01 / INTRO

How do you *know* if a skill covers what you need?

AI skills are curated knowledge bases that teach an assistant how to work in a specific domain — a codebase, a house style guide, a field of research, a body of legal precedent. Skill Cards audits them against the reference material they claim to cover.

Sruly Rosenblat · AI Disclosures Project

A sustainable marketplace needs a way to tell the *good* from the *bad*.

Skill quality differences are real and measurable.

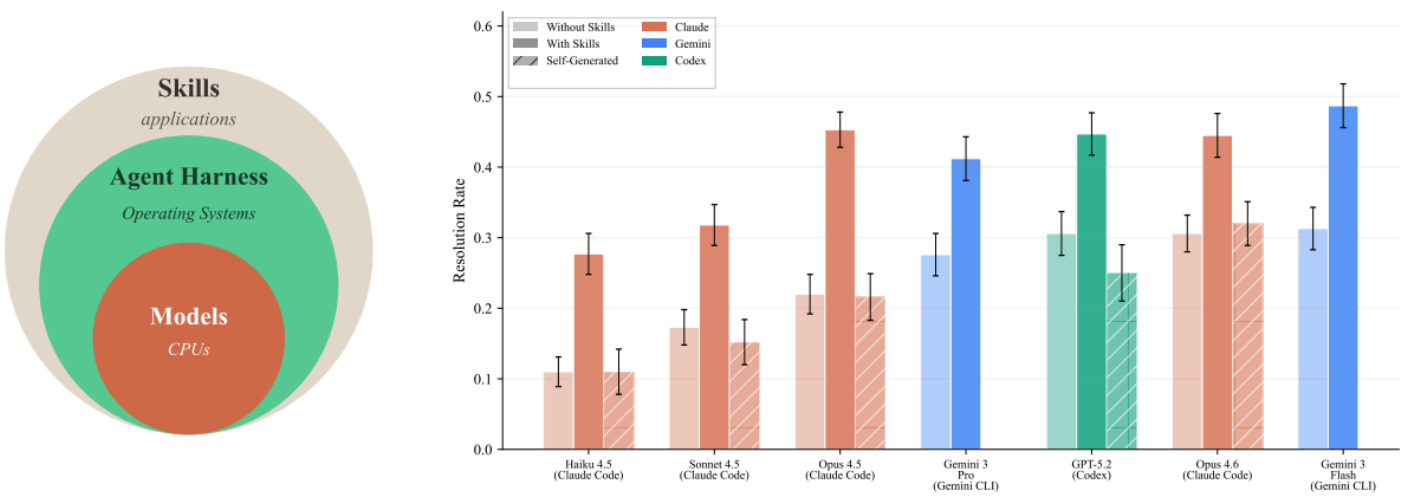


Figure 1: Agent architecture stack and resolution rates across 7 agent-model configurations on 84 tasks. Curated Skills (beige) improve performance by +16.2pp on average; self-generated Skills (amber) provide negligible or negative benefit.

In the study shown above, self-generated skills performed significantly worse than human-written ones. Browsing a marketplace doesn't show you any of that. arxiv.org/pdf/2602.12670

03 / THE PROBLEM

Skills on the same subject aren't *interchangeable*, but they are hard to *distinguish*.

SKILL.MD

fastapi-expert

```
---
name: fastapi-expert
description: Use
when building async
Python APIs with FastAPI
and Pydantic V2.
license: MIT
---

# FastAPI Expert

Deep expertise in async
Python, Pydantic V2, and
production-grade APIs.
```

5,453 words

?

?

?

DOCS/FASTAPI/ · 166,616 WORDS · 37 TOPICS

Query Parameters	OAuth & JWT	Async Programming
Response Handling	Dependency Injection	Streaming / SSE
Lifespan Events	Proxy Headers	Security Config
OpenAPI Callbacks	Request Body	Error Handling
SQLModel	Pydantic v2	Deployment
Settings / Env	Testing	Response Models
Forms & Files	*Custom Responses*	*+ 16 more topics*

03 / THE PROBLEM

Skills on the same subject aren't *interchangeable*, but they are hard to *distinguish*.

SKILL.MD

fastapi-expert

```
---
name: fastapi-expert
description: Use
when building async
Python APIs with FastAPI
and Pydantic V2.
license: MIT
---

# FastAPI Expert

Deep expertise in async
Python, Pydantic V2, and
production-grade APIs.
```

5,453 words

?

?

?

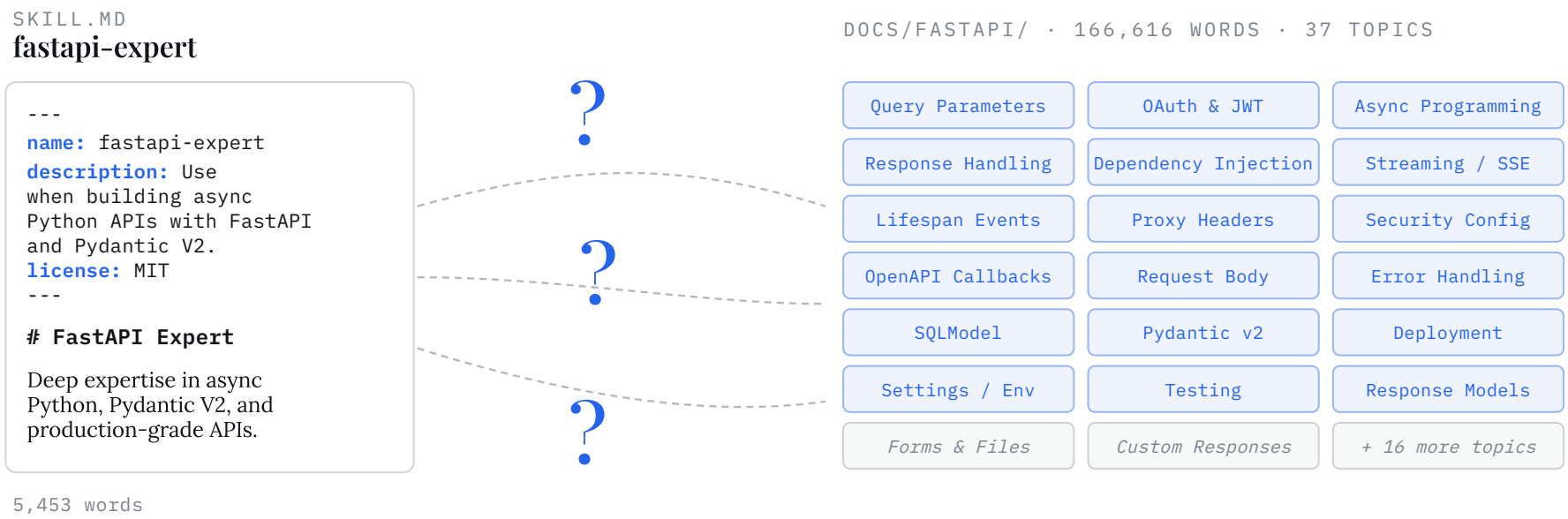
DOCS/FASTAPI/ · 166,616 WORDS · 37 TOPICS

Query Parameters	OAuth & JWT	Async Programming
Response Handling	Dependency Injection	Streaming / SSE
Lifespan Events	Proxy Headers	Security Config
OpenAPI Callbacks	Request Body	Error Handling
SQLModel	Pydantic v2	Deployment
Settings / Env	Testing	Response Models
Forms & Files	*Custom Responses*	*+ 16 more topics*

Does the left actually **cover** *the right?*

03 / THE PROBLEM

Skills on the same subject aren't *interchangeable*, but they are hard to *distinguish*.



Does the left actually **cover** *the right?*

There are dozens of skills for each topic. Nobody can check them all.

04 / BENCHMARKING TEXT

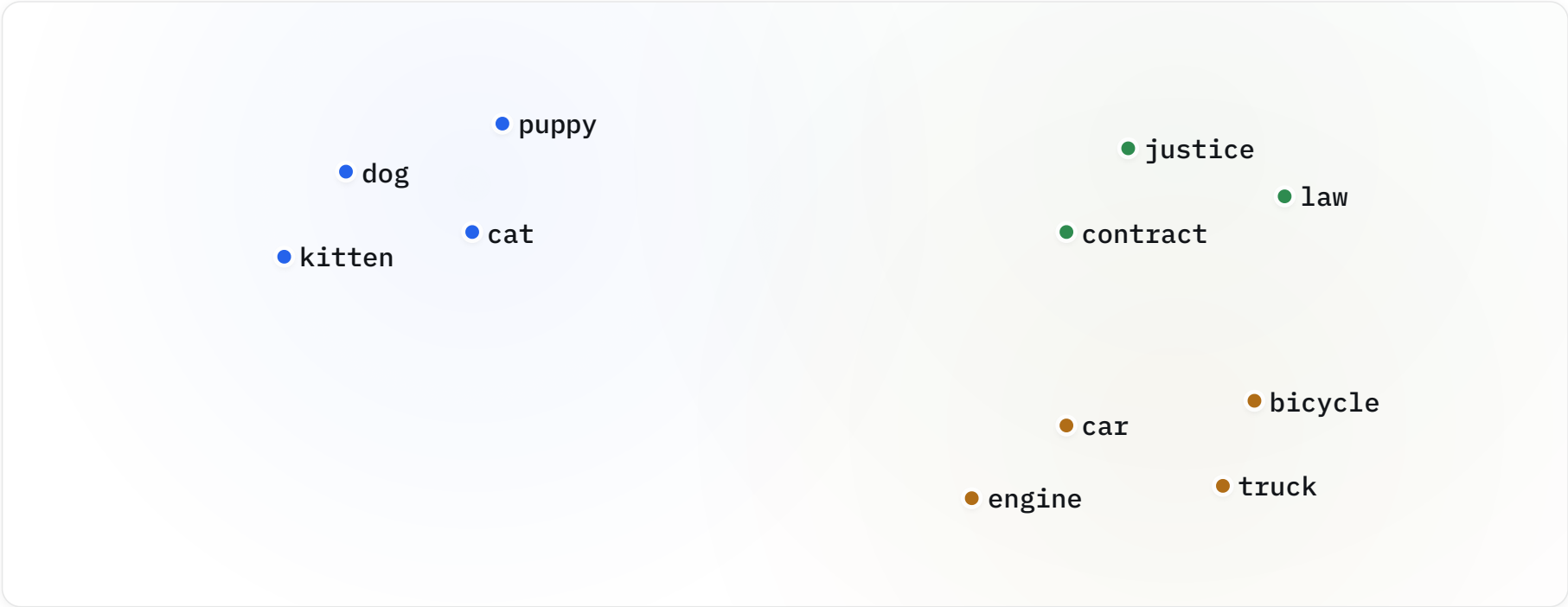
See the differences clearly without reading every skill.

FastAPI docs · 166,616 words · 37 topics, measured against three skills from the Claude skill marketplace.

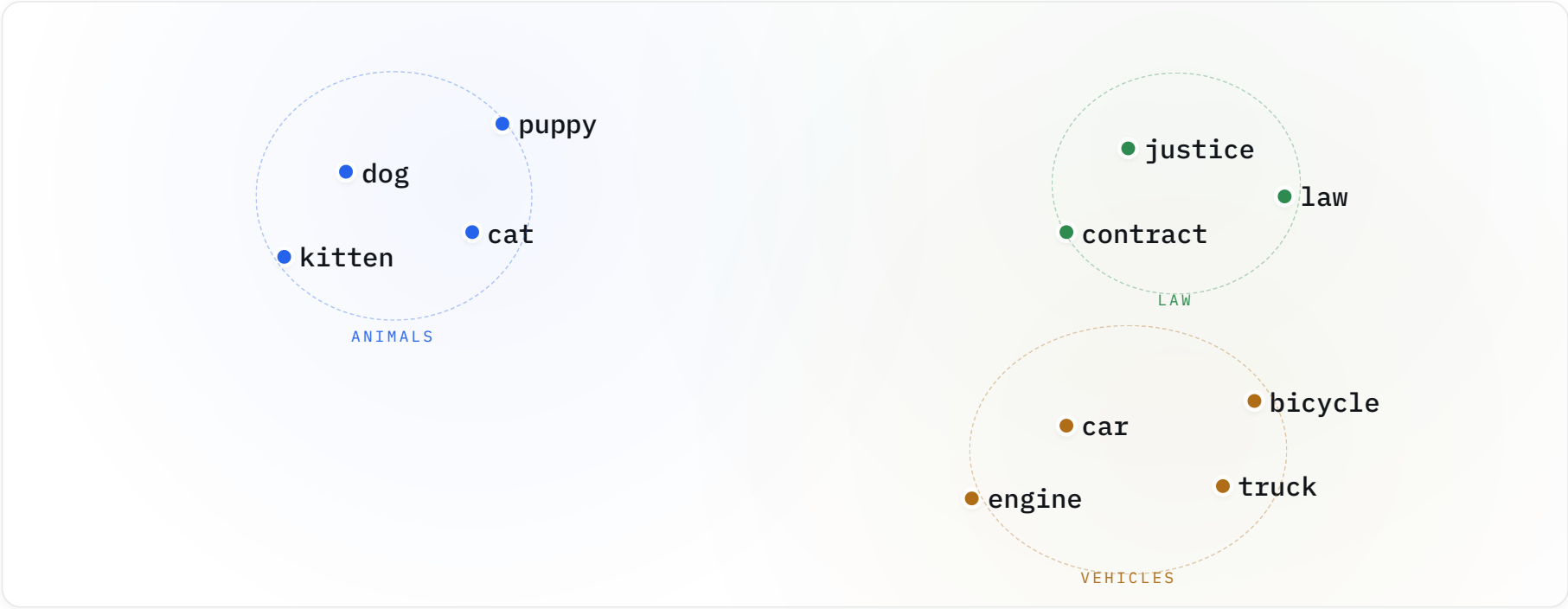
focus tier ● Heavy ● Moderate ● Light ● Trace ▢ Top focus per topic

TOPIC CLUSTER	FASTAPI-EXPERT 5,453 words	FASTAPI-PRO 901 words	FASTAPI-TEMPLATES 1,402 words
SQLModel Database Integration	35.7%	0%	39.1%
OAuth2 and JWT Authentication	23.6%	0%	28.3%
FastAPI Introduction and Basics	6.0%	27.8%	4.3%
Deployment and HTTPS Concepts	2.2%	16.7%	0%
FastAPI Security and Configuration	7.7%	0%	15.2%
Contributing and Community Support	0%	16.7%	0%
FastAPI Response Handling	0%	0%	0%

Language models can lay out text in a space where similar ideas end up *next to each other*.

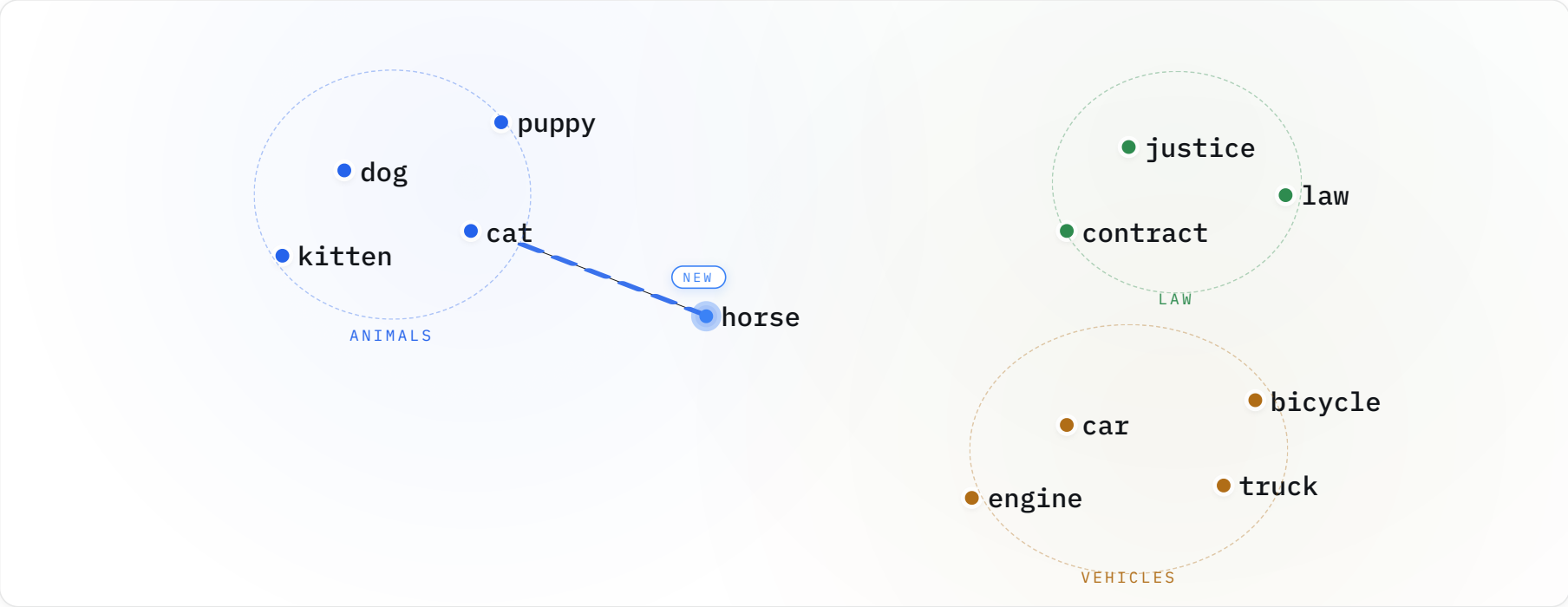


Language models can lay out text in a space where similar ideas end up *next to each other*.



Dots that sit close together mean similar things.

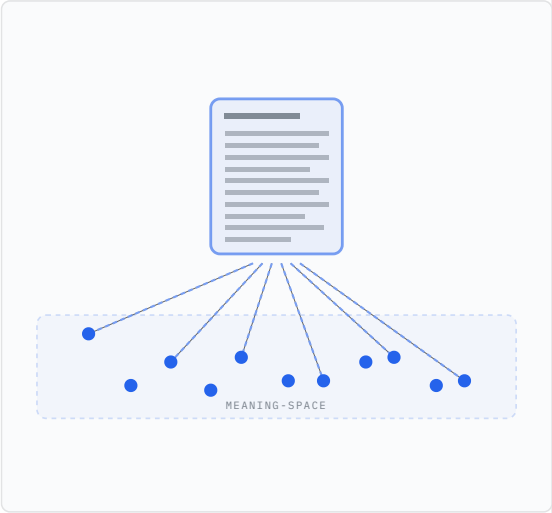
Language models can lay out text in a space where similar ideas end up *next to each other*.



Dots that sit close together mean similar things.

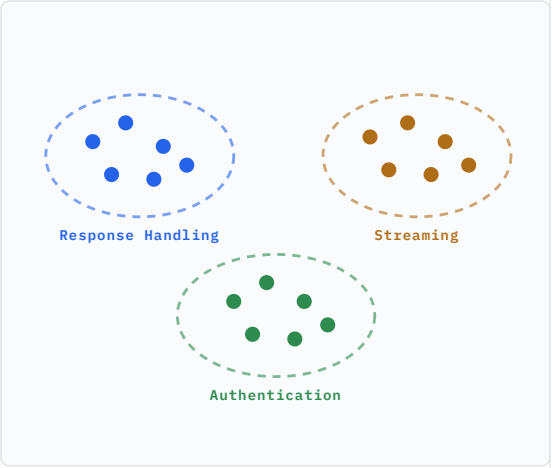
How to compare skills to reference material.

1
Break the docs into pieces.



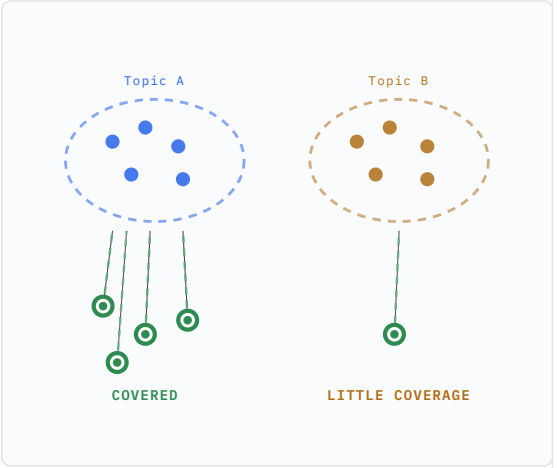
Each piece becomes a point in the meaning-space.

2
Cluster, then label.



Nearby points form topics. An LLM reads each cluster and gives it a name.

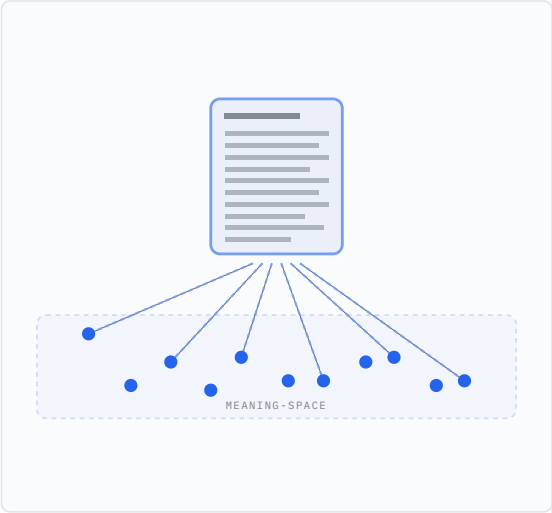
3
Measure the skill against each topic.



If a skill chunk lands inside a topic, it's **covered**. If nothing lands, it's a **gap**.

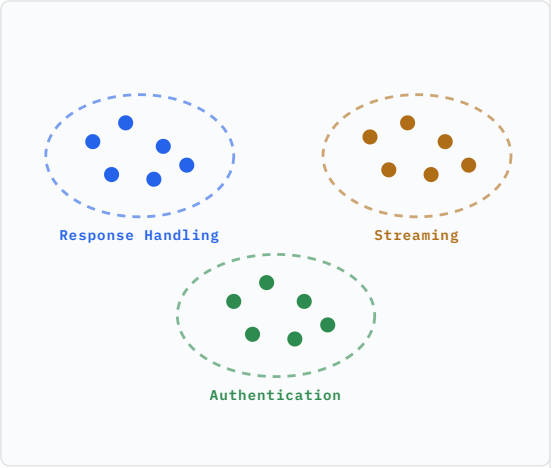
How to compare skills to reference material.

1
Break the docs into pieces.



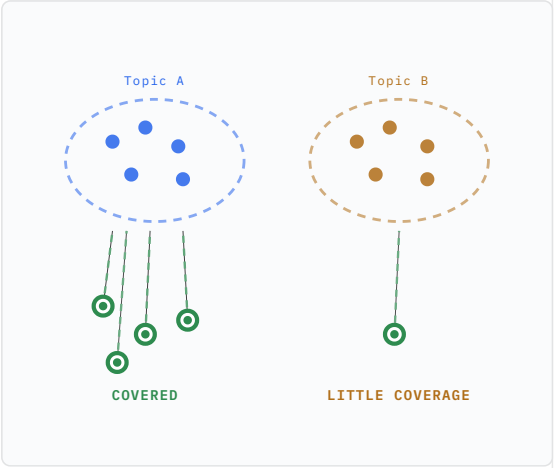
Each piece becomes a point in the meaning-space.

2
Cluster, then label.



Nearby points form topics. An LLM reads each cluster and gives it a name.

3
Measure the skill against each topic.



If a skill chunk lands inside a topic, it's **covered**. If nothing lands, it's a **gap**.

The skill goes through the same mapping process. Coverage is just geometry after that.

07 / THE FINE PRINT

A coverage score is not a quality score.

MISSES Whether the skill is **correct**.

MISSES Whether the skill is **well-written** or usable.

MISSES Whether a topic is **important** or niche.

FINDS Topics the skill **never addresses**.

FINDS Two skills that cover the **same topics**.

FINDS What topics the skill **misses**.

FINDS Where a skill puts its **focus**.

FINDS How a skill **compares** to others.

07 / THE FINE PRINT

A coverage score is not a quality score.

MISSES	Whether the skill is **correct**.
MISSES	Whether the skill is **well-written** or usable.
MISSES	Whether a topic is **important** or niche.
FINDS	Topics the skill **never addresses**.
FINDS	Two skills that cover the **same topics**.
FINDS	What topics the skill **misses**.
FINDS	Where a skill puts its **focus**.
FINDS	How a skill **compares** to others.

*Finds what's **missing**. Can't judge what's **there**.*

08 / BEYOND CODE

The same system works for anything made of text.

Code docs were one example. The math doesn't know what it's measuring — two piles of text, one mapping, one coverage score.

`a technical textbook`	`a coding skill`	*Which chapters' techniques does the skill actually know?*
`clinical guidelines`	`a medical skill`	*Which conditions does the skill leave out?*
`case law`	`a legal skill`	*Which precedents does the skill reference when drafting?*
`a research corpus`	`a literature-review skill`	*Which subfields did the skill's review skip?*
`a brand style guide`	`a copywriting skill`	*Does the skill stay on-brand across every voice it writes in?*
`your reference material`	`your skill`	*The same audit runs against whatever text you're writing against.*

Available tools

Read-only MCP server at `skill-cards-mcp.fly.dev/mcp/`.

`list_doc_sets`	Every doc corpus the server knows.	`list_skills`	All skills, or skills for one corpus.
`get_skill`	Shape of one skill (per-file chunk counts).	`get_doc_topics`	Topic clusters in one corpus.
`get_skill_coverage`	Per-topic coverage for one skill.	`compare_skills`	Side-by-side coverage for N skills.
`find_coverage_gaps`	Weakest topics for one skill.	`get_topic_detail`	Drill into one cluster.
`search_skill_chunks`	Lexical search over a skill.	`search_doc_chunks`	Lexical search over a corpus.

any agent → `skill-cards-mcp.fly.dev/mcp/` → coverage JSON

Thanks.

SCAN FOR THE LIVE DASHBOARD

skill-cards-mcp.fly.dev

MCP endpoint at `skill-cards-mcp.fly.dev/mcp/`

LIVE DASHBOARD + MCP

skill-cards-mcp.fly.dev

SOURCE CODE

github.com/SrulyRosenblat/skill-cards

CONTACT

sruly@aidisclosures.org

Sruly Rosenblat · AI Disclosures Project