiment loop

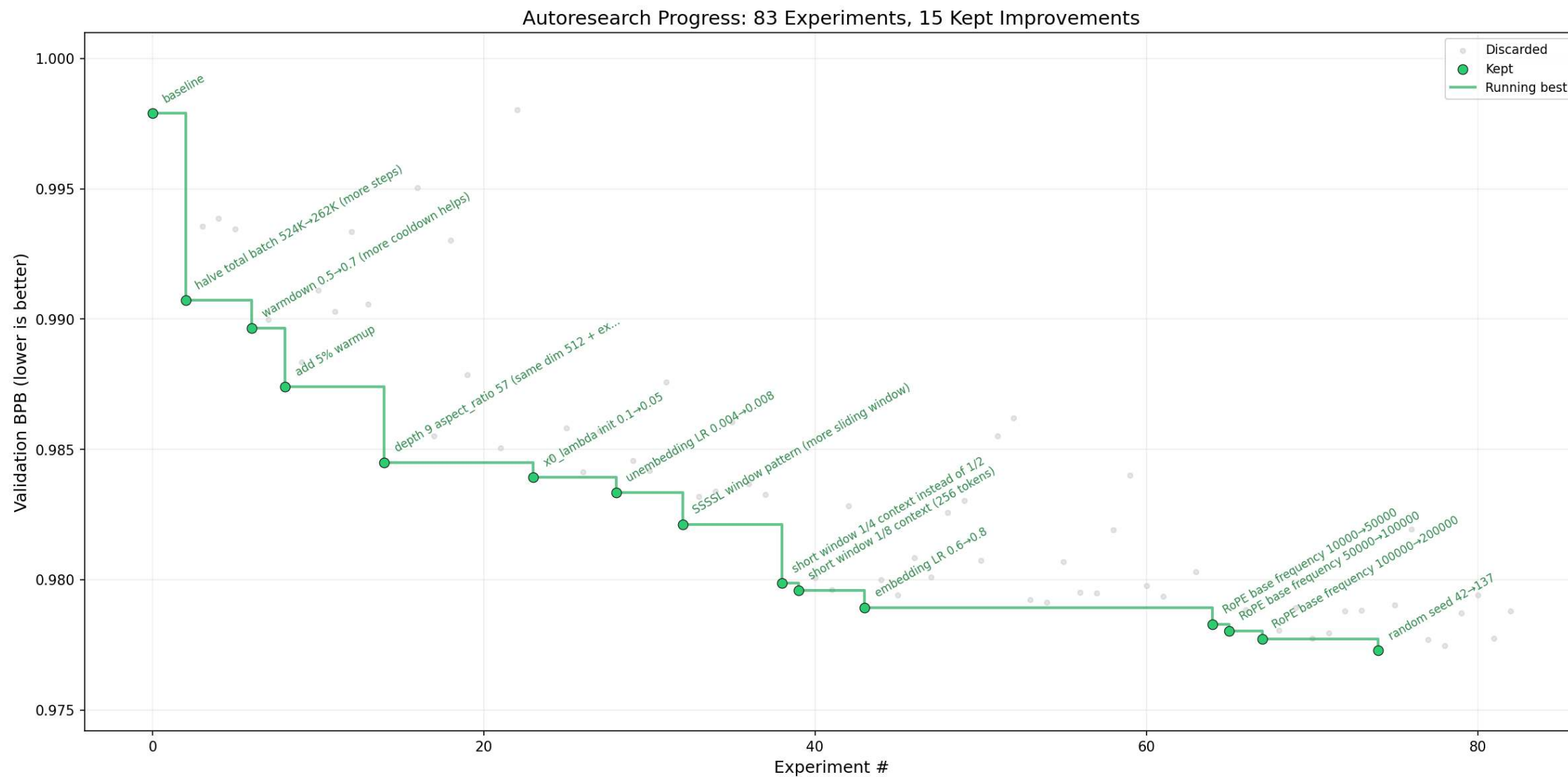
Production evidence and research evaluation
make autonomous work reviewable.

Karpathy's autoresearch experiment loop



Human owns objective, budget and held-out test

Autoresearch: progress across experiments



Agentic AI in Industry

THANK YOU

Thank you for your time

Happy to continue the conversation.

Ayush Kumar Shah

aykumars07@gmail.com,
shahayush@meta.com

linkedin.com/in/ayush7

shahayush.com

Agentic AI in Industry

DISCUSSION

Questions and discussion

- 01 Why agents
- 02 Context, skills and MCP
- 03 Everyday workflows
- 04 A second brain
- 05 Personal tools, voice and remote
- 06 Agent teams
- 07 Controlled autonomy
- 08 Production and research