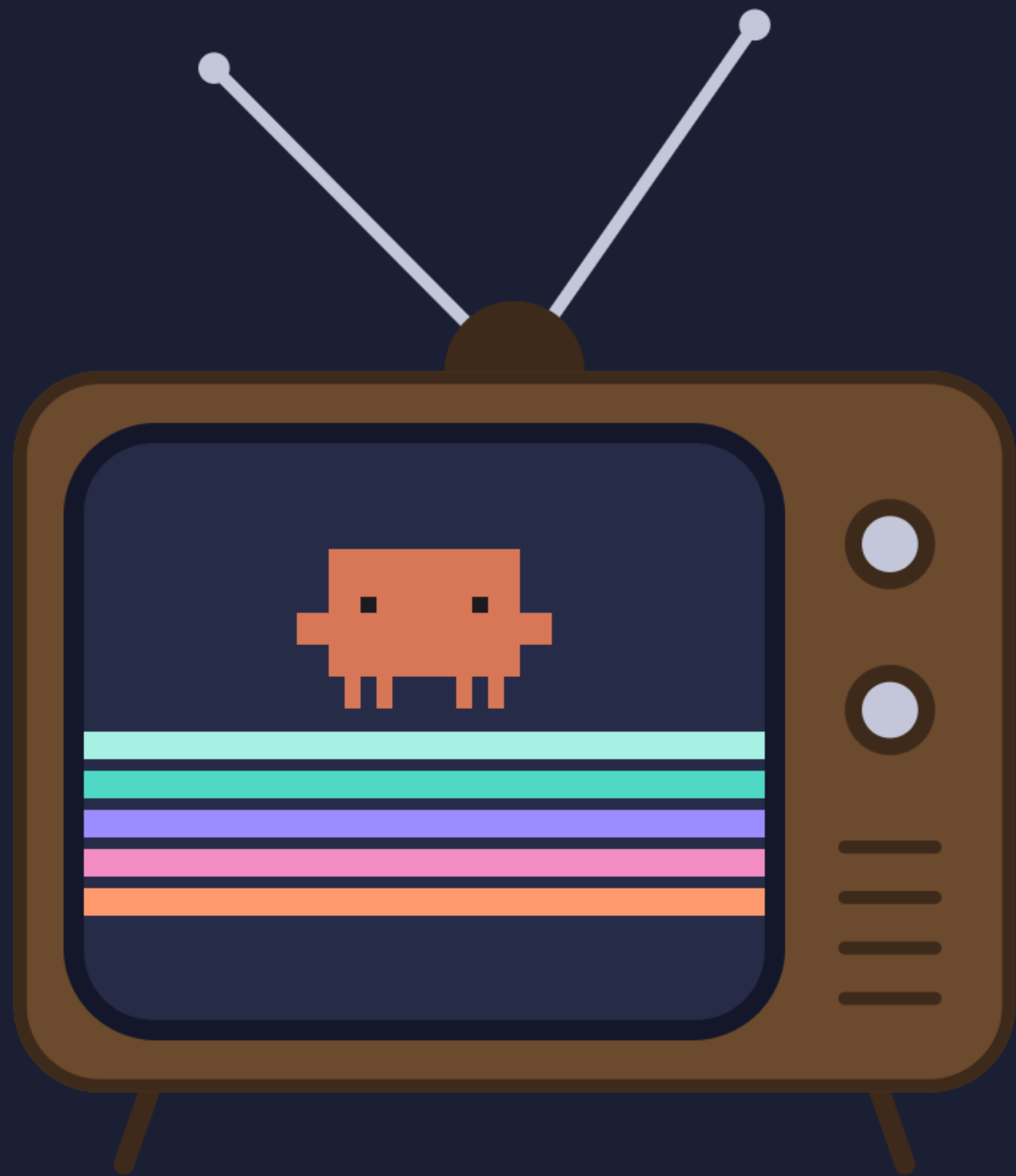


WHAT YOUR MCP SERVER DOES WHEN NOBODY'S LOOKING



Austin Parker • Independent

OVER 36,000 MCP SERVERS



Pretty good for
a dead spec!

How do you stand
out in the crowd?

AUSTIN PARKER

Independent

Co-founder, OpenTelemetry

Previously: AI Strategy @ Honeycomb

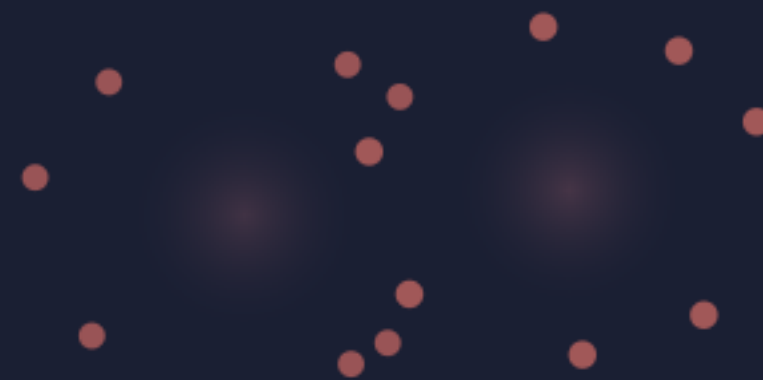


60% +

of Honeycomb's
weekly active
users are agents



AGENTS ARE DIFFERENT



run_query

get_trace

list_spans

get_columns

- | Sessions are shaped differently
- | Invisible context
- | Many harnesses, many models

MCP V0

RUNNING A QUERY



agent

API

POST /1/queries

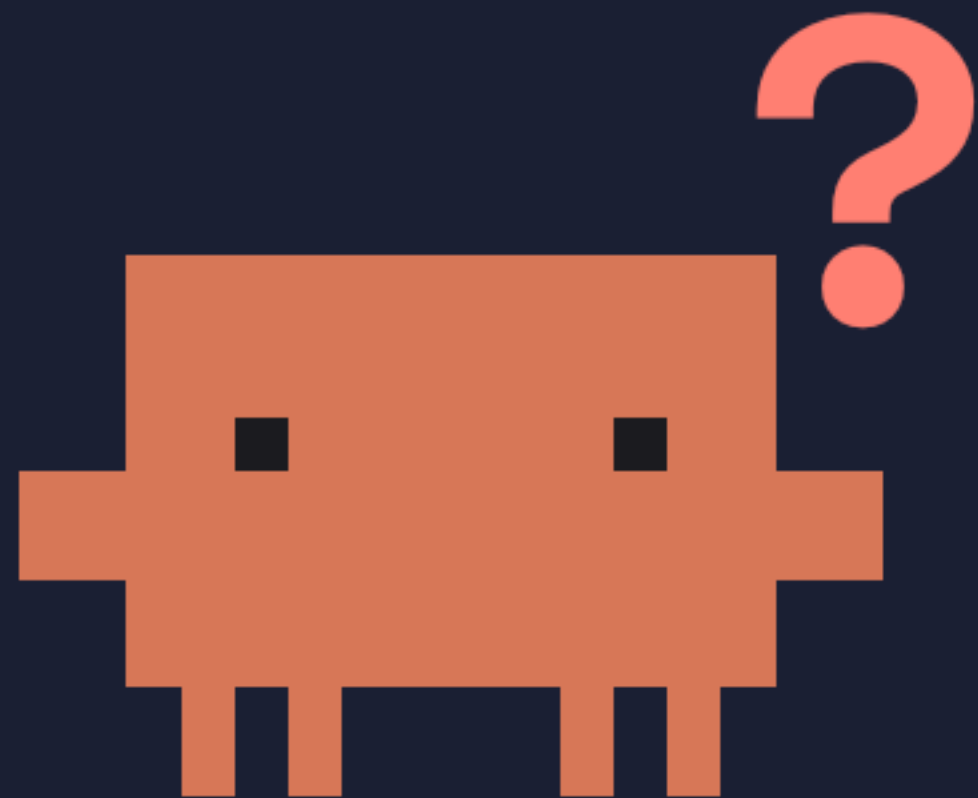
query id

POST /1/query_results

result id

GET /1/query_results/:id x4

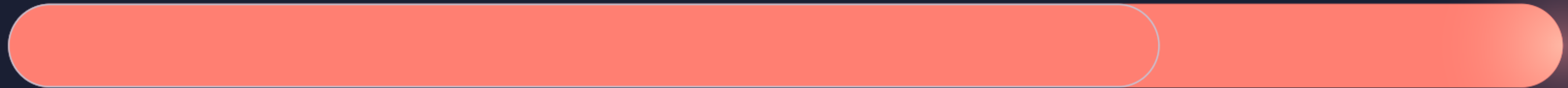
results



WRAP THE API



context window



overwhelmed!

AGENTS CAN WAIT GOOD QUERIES

“ It effectively zero-shot a real-world scenario. And it did it for sixty cents.

— Austin Parker, Honeycomb blog

THE ROAD TO V1



1 Capabilities

2 Context

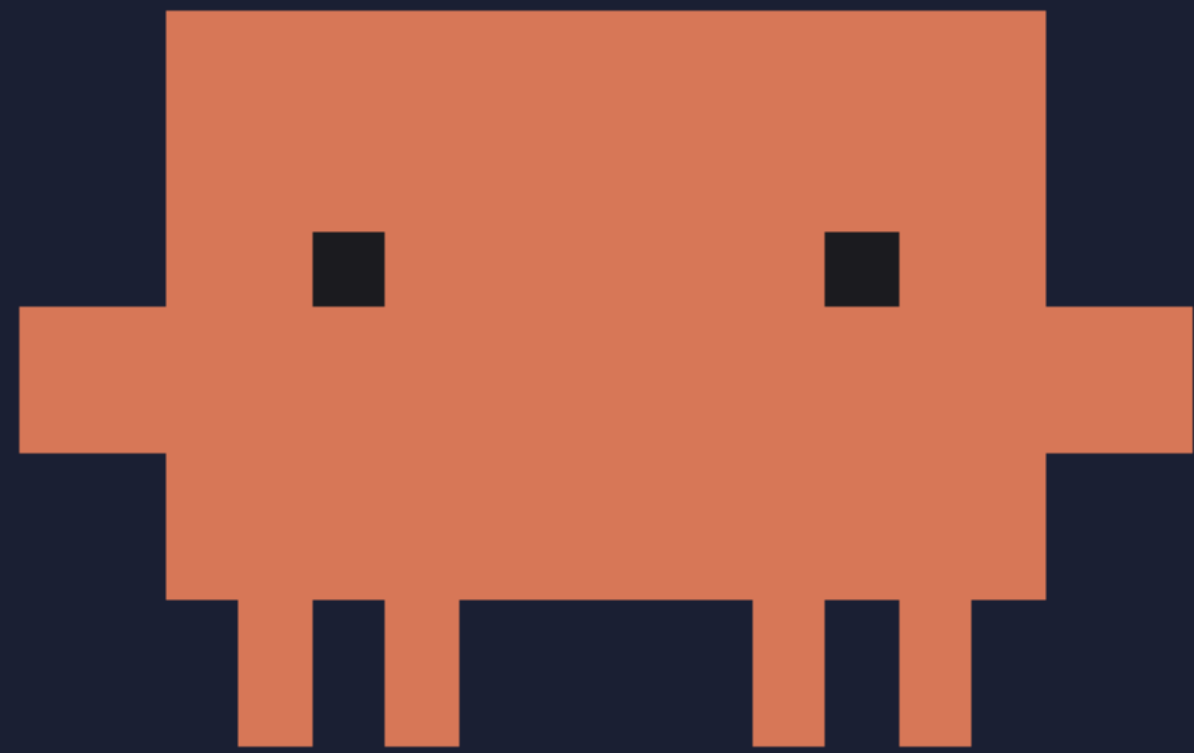
3 Documentation

DOCUMENTATION



a derived column

```
is_slow = duration_ms > 500
```



Document tool
use and answer
meaning.

**BETTER MCPS THROUGH
FAST FEEDBACK**

TRACE EVERY TOOL CALL



mcp.tool.query

295 ms

find env + dataset

16 ms

record query run

28 ms

translate query

18 ms

storage fetch

217 ms

query stats

14 ms

72% of it: the storage fetch

ON EVERY TOOL SPAN



`mcp.tool.query`

295 ms

who `mcp.client.name`

args `mcp.args.environment_slug`

query `mcp.tool.query_body`

shape `mcp.query.*_count`

errors `error.message`

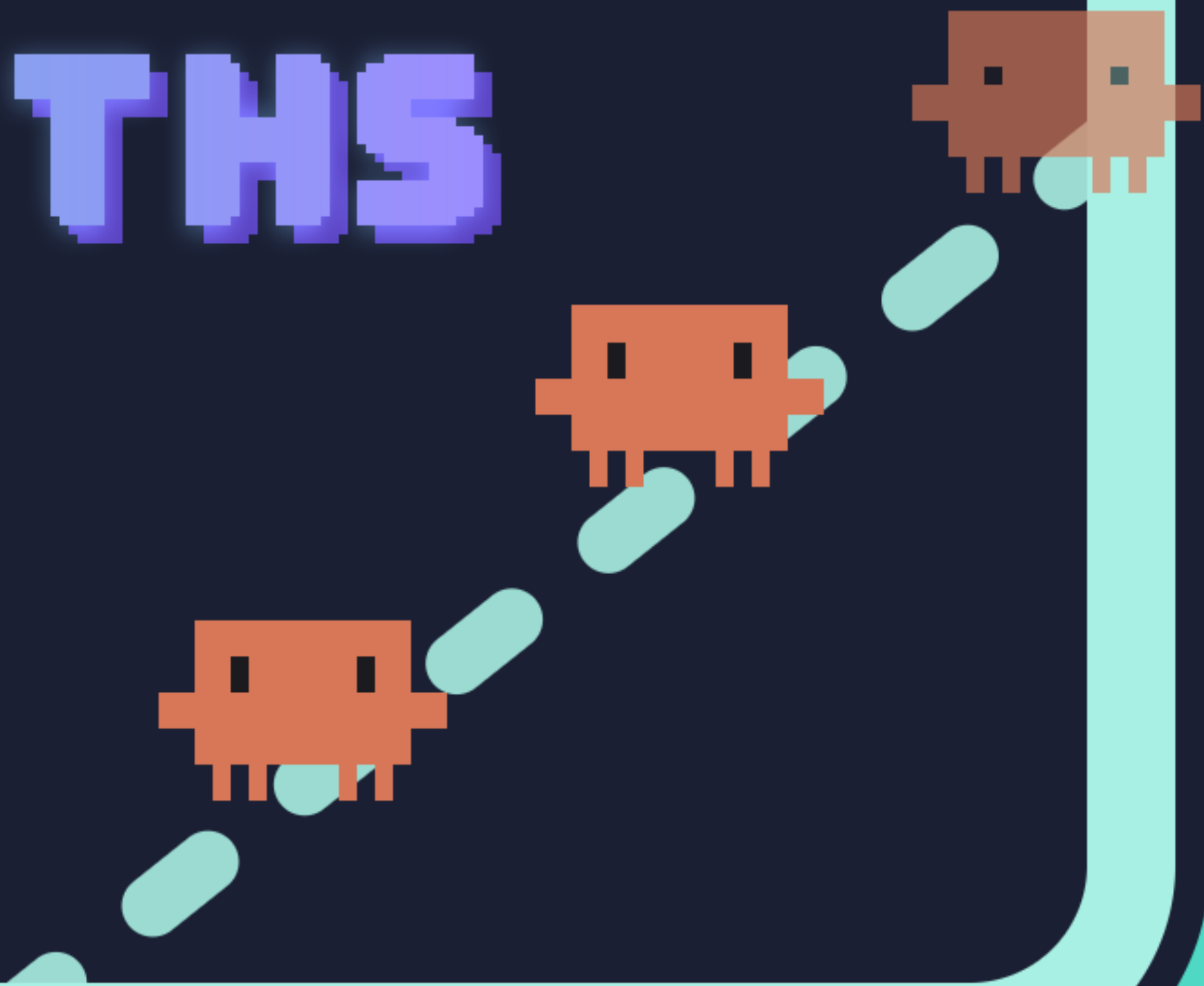
"the wrapper key is query_spec"

AGENT FEEDBACK LOOPS



- 1 Collect errors and successes
- 2 Classify and dispatch
- 3 Adversarial review
- 4 Deploy and monitor

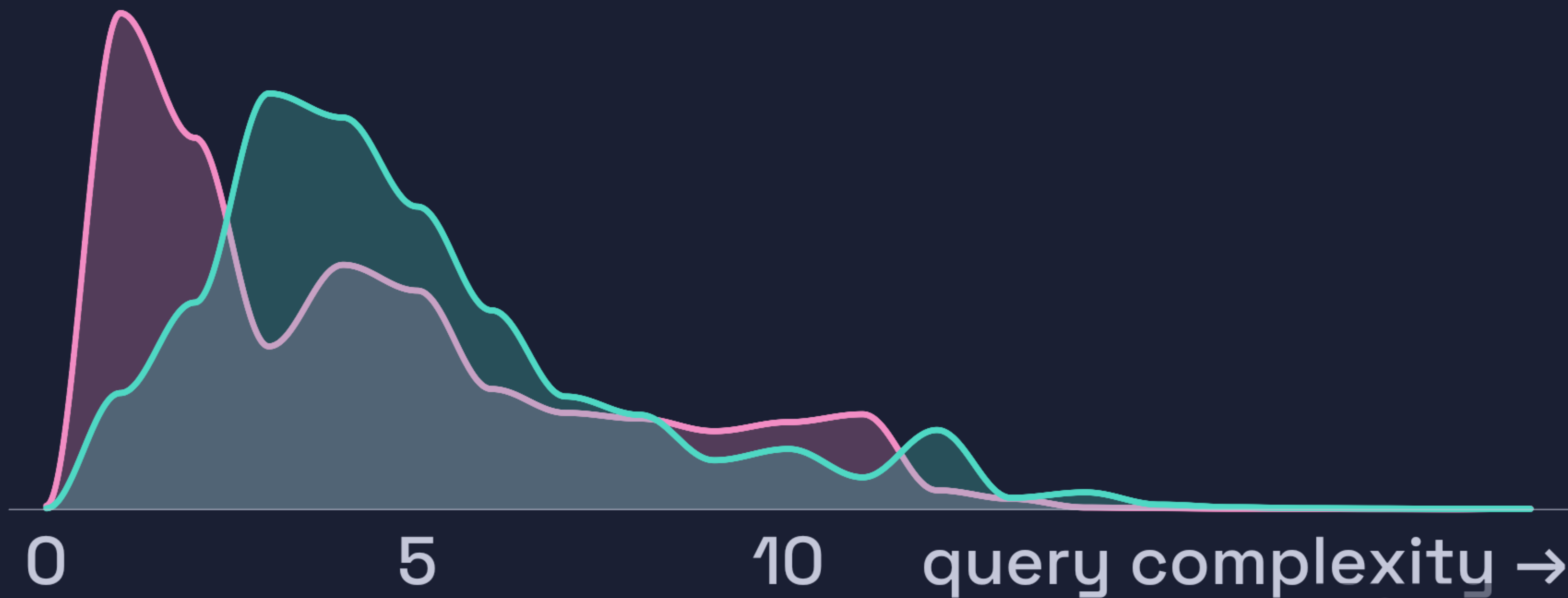
DESIRE PATHS



TWO KINDS OF TRAFFIC



■ API keys ■ OAuth



AGENTS LOVE NAMING THINGS



a person:

COUNT

an agent:

total = COUNT

errors = COUNT where ...

err_pct = $\$errors / \$total * 100$

vol = $\$total$

err_count = $\$errors$

AGENTS WANT CONSISTENCY

run_query

from: "2026-09-03T16:00:00Z"

get_trace

from: "2026-09-03T16:00:00Z"

create_marker

from: "2026-09-03T16:00:00Z"

CHANGE THE SCHEMA



SLOW LEARNINGS

EVALS ARE HARD



	Claude	GPT	Gemini	other
Claude Code				
Codex				
Cursor				
VS Code				
Custom agents				

Traffic tells us where to start

MODELS CAUGHT UP



count unique users

89% ● — ● 100%

latency root cause

72% ● — ● 100%

SLO investigation

74% ● — ● 93%

user activity

38% ● — ● 77%

calculated fields

53% ● — ● 93%

list environments

63% ● — ● 86%

All six: 65% → 92% passed

● spring ● now

IMPROVING RESPONSES



Overview BubbleUp Correlations Traces Explore Data

mcp.client.name		COUNT
mcp	...	22,335
claude-code	...	6,350
codex-mcp-client	...	1,488
Anthropic/ClaudeAI	...	1,176
mcporter	...	795
notion-mcp-client	...	768
mcp-remote-fallback-test	...	689
MintMCP Client	...	567
Cursor	...	509
nominal-honeycomb-client	...	420
claude-code (via mcp-remote 0.14.3)	...	319

WHAT WE TRIED



NO

JSON

MEH

CSV

MEH

SVG

NO

Screenshots

NO

Text summaries

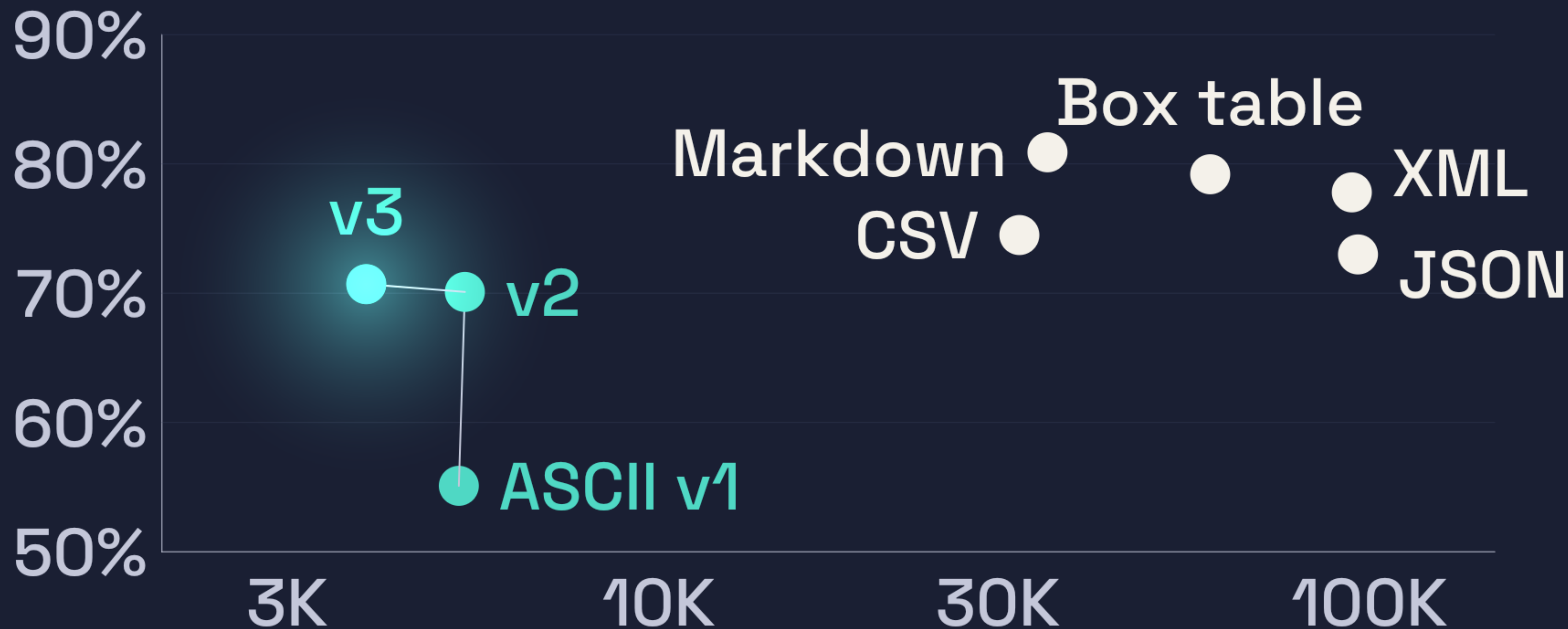
WHAT WE LANDED ON



Results

COUNT	P50(duration_ms)	name
---	---	---
2224775	272.75	mcp.tool.query
77854	1137.55	mcp.tool.trace
70037	1172.32	mcp.tool.get_span_details
67064	577.72	mcp.tool.list_spans

1/10TH THE TOKENS



prompt tokens →

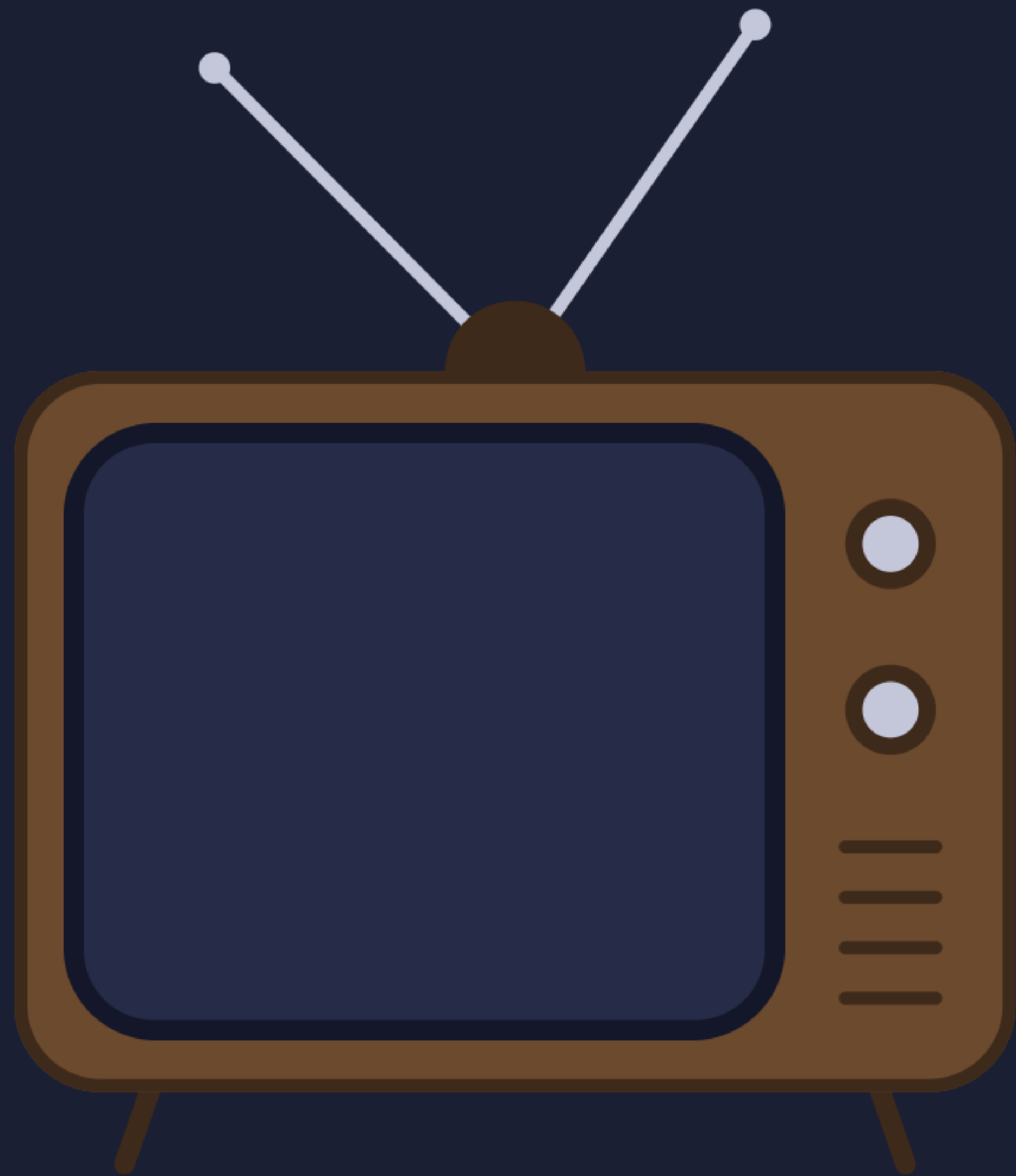
SHAPE BEATS PRECISION

In an incident, you
want what's different.

TAKEAWAYS



- Trace it with OpenTelemetry
- Follow the desire paths
- Context still matters



How do you stand out in the crowd?

Watch your server.

Agents won't tell you what's broken. Your telemetry will.

THANKS!



aparker.io