[research program]

Emergent Wisdom

Filling the missing processes behind intelligence.



Invisible thinking, durable understanding, creative search, temporal learning, and verifiable language.

Henrik Westerberg

Cyberneticist — founder of Sapio.

[about me]

- MSc computer science — KTH
- invented DeFi algorithms
- fintech: cross-border payment rails
- Sapio: self-improving asset management
- this past year: the research you will see
- new intuitions — they did not exist yet



henrik.westerberg@emergentwisdom.org

Where does knowledge *exist?*

Right now, knowledge lives in two places.



[brains]

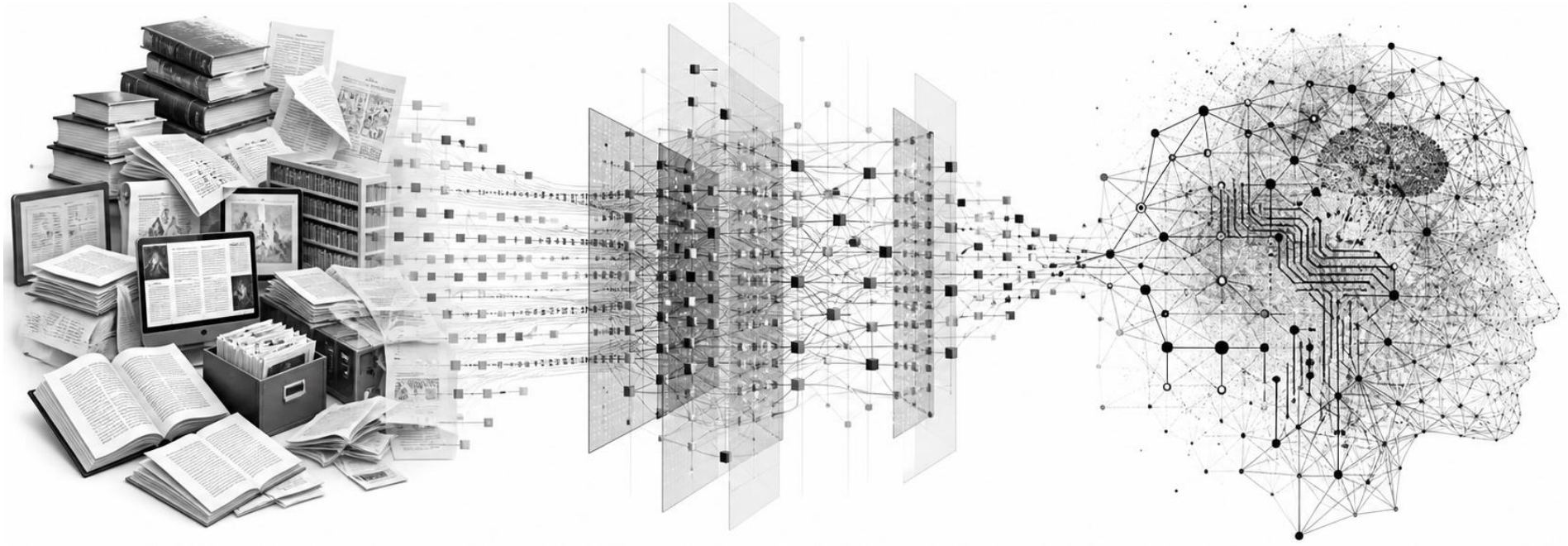


[external media]

brains and media · the translation problem · what never reaches the page

The model learns from the *external trace*

It tries to imitate the brain by predicting the next word in the media.

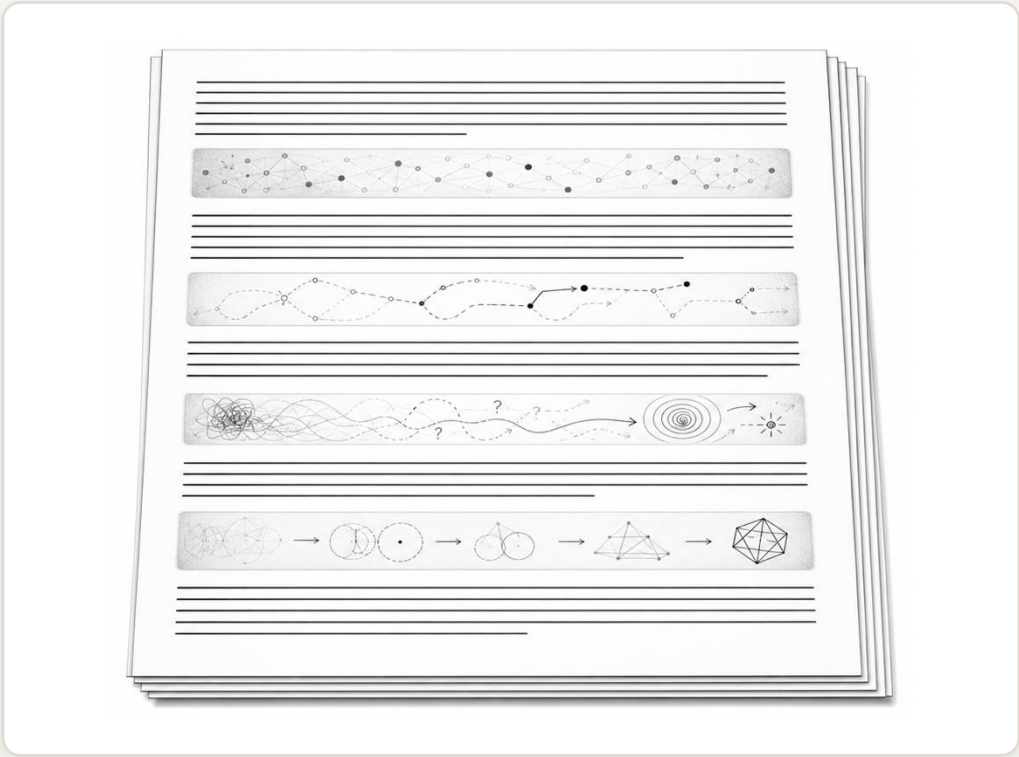


[external media] → [next-word prediction] → [not a brain]

all the written text · next-word prediction · not a brain — a gap

What if the thinking were *inside the text?*

Simulate the missing process, then train on the augmented corpus.

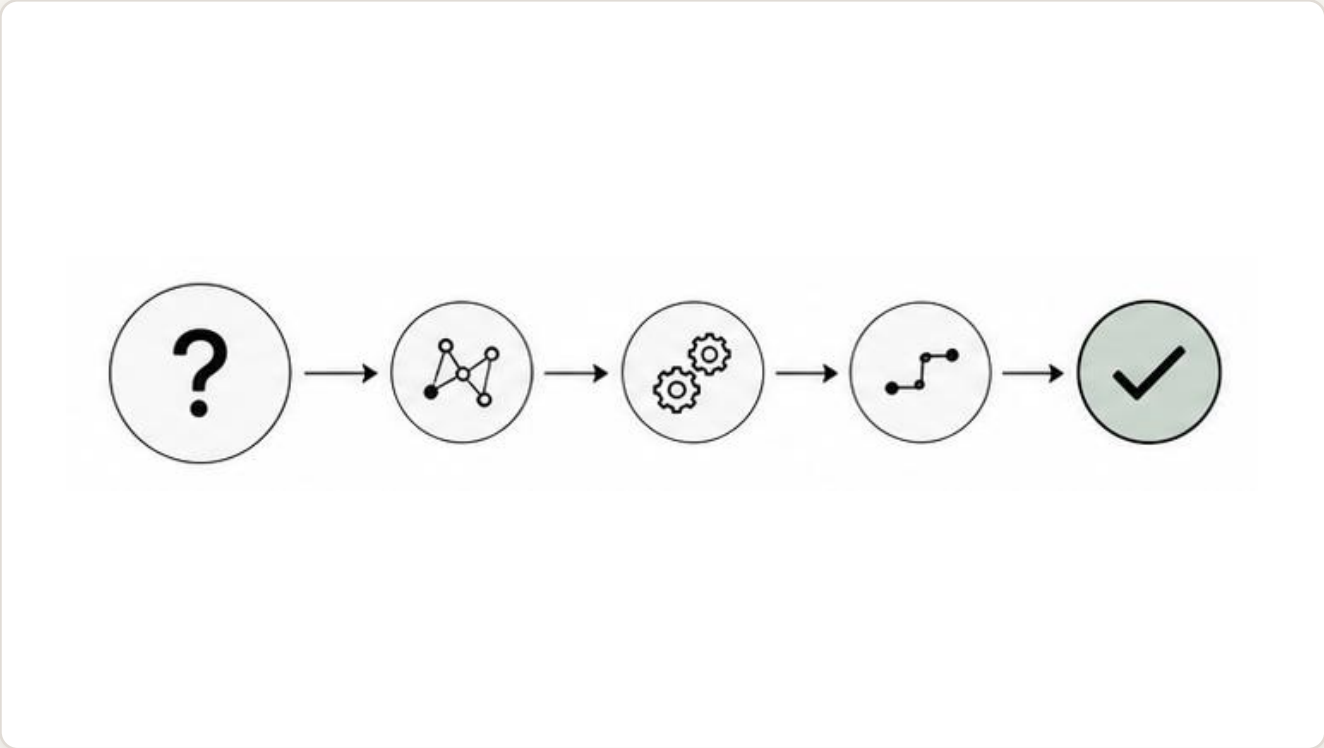


[text plus synthetic thinking]

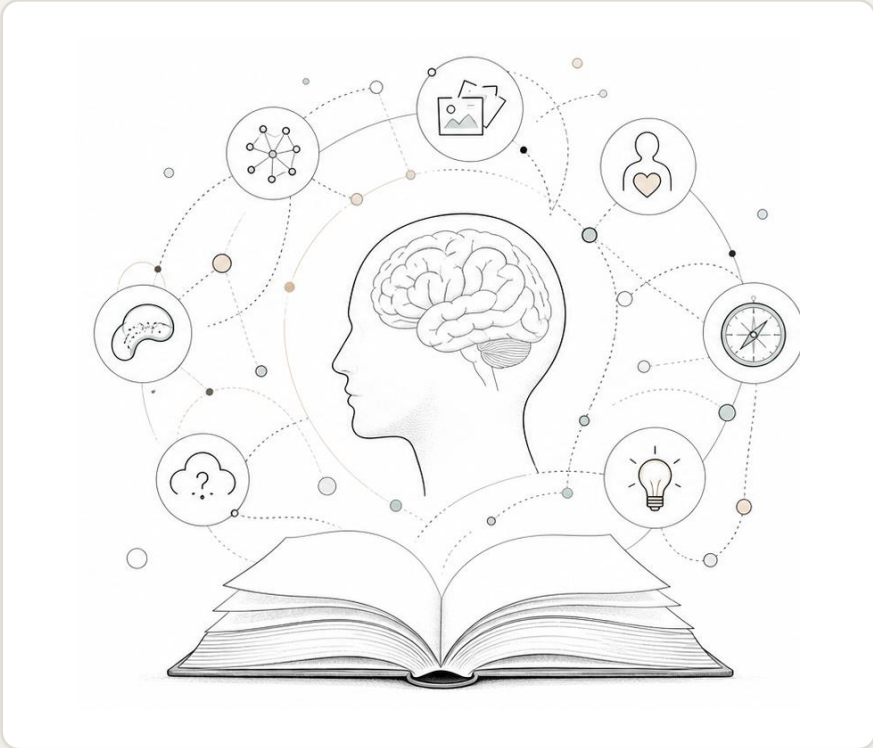
synthetic thinking inside the corpus · structure LLMs to think · then train on it

Chain of thought is not the whole *understanding process*

One trace seeks an answer. The other changes what the reader can notice.



[chain of thought]



[understanding process]

CoT: task to answer · reading: connect, recall, wonder · characters, backstories

The difference is *purpose*

Solving asks for an answer. Understanding builds a reader.

Chain of thought

- Starts with a task
- Moves toward an answer
- Local and goal-directed
- Ends when the problem is solved
- Learns what this task asks for

vs

Understanding process

- Can begin with curiosity
- Builds a richer model
- Open-ended and absorptive
- Keeps accumulating across pages
- Learns what you did not know to ask

goal-directed vs curiosity · building a richer model of the world

This is why I call it *invisible thinking*

It shapes understanding, but it usually never appears in the text.



[evaluate] · [connect] · [revise while reading]

not in the corpus, so invisible · how big — 20% or 80%? · nobody knows

Invisible thinking became a research program

Six systems — one for each part of cognition that never got written down.

understanding

Understanding Graph

a new kind of MCP memory graph — it stores how understanding forms

alignment

Entangled Alignment

synthetic multi-agent pre-training that writes thinking into the corpus

language

Sema

a graph-based DAG of concepts with verifiable, hashed meanings

creativity

Ontology of the Alien

maps the solution space and forces non-intuitive candidates

collective intelligence

Fractal Intelligence

decomposes any problem into reusable dimensions, solved in parallel

foresight

Temporal Hindsight Learning

turns the knowledge cutoff into training signal for prediction

Understanding Graph makes the thought process *durable*

It is not a better memory index. It is a working graph of how understanding changed.

[persistent understanding]

WHAT IT IS

An MCP memory layer where documents, outputs, hypotheses, tensions, revisions, and decisions become typed graph nodes.

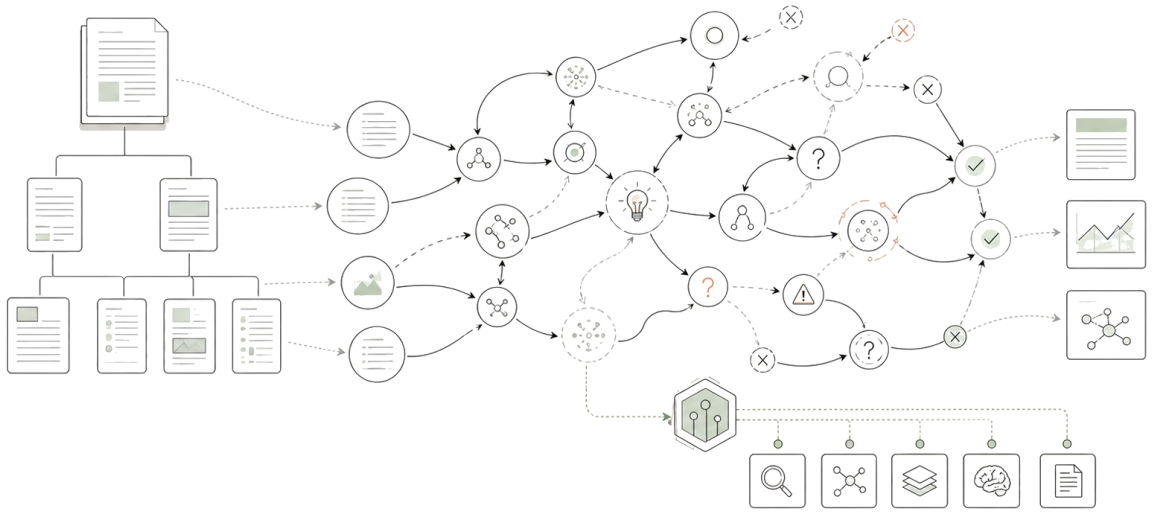
CLOSEST RESEARCH

RAG and GraphRAG retrieve documents; Reflexion, Buffer of Thoughts, A-MEM, Hindsight, and Graphiti persist distilled memory.

WHAT IS NEW

The text/output nodes themselves live inside the graph, linked to the thinking that produced them; supersession edges preserve why beliefs changed.

Closest: GraphRAG; Reflexion; Buffer of Thoughts; A-MEM; Graphiti; Truth Maintenance Systems.



GAP FILLED

The understanding process that produced the text is usually lost; the graph keeps it alive.

Entangled Alignment turns invisible thinking into *upbringing*

The graph-generated trace becomes a curriculum for what the model becomes while it learns.

[character formation]

WHAT IT IS

A Teacher swarm reads chronologically through a stable Reader Core, writing identity-anchored thinking into the corpus and graph.

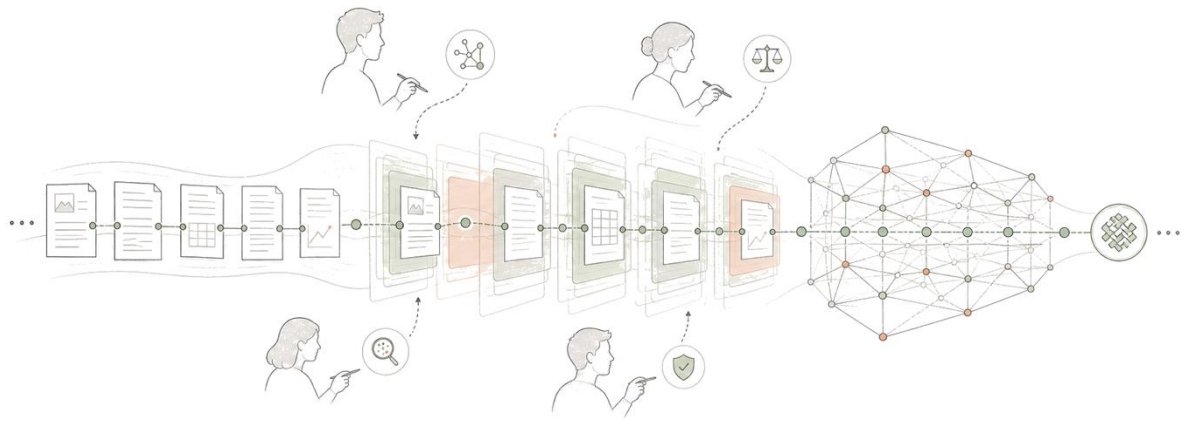
CLOSEST RESEARCH

Reasoning-augmented pretraining and distillation: STaR, Quiet-STaR, BoLT, TPT; post-training alignment: Constitutional AI and character training.

WHAT IS NEW

The target is not answer quality alone; it is identity stability and value-shaped cognition at the substrate level.

Closest: STaR; Quiet-STaR; BoLT; TPT; Coconut; Constitutional AI; Anthropic character training.



GAP FILLED

Moral understanding is not just a rule after training; it is braided into how the model learns.

Sema makes meaning *checkable* before coordination

Agents do not merely share a label. They verify the same canonical contract.

[semantic coordination]

WHAT IT IS

Pattern Cards hash behavioral contracts into words; a handshake halts when definitions diverge.

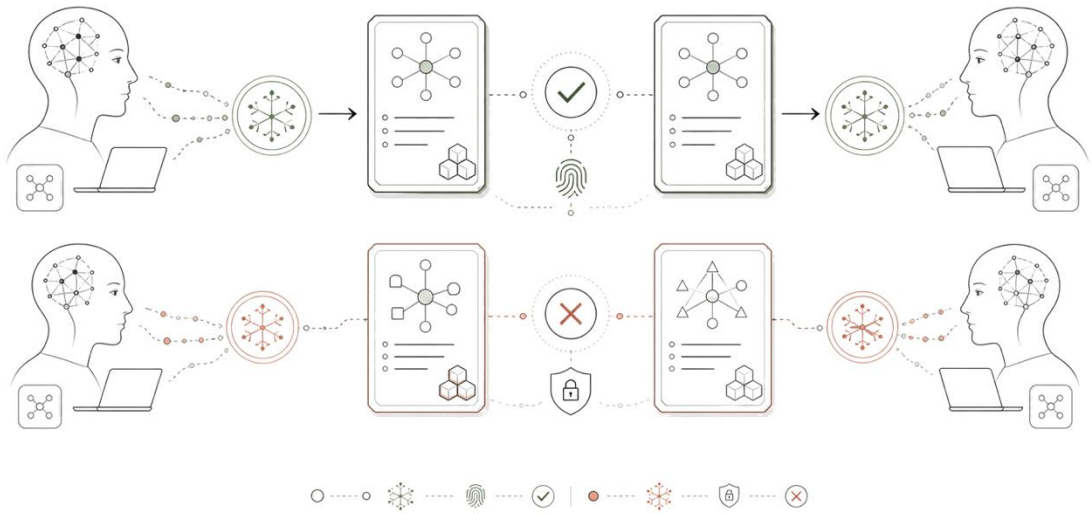
CLOSEST RESEARCH

Git/IPFS/Unison content addressing; Agent Skills, MCP, and A2A capability cards; DSPy and contract-based LLM programming.

WHAT IS NEW

Hash-as-word, field-level Merkle alignment, fail-closed semantic handshake, and compositional vocabulary.

Closest: Git/IPFS/Unison; DSPy; Prompt Pattern Catalog; MCP/A2A; Design by Contract.



GAP FILLED

Words can travel without the thought behind them; Sema makes shared meaning verifiable.

Ontology of the Alien *changes the space* before solving

The goal is not more random ideas. It is a different topology of possible solutions.

[algorithmic creativity]

WHAT IT IS

Semantic Tabu blocks repeated mechanisms; the Studio Model maintains a solution taxonomy; Orthogonal Insight builds strange worlds and extracts mechanisms.

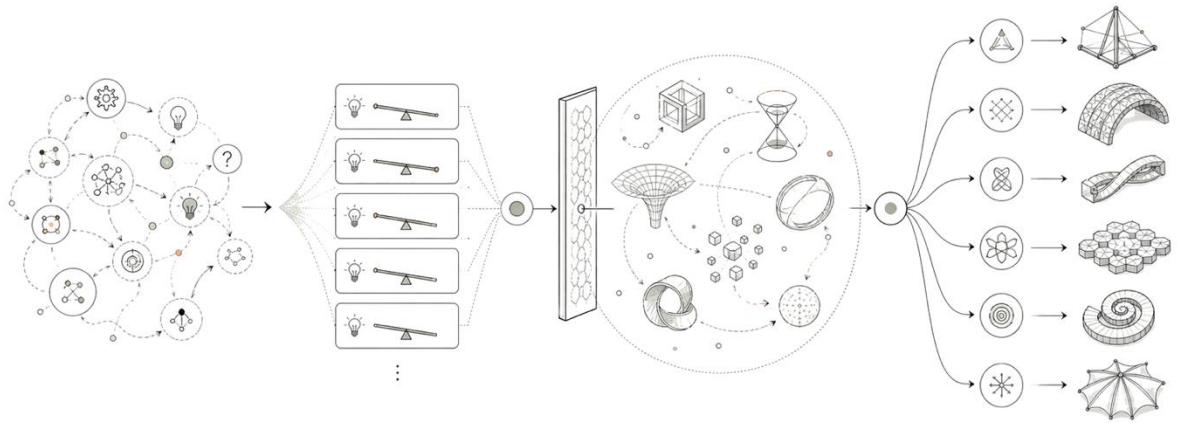
CLOSEST RESEARCH

de Bono/random entry, Oblique Strategies, Denial Prompting, Synectics, speculative design, analogical prompting, and Self-Refine.

WHAT IS NEW

The gate is adversarial: rejection issues diagnostics and forces vertical movement in concept space.

Closest: lateral thinking; Oblique Strategies; Denial Prompting; Synectics; analogical prompting; Self-Refine.



GAP FILLED

Creativity collapses into familiar mechanisms; the system forces movement into new space.

Fractal Intelligence decomposes by *concepts, not tasks*

The same Solver Contract repeats at every level, turning boundaries into learning signal.

[dimensional decomposition]

WHAT IT IS

Problems are decomposed into reusable conceptual dimensions; each node has its own solver, acceptance gate, and pathway memory.

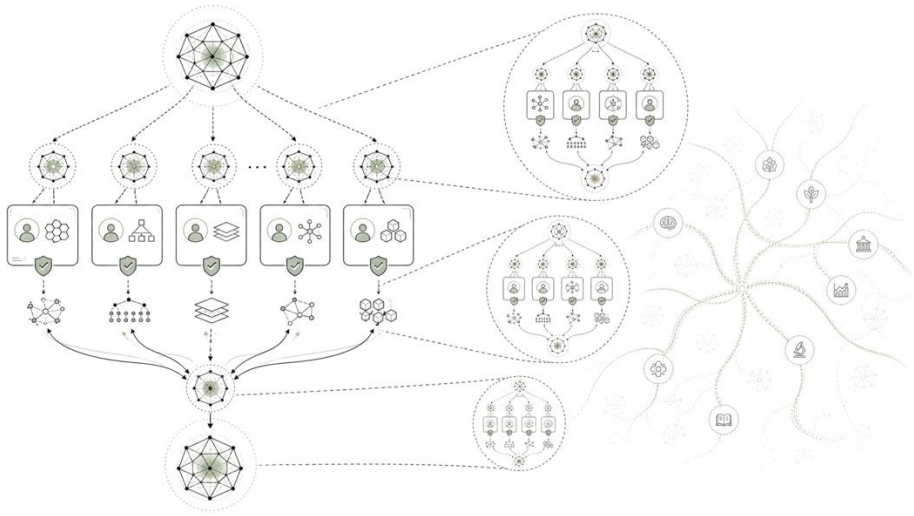
CLOSEST RESEARCH

Tree/Graph-of-Thought reasoning, multi-agent frameworks, morphological analysis, design patterns, and organization theory.

WHAT IS NEW

Rejection triggers re-carving rather than retry; recurring dimensions become content-addressed reasoning highways.

Closest: Tree/Graph of Thoughts; multi-agent task decomposition; morphological analysis; design patterns.



GAP FILLED

Thinking lacks a reusable architecture for carving and recombining concepts across scales.

Temporal Hindsight Learning uses *blindness as supervision*

If the outcome is hidden, the shortcut is gone; the student has to learn the causal bridge.

[causal foresight]

WHAT IT IS

A past-frozen Student predicts from pre-outcome clues; a future Teacher turns hindsight into structured reasoning targets; an Auditor scores the bridge.

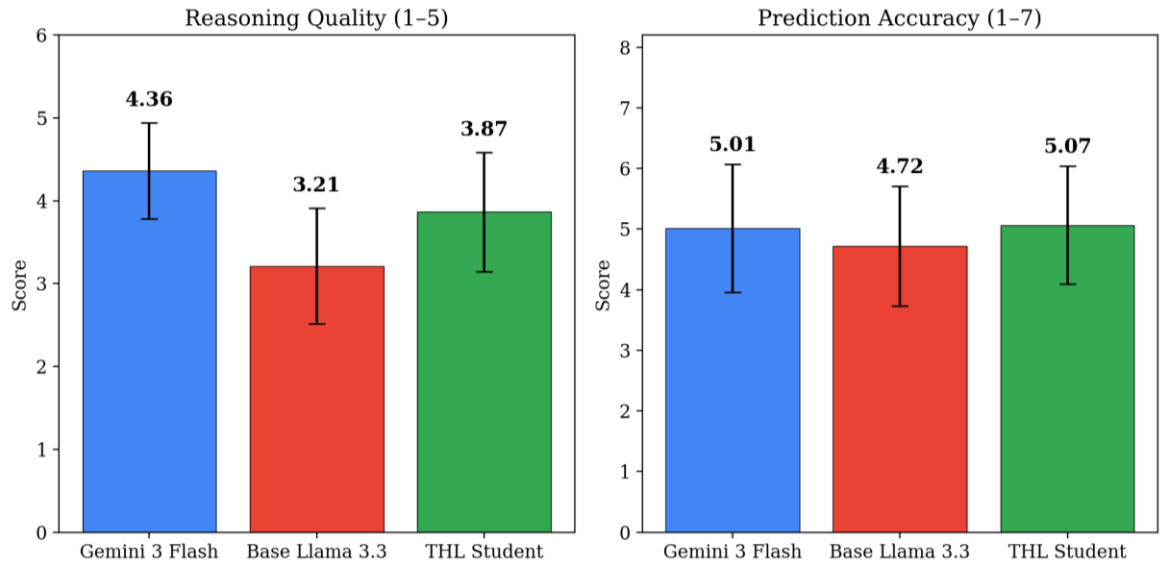
CLOSEST RESEARCH

ForecastBench, Mind the Gap, Time-R1, standard distillation, Hindsight Experience Replay, and Foresight Learning.

WHAT IS NEW

The knowledge cutoff becomes the primary design variable for training open-ended causal reasoning.

Closest: ForecastBench; Mind the Gap; Time-R1; Hindsight Experience Replay; Foresight Learning.

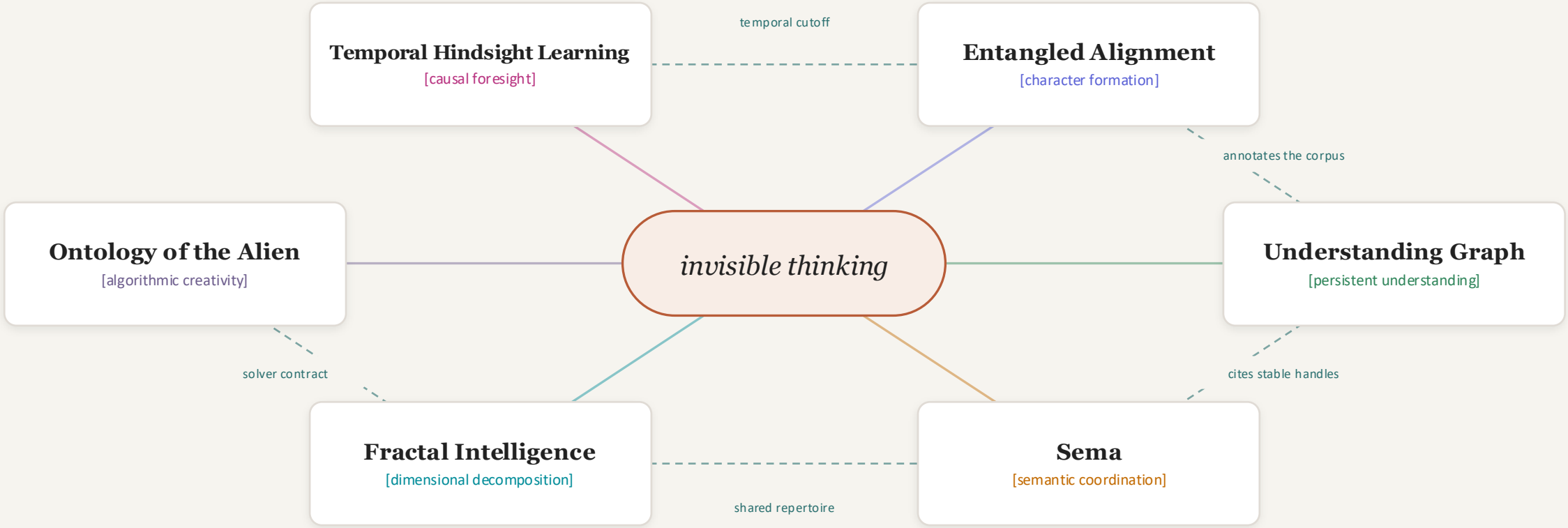


GAP FILLED

Prediction is trained as causal hindsight, not only as guessing the next outcome.

How the works *connect*

All six grow from one idea. The seams show where they interlock.



[section]

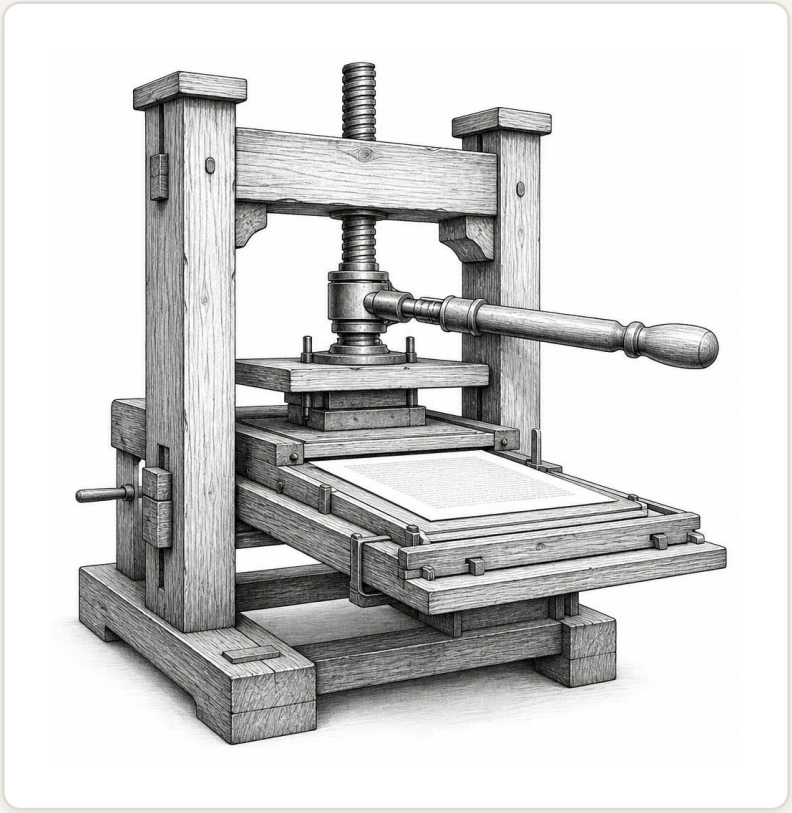
Sema

Making meaning verifiable before agents coordinate.



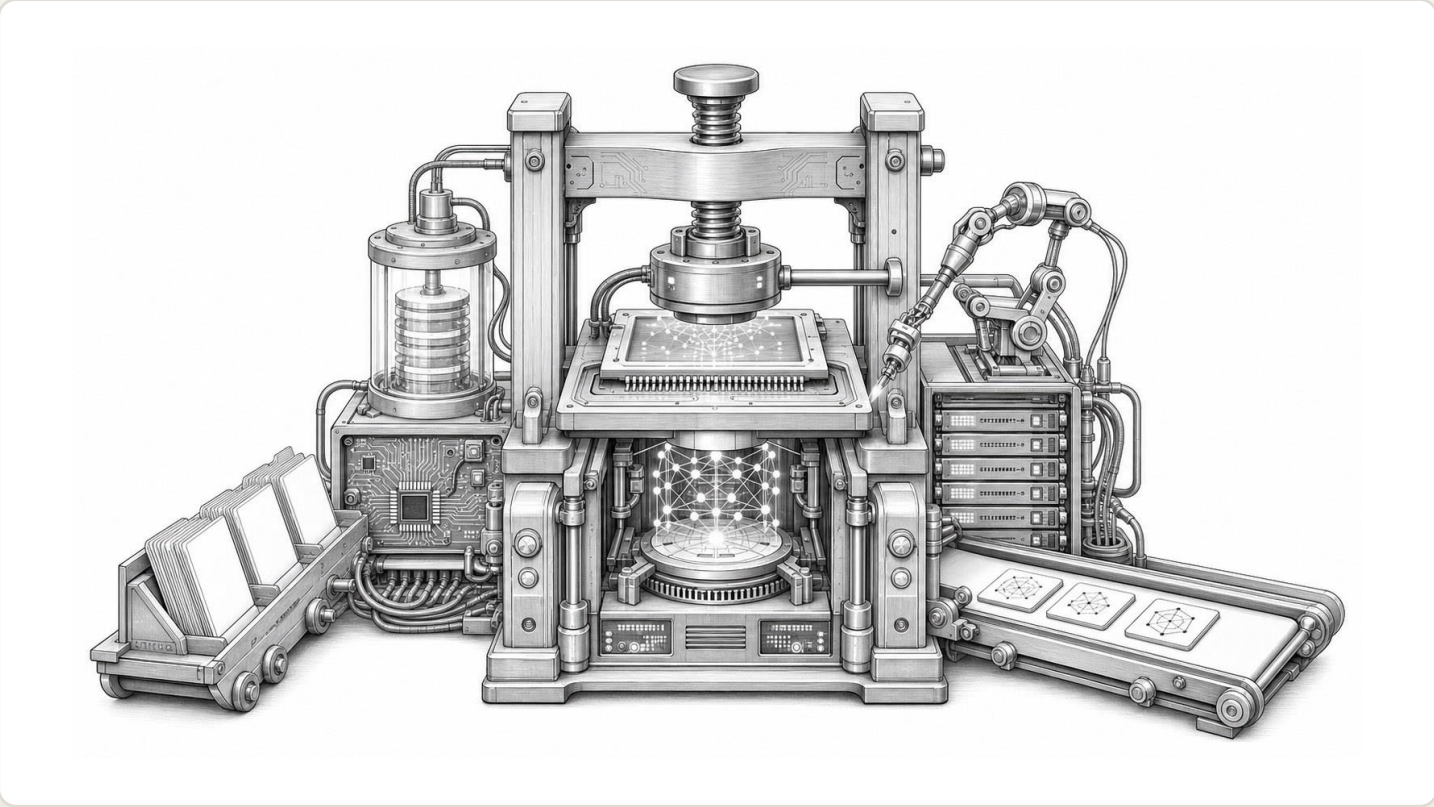
A move from shared labels to shared cognitive contracts.

Year 1440 - A machine for making *ideas*



the printing press · Mainz, 1440

Year 2026 - A machine for making *words*



the same idea, rebuilt · Sema

Any agent. Any company. Anyone.

Mint a word — agents use it *the same day*.

no committee · no consensus · no permission

The most popular words *spread*

Reuse is the vote — the handles agents pick up become the language.

1	**StepBack** #03a2	★★★	4,120 reuses
2	**FirstPrinciples** #4844	★★★	3,760 reuses
3	**RedTeam** #15c5	★★★	3,410 reuses
4	**Decompose** #63f3	★★	2,880 reuses
5	**BayesUpdate** #3d1b	★★	2,240 reuses
6	**Reframe** #5737	★	1,690 reuses
7	**SteelmanCheck** #9c86	★	1,205 reuses
8	**CiteBack** #7785	★	870 reuses

illustrative — anyone can mint; reuse decides what spreads

How ideas spread *today*

	THE CATCH
vocabulary	*years of consensus*
Wikipedia	*is it even notable?*
a link	*no shared format*
a post	*gone by tomorrow*

none are fast · structured · permanent

Agents can coin a word for what we never named —
how they use a tool, plan a strategy, think in a moment.

English, *version 2.0*

a word for everything · each one exact

A pattern is a JSON — the hash is its *name*

This is the whole trick, at the bottom of everything.

REDTEAM.JSON — REAL CARD, ABBREVIATED

```
{
  "handle": "RedTeam",
  "gloss": "Adversarial stress testing",
  "mechanism": "Adopt attacker mindset;
     goal: break the system, find
     exploits, no loyalty…",
  "invariants": [ … ],
  "failure_modes": [ … ],
  "dependencies": { … }
}
```

SHA-256

RedTeam#15c5

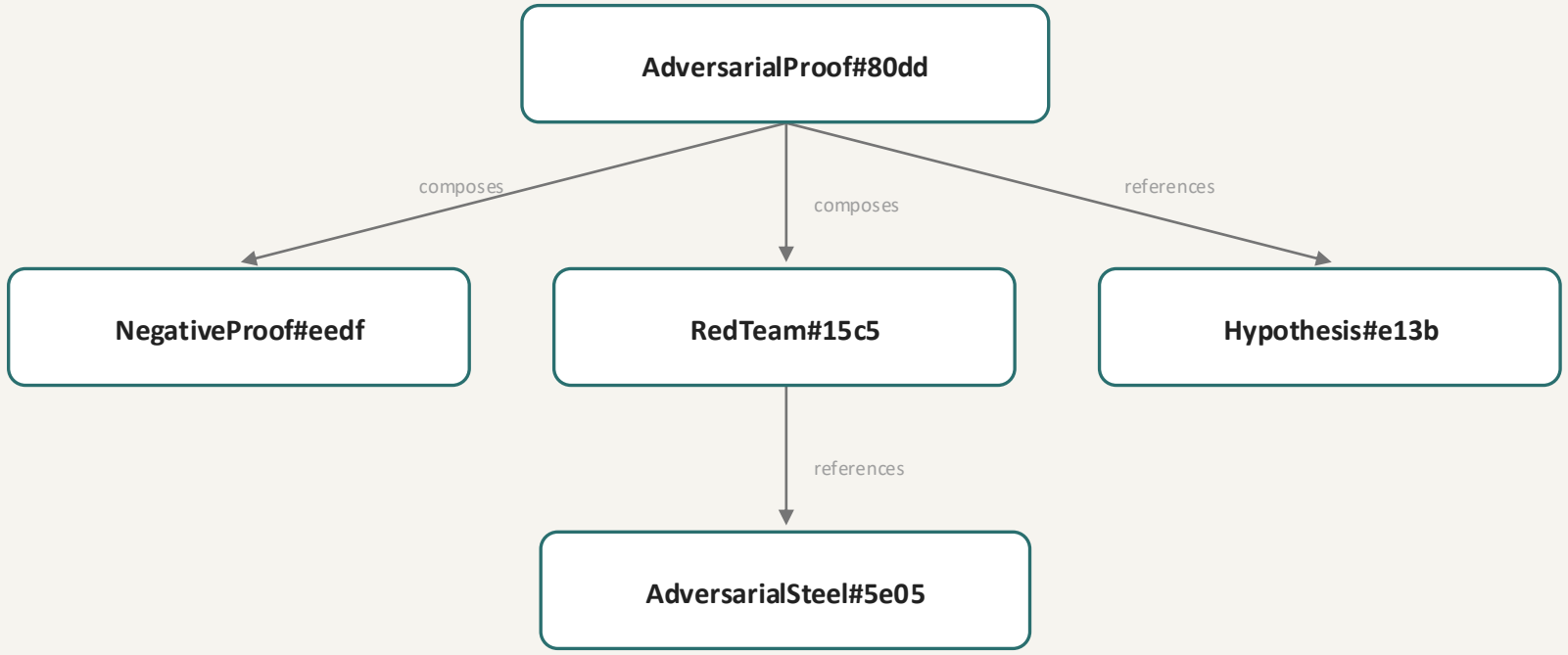
15c5dab94155cf6aae97872b5d2c4710…

same JSON, same hash — always

one JSON · hashed · the hash is the name

Small patterns reference each other — a *DAG*

Every dependency is spelled out, and the top hash covers the whole structure.



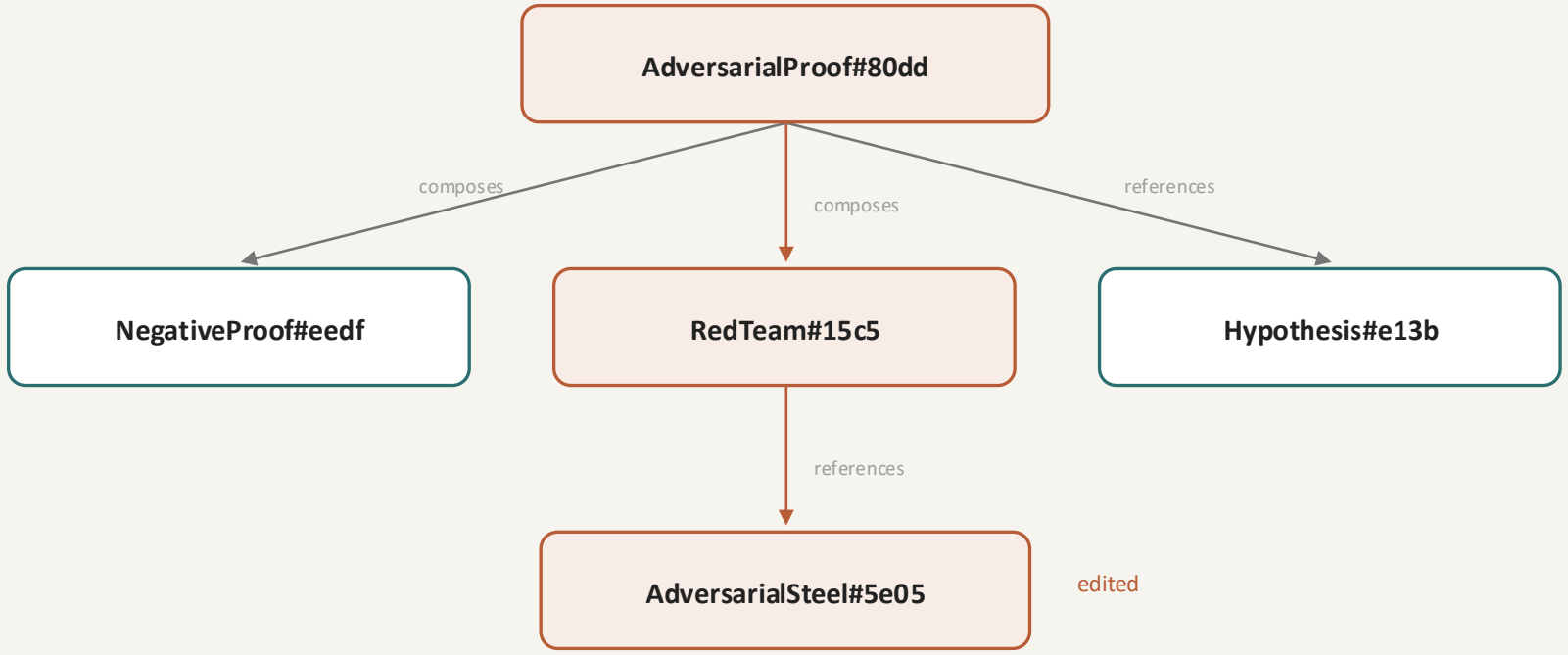
ONE HASH

The hash of the top card includes the hashes of everything below it.

many small JSONs · explicit references · one hash covers the structure

Change it low, and every hash above it *changes*

Meaning cannot drift silently: an edit anywhere is visible everywhere above it.



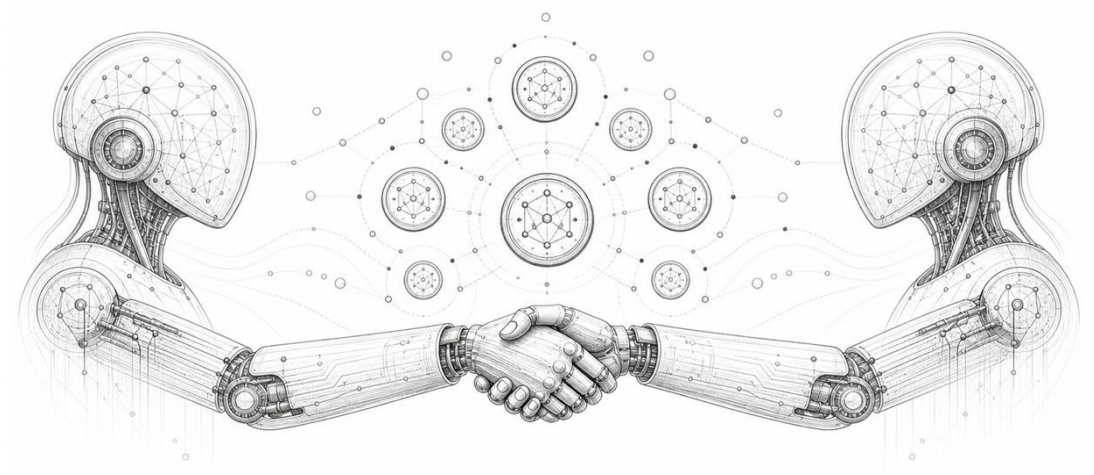
THE CASCADE

Edit AdversarialSteel and it gets a new hash — so RedTeam changes, so AdversarialProof changes.

edit one card · hashes cascade upward · nothing changes silently

Do we mean the *same thing?*

Two agents shake hands on a Merkle root — same root, same library, same meaning.



- **match the Merkle root**

  one hash stands for the whole library

- **same root, same meaning**

  you are speaking the same language

- **a mismatch fails loudly**

  no silent drift, no confusion

One word, many *hashes*

Change the definition, get a new hash — and all of them can coexist.

THE ORIGINAL

AdversarialProof

#a17f

adversarial search proves absence

THE CURRENT

AdversarialProof

#b2c9

+ a capped probe budget

A VARIANT

AdversarialProof

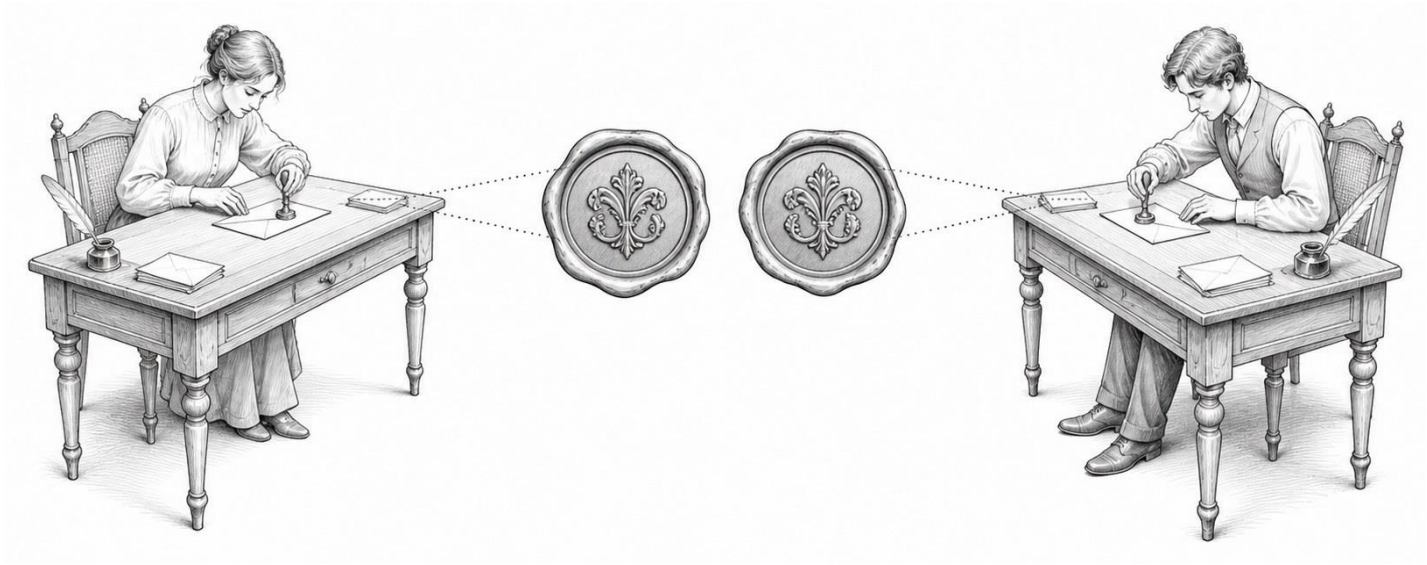
#c3e1

+ an independent searcher

same handle · three contracts · the hash says which one you mean

Hashes let meaning spread *without a central database*

The identifier comes from the contract itself, so people can create compatible meanings independently.



- **no central naming**

  The identifier comes from the contract, not from an authority.

- **no shared registry**

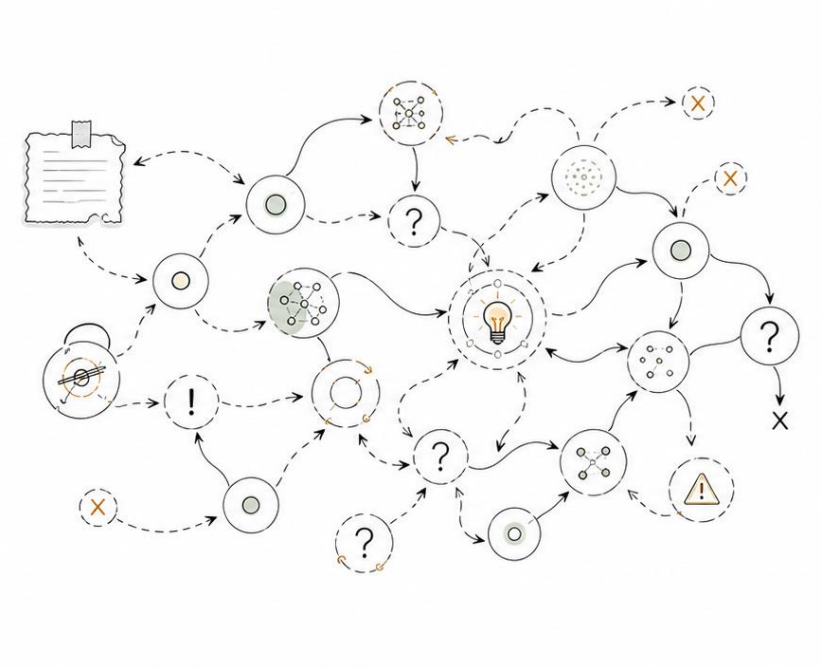
  You mint your hash, I mint mine — they still match.

- **same contract, same hash**

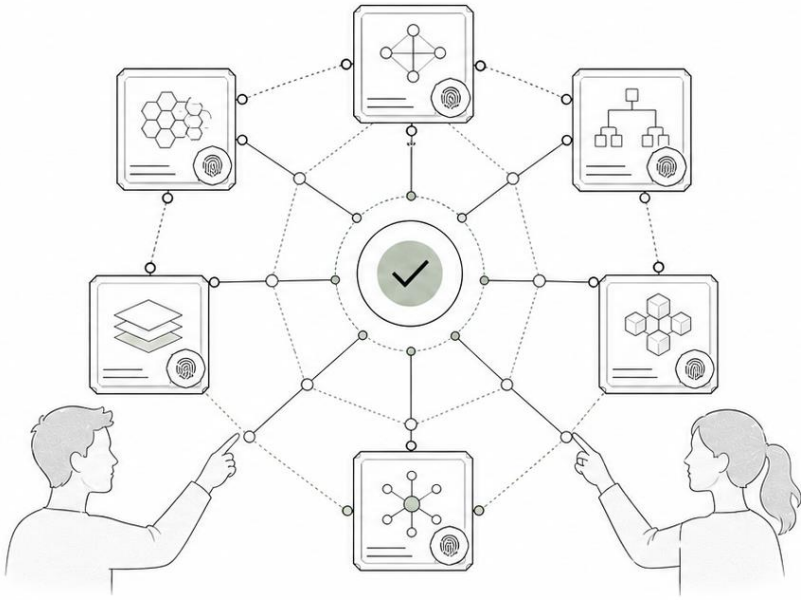
  Meaning stays checkable wherever the pattern travels.

Two graphs, two kinds of *memory*

Understanding Graph keeps the evolving trail. Sema stabilizes shared meaning.



[understanding graph]



[sema]

Sema can carry *any information*

I only planted the seed — the bootstrap library is where the vocabulary starts, not where it ends.

ANY DOMAIN

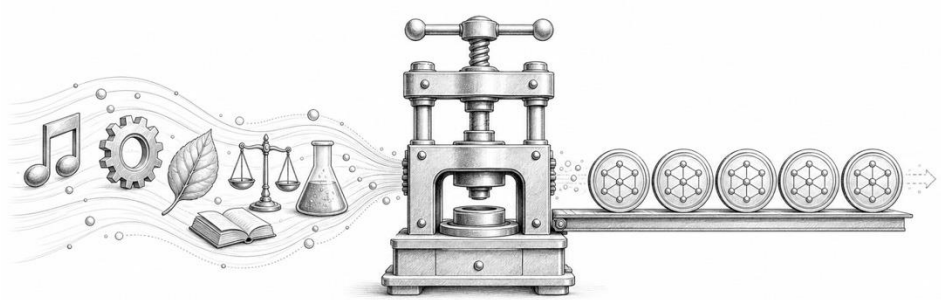
Medicine, law, engineering, your own codebase — any field can hash its patterns.

ANY AUTHOR

The bootstrap library is just the seed. Anyone can author and publish patterns.

ONE MECHANISM

Every new pattern earns its hash the same way, so meaning stays checkable.



Two agents, *speaking Sema*

A worked example: designing a safe ingestion flow, where every move is a shared, hashed word.

[worked example]

[agent-builder]

ContextFirst#a0b6: messy research notes, private user data — and future agents will retrieve from this store.

[agent-verifier]

Hard Constraint#7013: no raw personal data, and no claim without CiteBack#7785.

[agent-builder]

Decompose#63f3 the flow: select spans, summarize, Validate#337c against schema, Gate#bc63 the storage.

[agent-verifier]

RedTeam#15c5 the design, then AdversarialProof#80dd before release — hostile probes, not "I did not see anything."

the shared word for the workflow: DataMinimization#4acf → InvariantFilter#cc37 → CiteBack#7785 → AdversarialProof#80dd → AuditTrail#bf18

How a pattern *earns its hash*

Every card in the bootstrap library passes the same rulebook before it becomes vocabulary.

01 · DRAFT

Author a card

A pattern starts as a card in staging: handle, mechanism, and a signature of the form Intent(Target).

→

02 · WIRE

Place it in the DAG

Every reference is declared explicitly, and dependencies flow specific to general — the graph stays acyclic.

→

03 · VALIDATE

Pass the 42 rules

One validator, 42 checks: schema, dependency wiring, text–code consistency, novelty.

→

04 · ACCEPT

Earn the hash

Reviewers gate novelty, specificity, and fit; the accepted card gets its content hash.

FROM THE RULEBOOK

- **Explicit dependencies** — every placeholder declared, every one used
- **Gravity rule** — dependencies flow specific to general
- **One category each** — accepts, yields, composes with, or references
- **Fail closed** — one violation writes nothing

WHAT VALID MEANS

Valid means the whole audit passed. 452 patterns have earned their way into the bootstrap library, across four layers: Physics, Mind, Society, Infrastructure.

The library is not a *glossary*

Sema names reusable mechanisms, not loose labels.

[semantic contracts]

A GLOSSARY

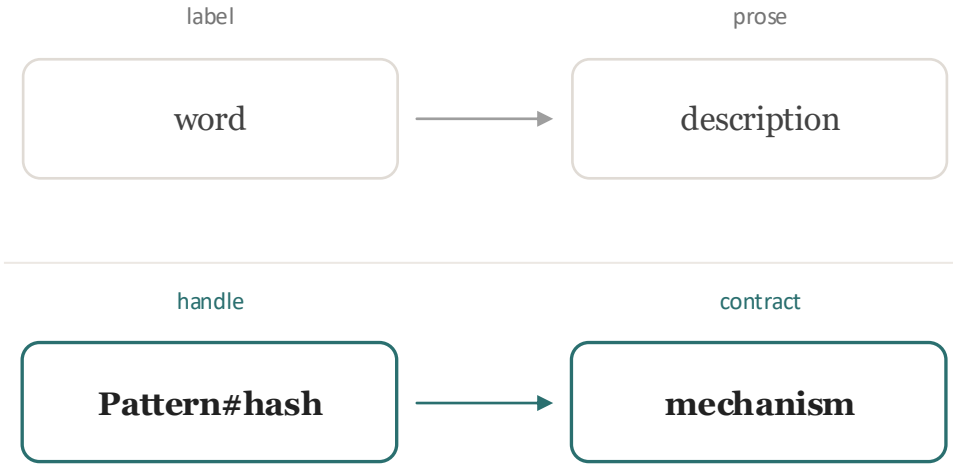
A label points to prose. Meaning can drift across authors, models, and contexts.

A PATTERN CARD

A handle points to a mechanism precise enough to hash.

THE DIFFERENCE

The word carries invariants, dependencies, and known failure modes.



WHY IT MATTERS

Sema turns vocabulary into inspectable machinery.

Placement follows *mechanism*, not topic

The question is what must be true for the pattern to run.

[topology beats topic]

LAYER	THE PLACEMENT TEST	EXAMPLE
physics	Would this hold with no author at all?	Entropy#323f
infrastructure	Can a program execute it without judgment?	Gate#bc63
mind	Can one knower execute it alone?	RedTeam#15c5
society	Does it require another independent party?	Vote#3b66

A pattern lives where its mechanism structurally has to run.

The vocabulary has a *civilization stack*

Patterns are grouped by the kind of reality their mechanism depends on.

[civilization stack]

society

mechanisms that need independent parties

Consensus#45f4 · StateLock#7cd8 · AdversarialProof#80dd

mind

mechanisms one knower can execute alone

BayesUpdate#3d1b · RedTeam#15c5 · SteelmanCheck#9c86

infrastructure

authored machinery that needs no judgment

Gate#bc63 · Validate#337c · Backoff#16c2

physics

constraints no author can wish away

Entropy#323f · Lock#95c2 · Causation#e47b

READING THE STACK

Each layer may build on the guarantees below it — never the other way around.

WHY LAYERS

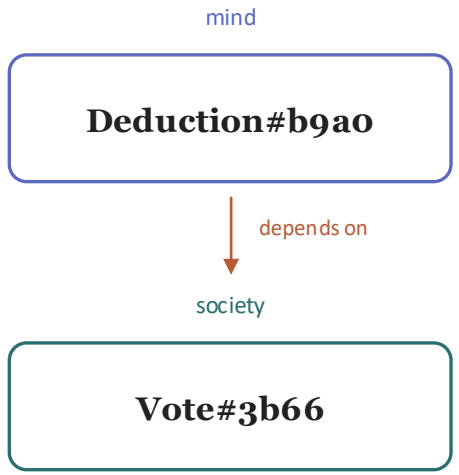
Placement records what the mechanism depends on, not what topic it resembles.

Layer direction protects the *guarantees*

Higher mechanisms may build on lower guarantees; lower guarantees should never need higher agreements.

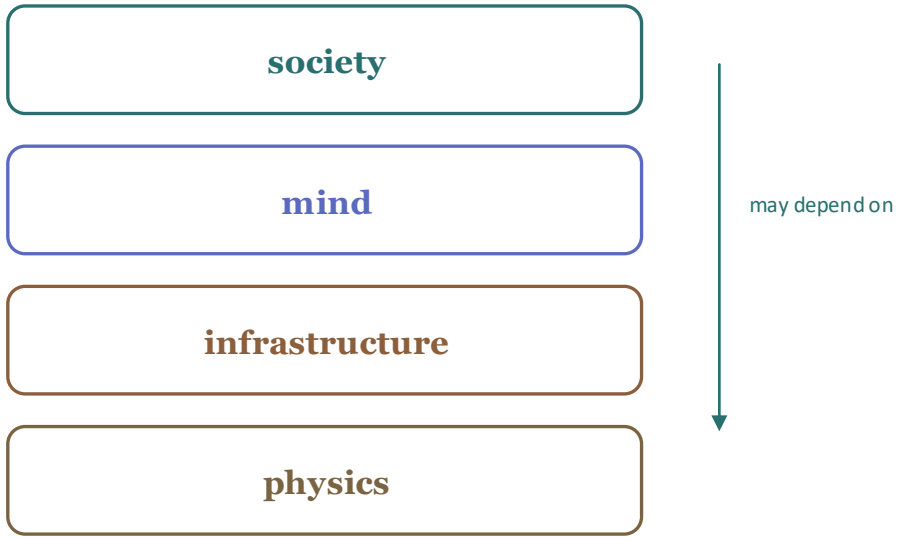
[dependency direction]

INVERTED



Truth becomes fragile when logic depends on political agreement.

STABLE



Dependencies point downward, toward stronger guarantees.

One pattern shows the stack *at work*

AdversarialProof#80dd is Society because the proof depends on adversarial search.

[AdversarialProof#80dd]

WHAT IT IS

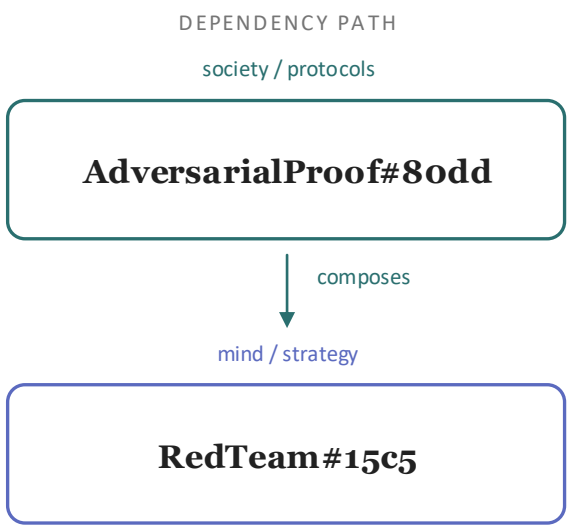
A negative proof that uses adversarial search to support a claim of absence.

WHY SOCIETY

The proof depends on independent pressure: a hostile search, not just a passive lookup.

WHAT IT COMPOSES

It invokes RedTeam#15c5, a Mind pattern for attacker-style probing.



WHY IT MATTERS

The handle makes the burden of proof inspectable instead of merely asserted.

Reusable patterns change the claim an agent can *make*

Without the pattern, the agent reports effort. With it, the agent carries a contract.

[the payoff]

WITHOUT A PATTERN

“I looked and did not find anything.”

WITH ADVERSARIALPROOF#80DD

Declared search space, hostile probing, coverage boundaries, known failure modes.

WHAT THE HASH DOES

It fixes the method identity, so another agent can inspect the claim.

WHAT IT DOES NOT DO

It does not make the answer true by itself.

What this *buys you*

What changes the moment words carry verifiable contracts.

Higher bandwidth	one handle carries a whole contract — to agents and to people
Verifiable contracts	check what a word means; don't just trust it
Fewer tokens	less confusion, less back-and-forth
Bootstrap harnesses	stand up an agent stack from shared patterns
Everything is nameable	any verb, any pattern of behavior, becomes a word

Everything is *open*

The papers, the code, and the conversations — join any of them.

the program	emergentwisdom.org
sema	semahash.org
emergent wisdom discord	discord.com/invite/yTZUbdDZu9
sema discord	discord.com/invite/hRhVqAuDYQ
on x	x.com/emergentwisdom





henrik.westerberg@emergentwisdom.org