Hermeneutic Ontology Engineering: LLM-Assisted Schema Induction for Oral History Knowledge Organisation

Jiajie Zhang[1], Andreas Vlachidis[1], Julianne Nyhan[1,2]

AFFILIATIONS

University College London, UK
Technische Universität Darmstadt, Germany

What testimony sounds like

NYHAN

...it is **the physicality of the operation** that always strikes me. That is something we have lost.

MALLOY

Indeed, we have lost that. ... I used to think that the slot accessed some long tunnel to the Goddard Space Flight Center, but when I was looking at the map I saw that **the Space Center** was actually not that close ... **They must have gone to a truck or something**.

INTERSUBJECTIVE — the interviewer's remark shapes what is told next

SITUATED — remembered from now, corrected by a map seen years later

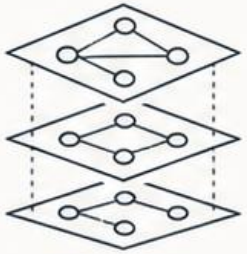
PERFORMATIVE — meaning made live: 'must have gone to a truck'

Oral history records not simply what happened, but how individuals construct meanings from the past.

AFTER PORTELLI, 1991

Research questions

01



RQ1 — MODELLING & REPRESENTATION

How can network structures represent the narrative and temporal dimensions of oral histories?

02

RQ2 — ANALYSIS & PATTERNS

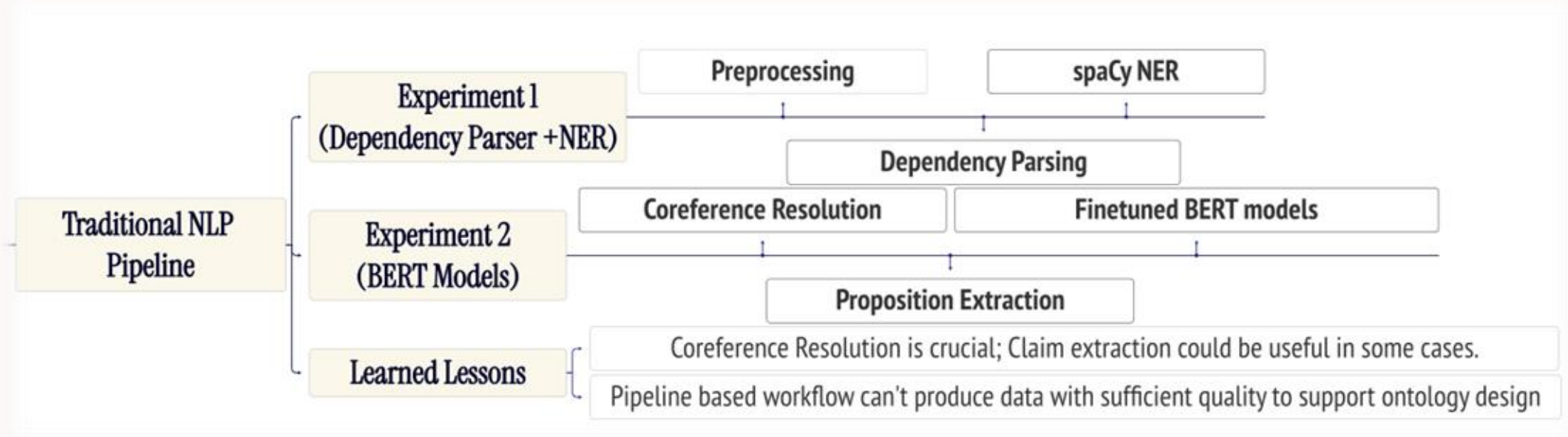
What narrative patterns of knowledge creation, mobility and influence emerge when we analyse these recollected relationships?

03

RQ3 — BRIDGING METHODS

How knowledge graphs create a defensible and transparent bridge between qualitative narrative and quantitative network analysis?

Information Extraction from Transcript



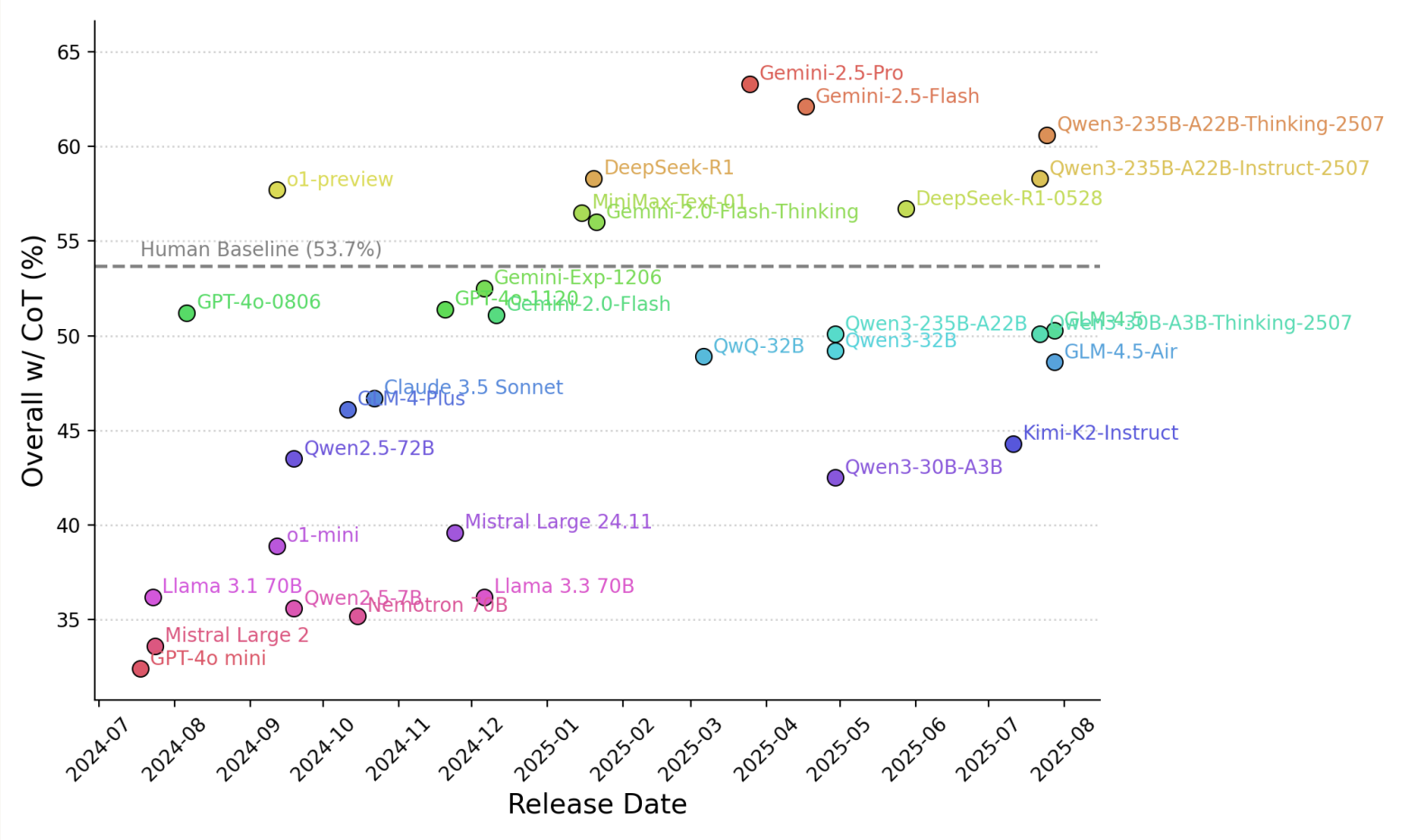
Traditional NLP Pipeline (Traditional ≈ before 2021)

I / IV — THE PROBLEM OF REMEMBERING

What generic AI

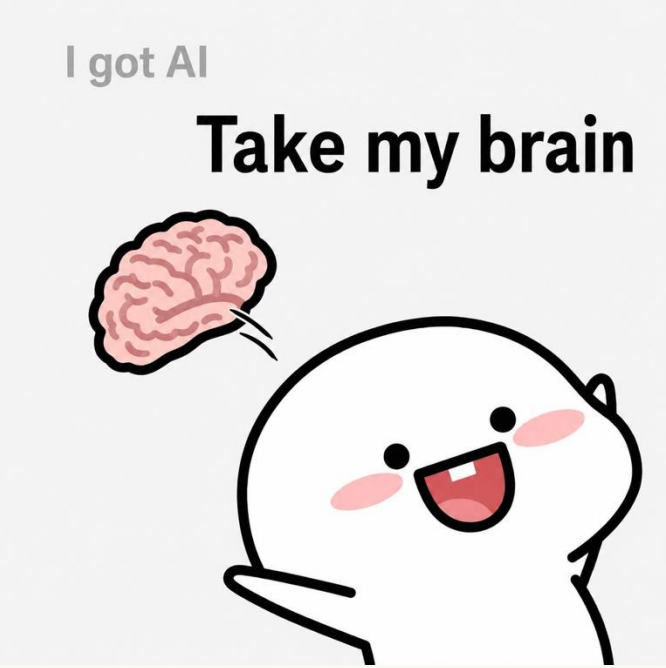
LongBench v2

Benchmarking Deeper Understanding and Reasoning on Realistic Long-context Multitasks



I / IV — THE PROBLEM OF REMEMBERING

What generic AI ?



What generic AI smooths over

Three ways a generic pipeline fails testimony —

01 PARAMETRIC KNOWLEDGE OVERRIDE

> "I had access to the University of California's libraries."
>
> — MALLOY

↓

Malloy — educatedAt → University of California

NEVER SPOKEN

a 'Wikipedia-shaped' pattern from pretraining overrides the sentence in front of it

02 FLATTENING THE HEDGE

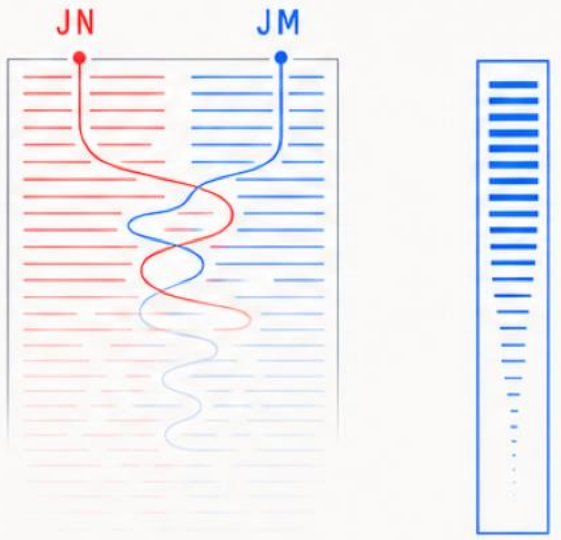
> "My first significant grant was, if I remember correctly, for $242,000, from the NEH."
>
> — NITTI

↓

Nitti — receivedGrant → $242,000 (NEH)

a bare triple has no field for epistemic stance — the doubt is erased

03 CONTEXT ROT



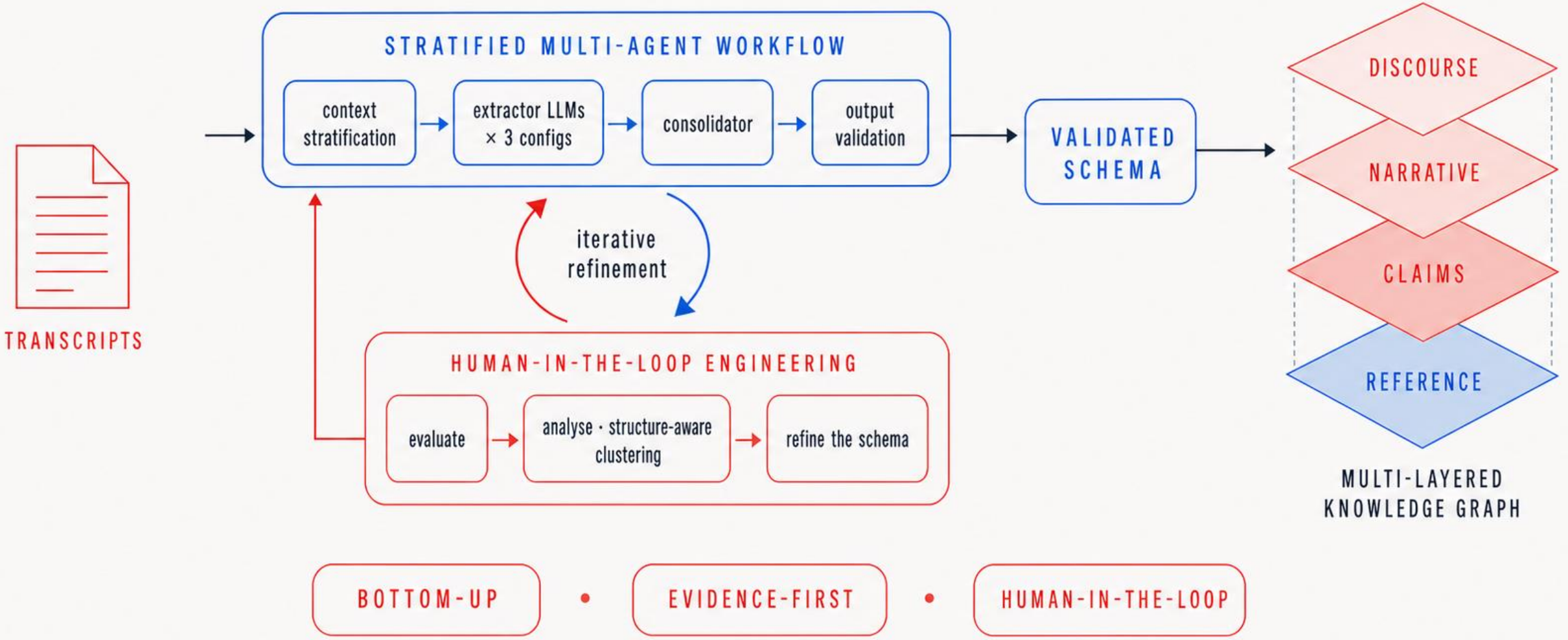
a full-length transcript dilutes attention: details blend, speaker attributions fail

fatal where meaning lives in nuances scattered across the whole interview

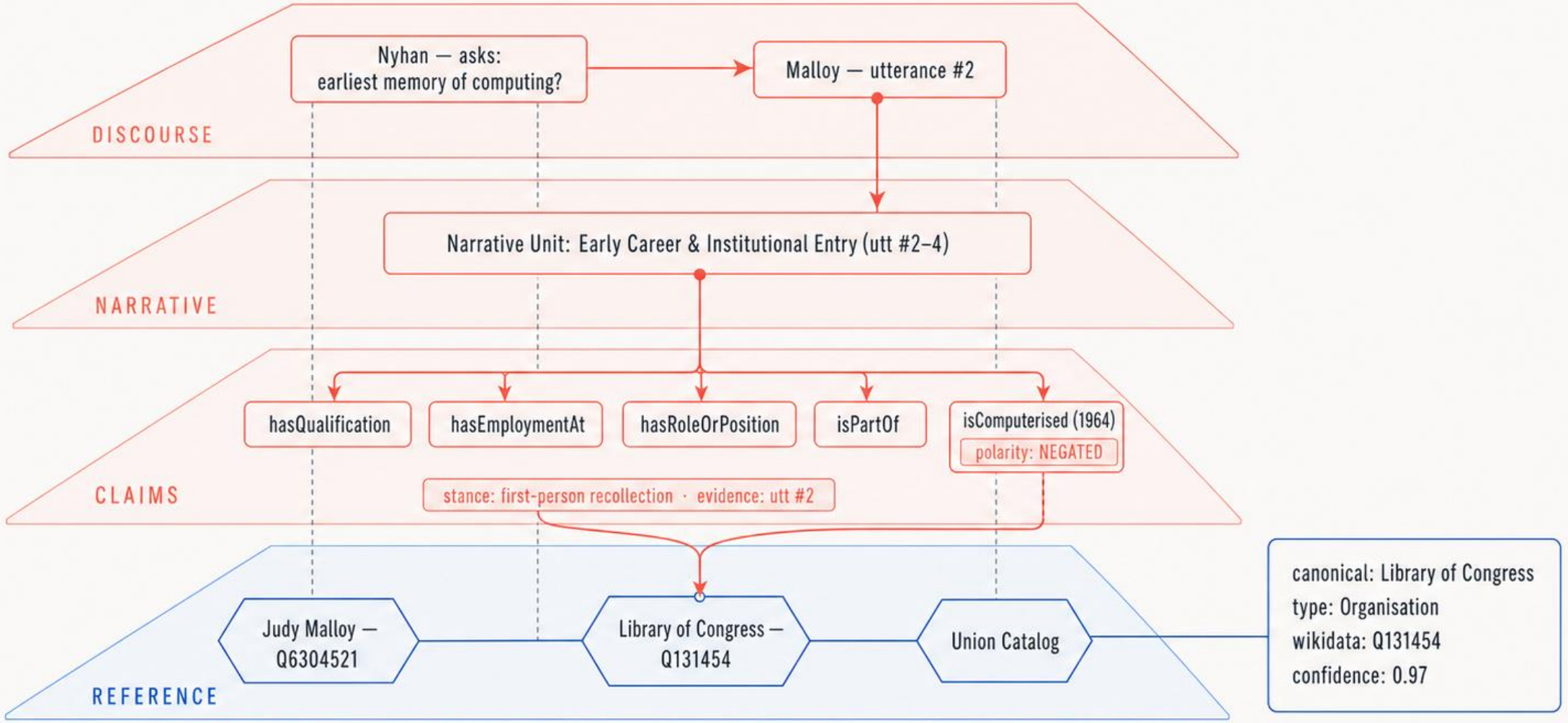
Three mechanisms, one common loss: fidelity to what was actually said. — our zero-shot runs · cf. Ji et al., 2023 · Hong et al., 2025

LLM-driven Hermeneutics Framework

The ontology as an instrument of reading, not a verdict — the LLM as a tireless reader, not an oracle.



Layers of information



One utterance → five verifiable claims — hedges, roles and negations kept. Nothing stored as bare fact.

III / IV — THE METHOD

Claims, not facts

A FLAT FACT

Malloy — worked at → Library of Congress

speaker, stance, evidence: gone

“ ⟨ Malloy — hasEmploymentAt → Library of Congress ⟩ ”

STANCE

- polarity: affirmed
- type: direct_experience
- certainty: certain
- temporal: distant_memory

PROVENANCE

- speaker: Judy Malloy · interviewee
- utterance #2

My first job … was at the Library of Congress …

VERIFICATION

- confirmed — 0 of 3 refute votes
- extractor: claude-opus-4-8

CONTESTED MEMORY, PRESERVED — WHERE DID THE TEI BEGIN?

⟨ TEI — takesPlaceAt → Modern Languages Association ⟩

said by RUTIMANN · certainty: probable

≠

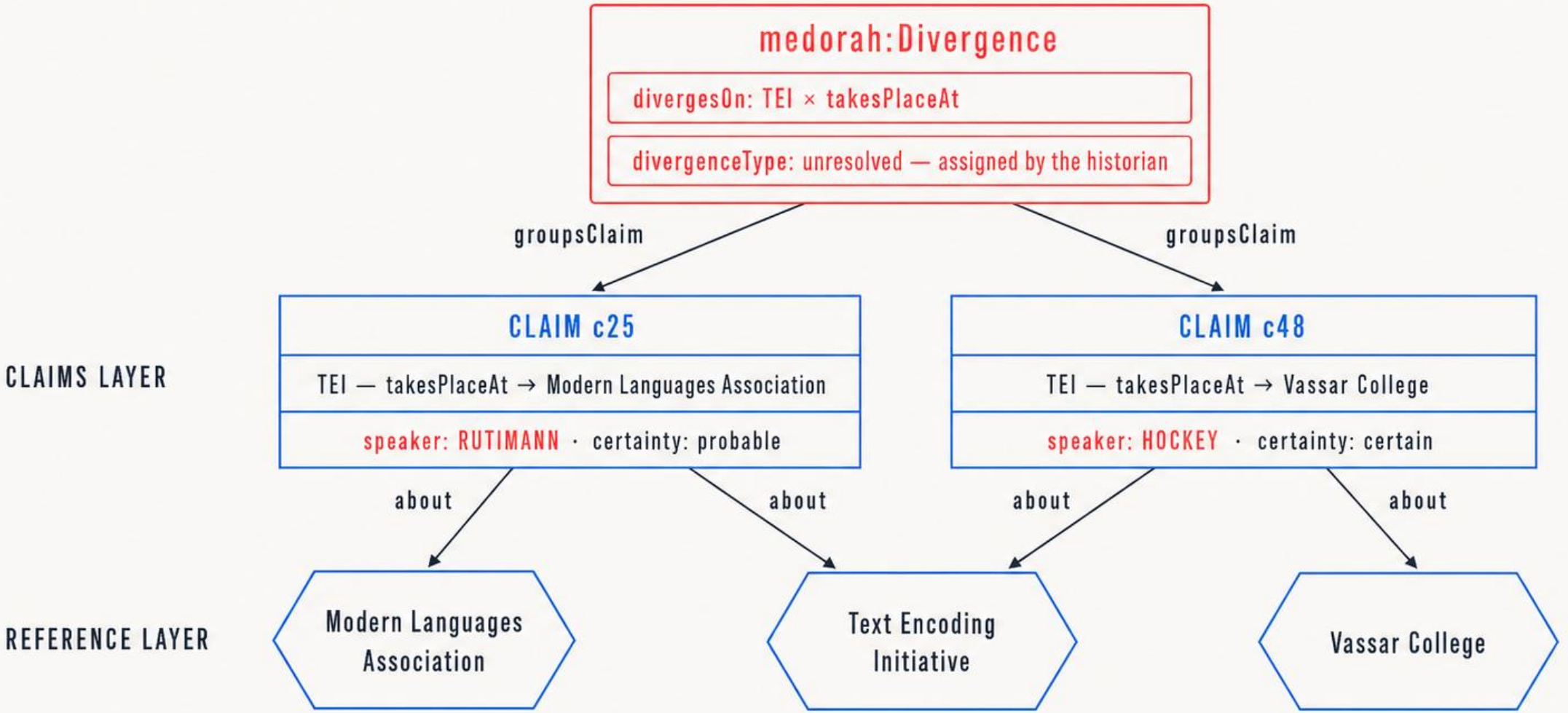
⟨ TEI — takesPlaceAt → Vassar College ⟩

said by HOCKEY · certainty: certain

Both stay — attributed, stanced, queryable. The graph preserves the divergence; the historian interpretes it.

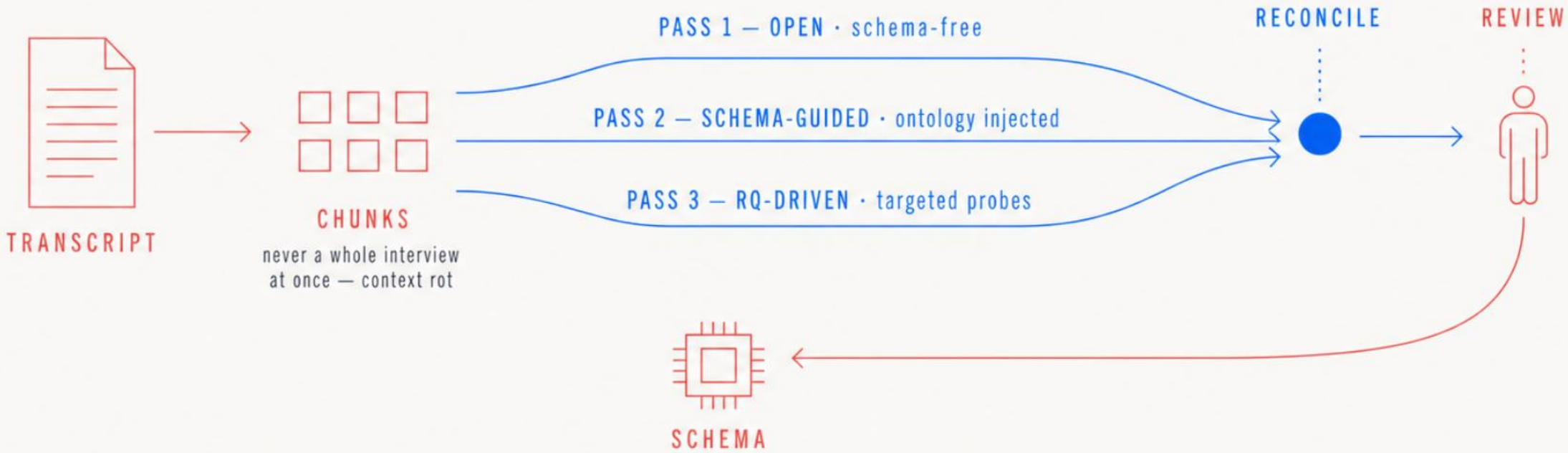
The graph knows how it knows.

The schema of a disagreement



Disagreement becomes a first-class, queryable object — noticed by the machine, typed by the historian, resolved by no one.

Three passes, designed to disagree



convergence → confidence

divergence → the schema's gap

model-agnostic — Gemini 2.5 Pro · Gemma 3 · Qwen 3

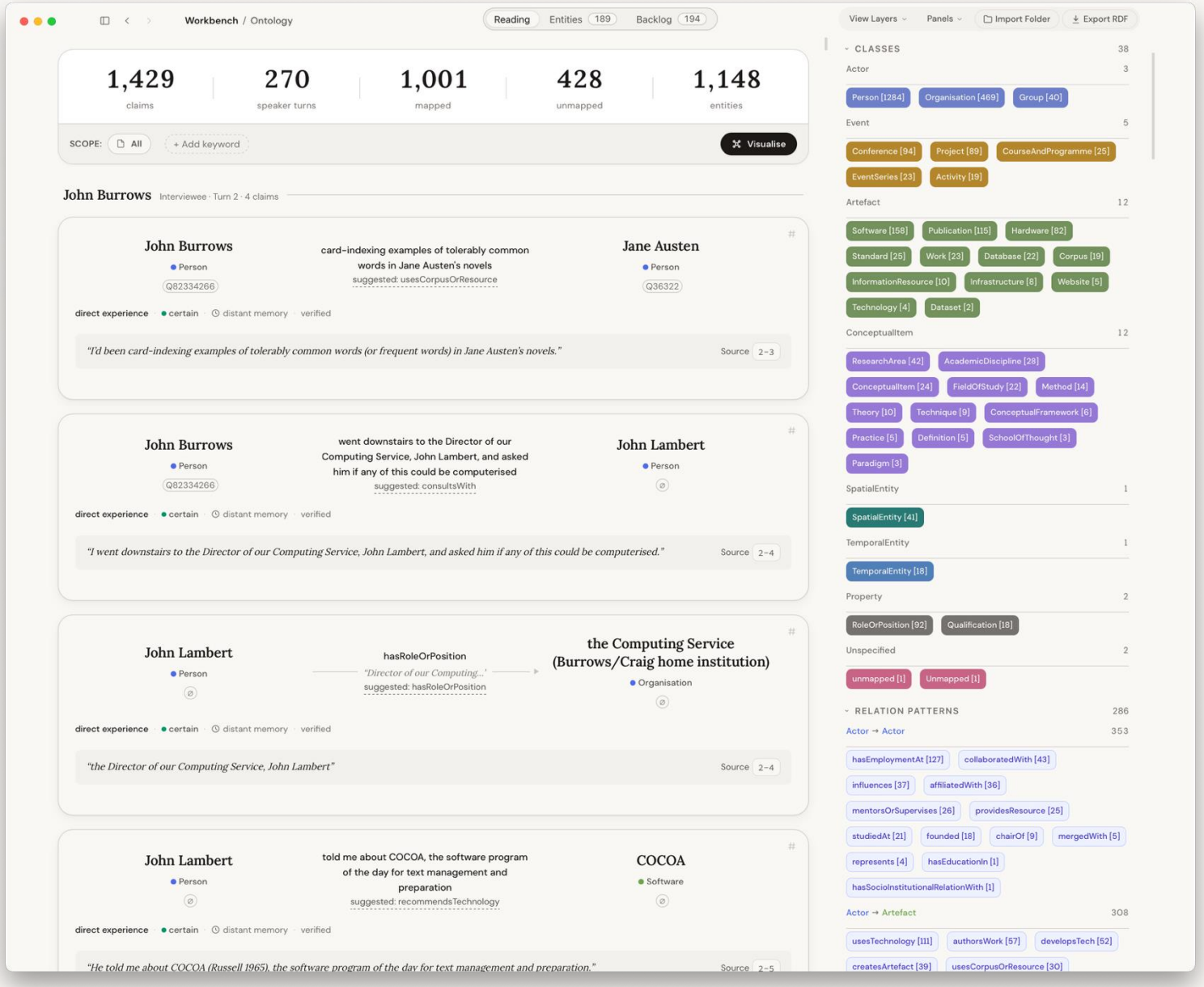
III / IV — THE METHOD

An interactive workbench

Judgement needs an instrument.

Macro View: The sidebar provides aggregated analytics, revealing frequent entities, relationships, and structural patterns across the entire dataset.

Micro View: The main panel grounds every extraction in its original narrative context, showing the speaker and the



III / IV — THE METHOD

An interactive workbench

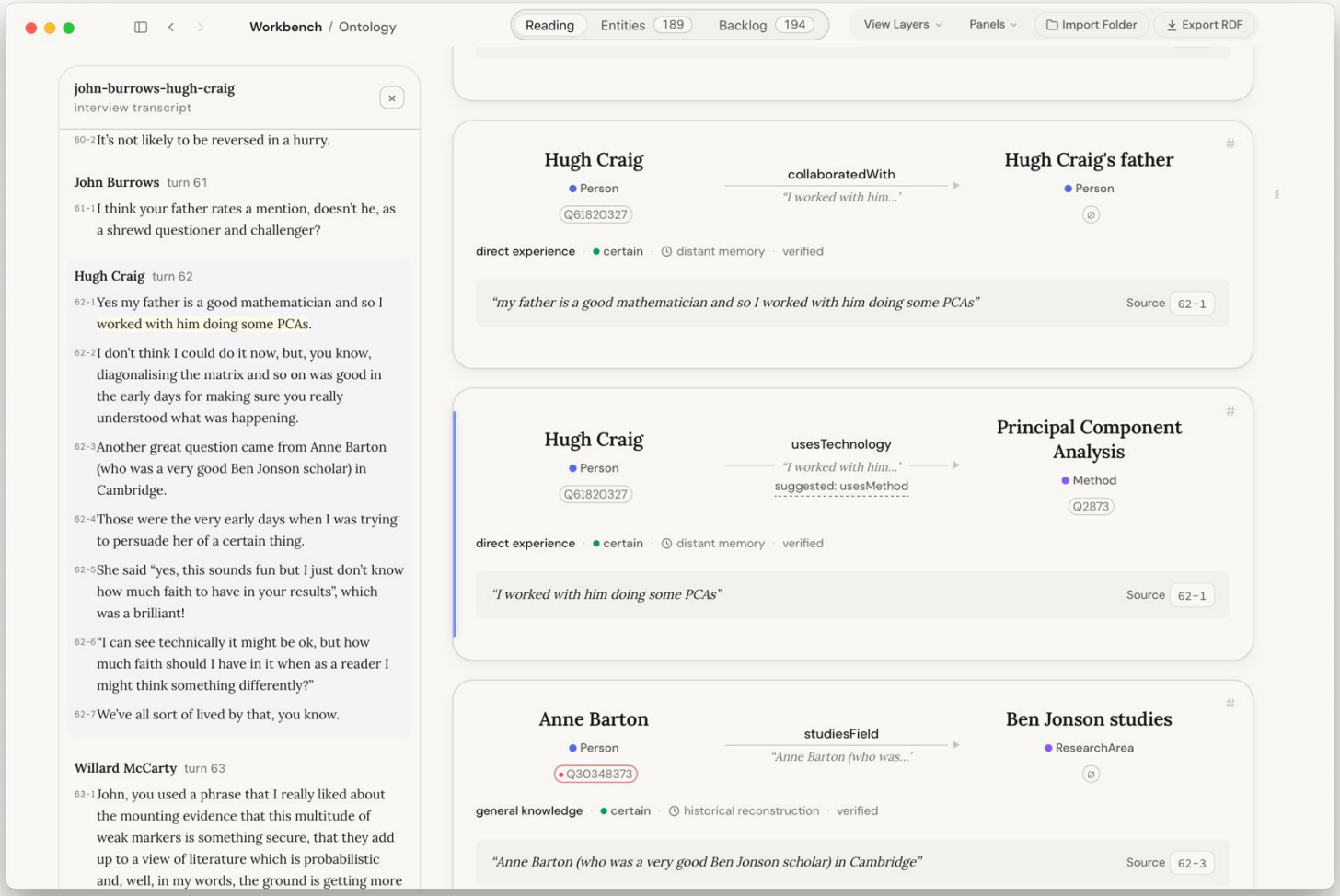
Judgement needs an instrument.

provenance-linked inspection — every claim beside its source span

structural analytics — how relations distribute

predicate clustering — candidate templates to accept or reject

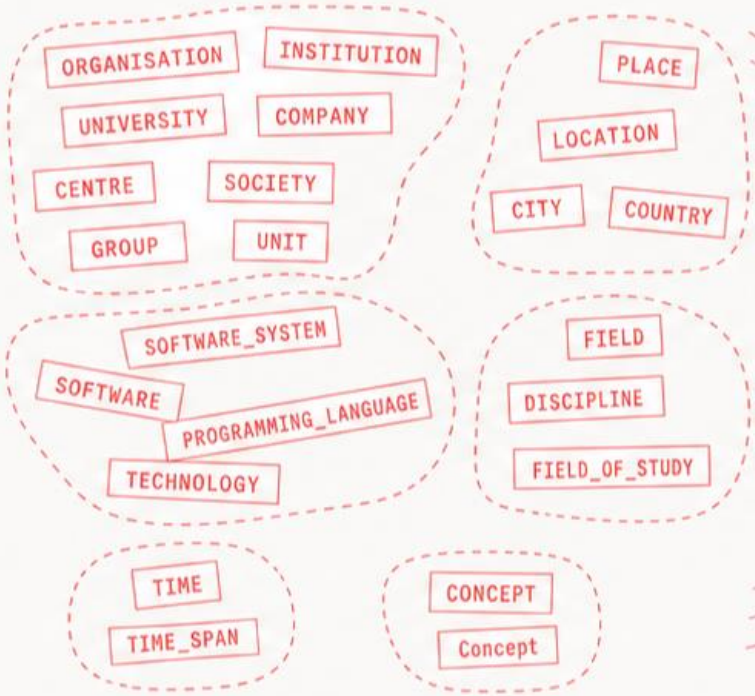
OPEN SOURCE



From 160 raw types to a governed ontology

160

raw entity types, schema-free pass — real labels:

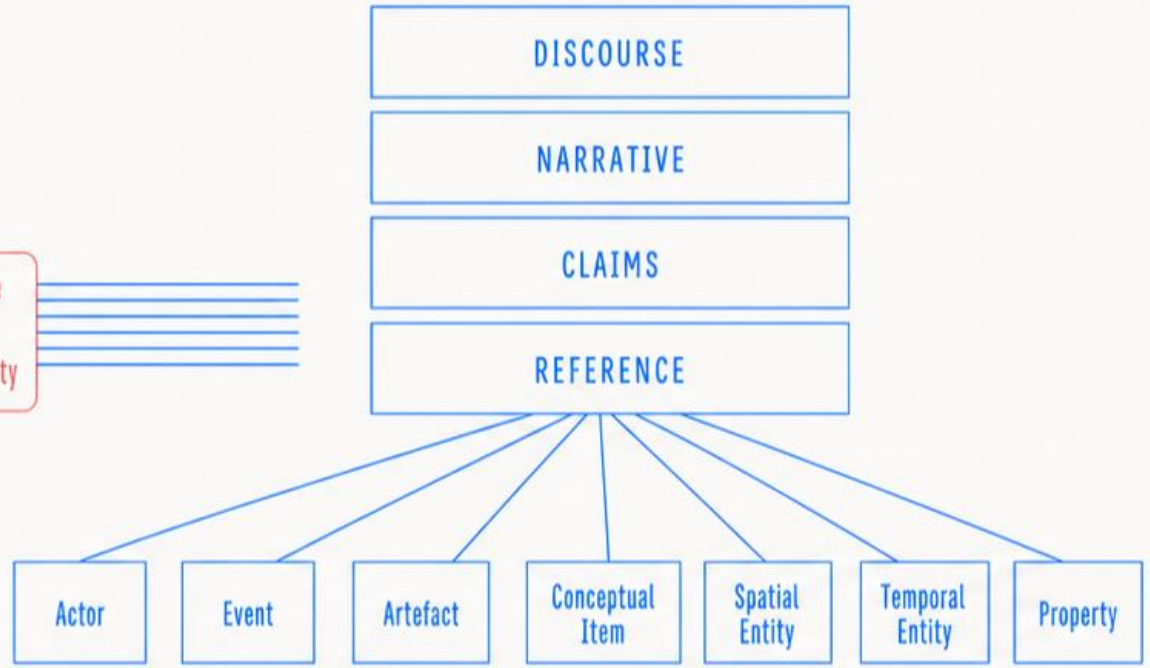


eight labels, one idea — the model mints a new type for every new word

7 : 1 — surface variation, not conceptual diversity

22

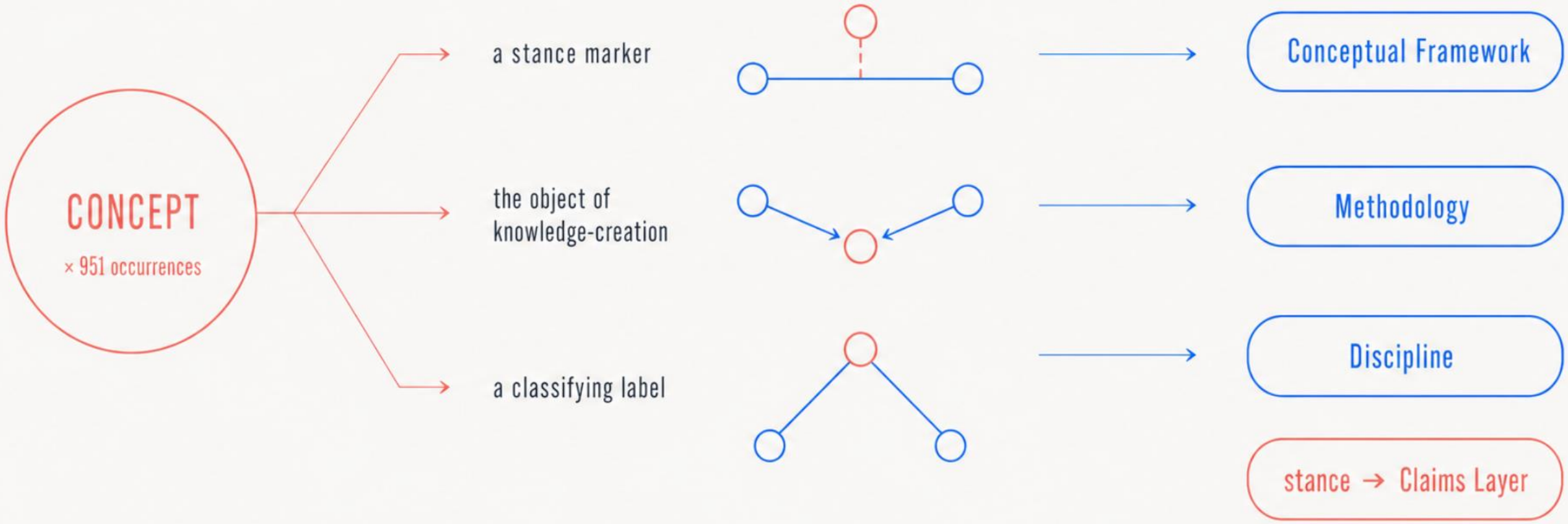
classes — 7 top-level + 15 sub



11 relation patterns · informal alignment: CIDOC-CRM · FOAF · Dublin Core

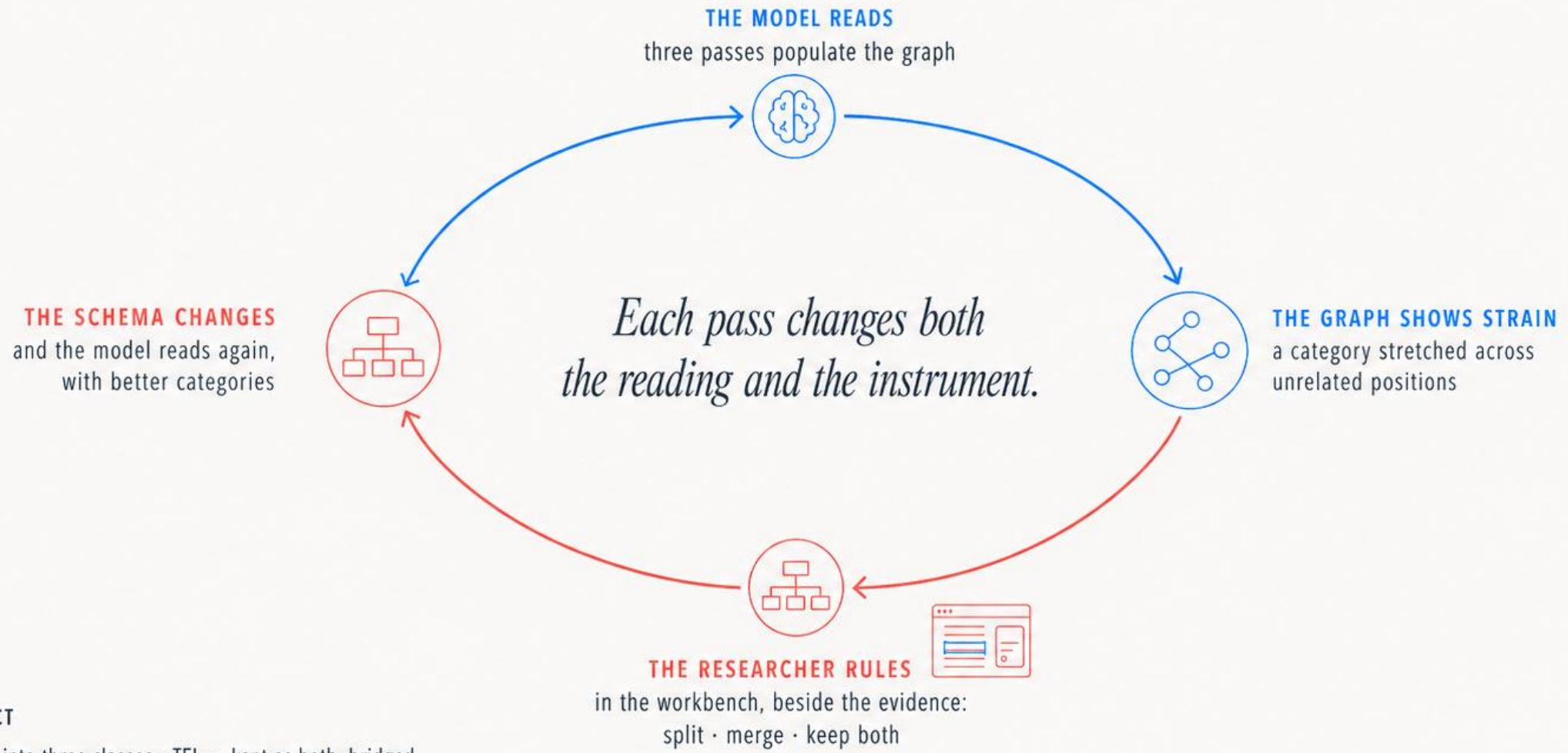
formal mappings deferred — deliberately: the schema is still moving.

The case of "Concept"



Invisible in frequency. Obvious in graph position.

The hermeneutic loop, in practice

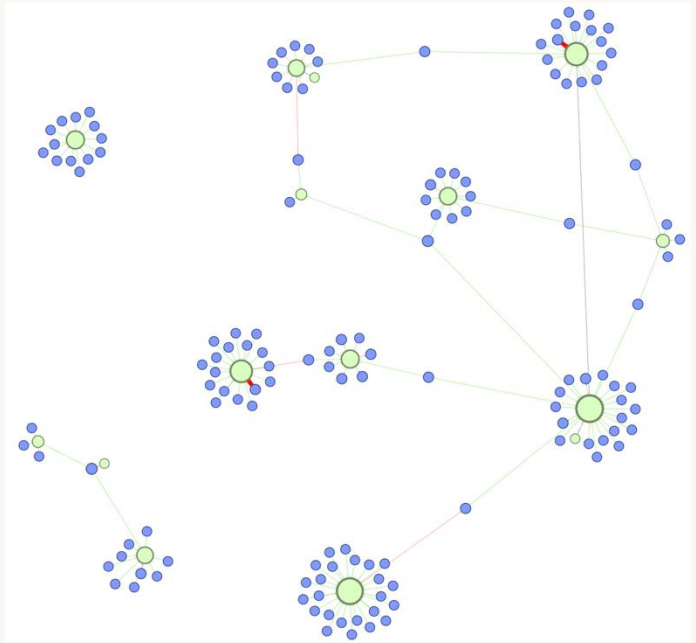


IN THIS PROJECT

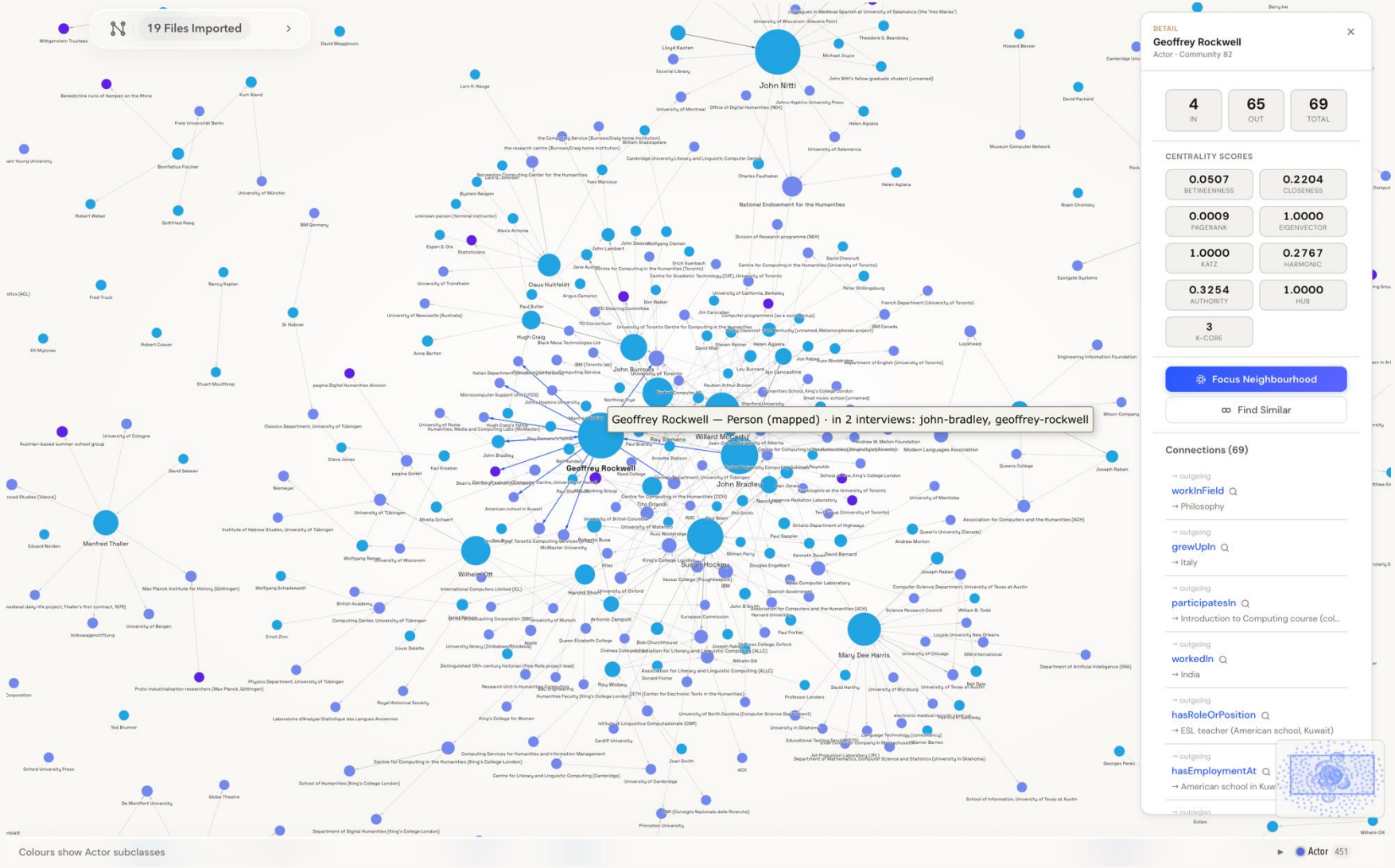
'Concept' → split into three classes · TEI → kept as both, bridged

Extraction stops consuming the schema – and starts testing it.

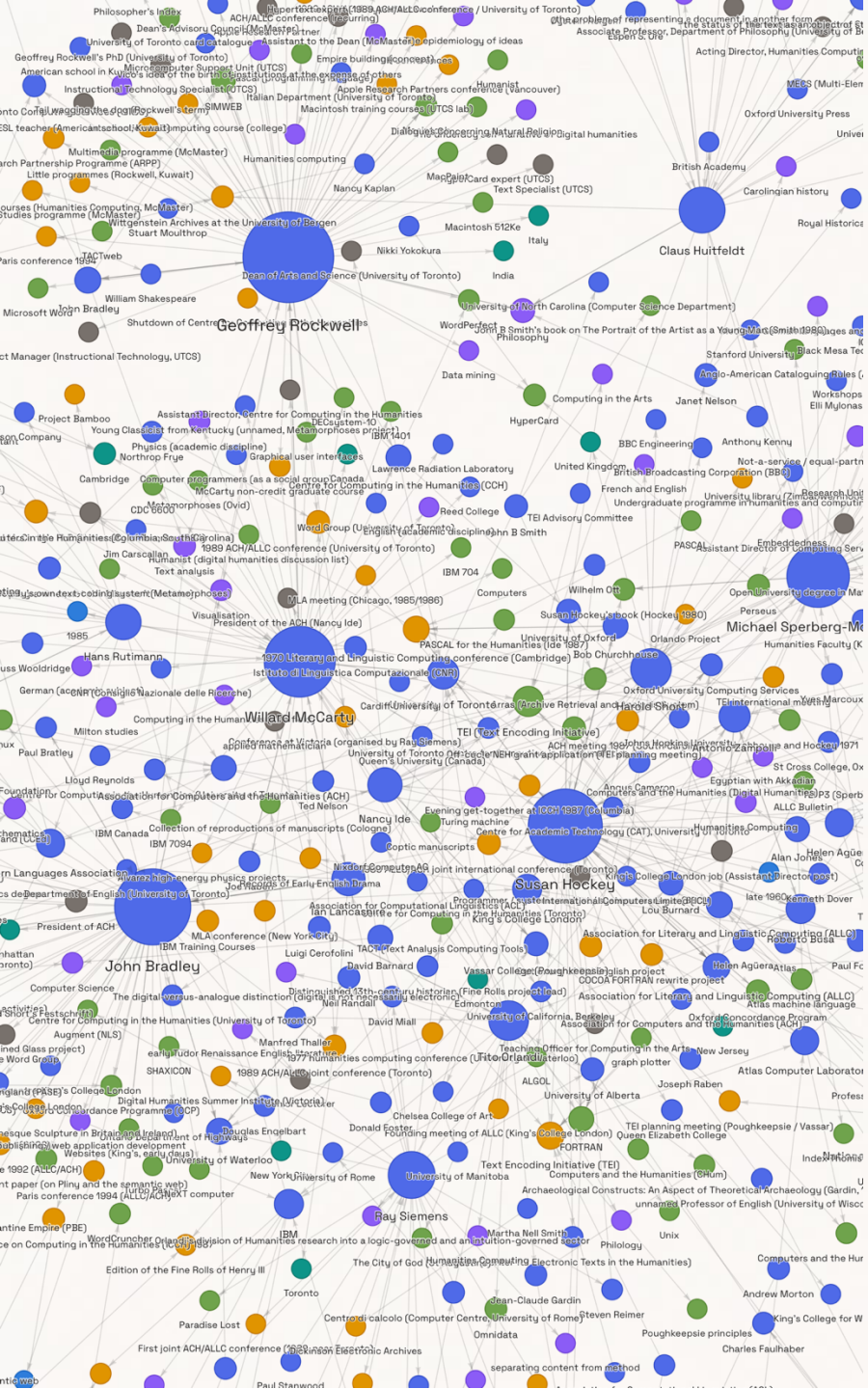
Expert Interpretation & Automated Extraction



Manual Annotation



LLM driven extraction with human in the loop



IV / IV — WHAT THE STRUCTURE REVEALED

The populated knowledge graph

Same questions, two readings — hand-coded 2025 vs the claims graph 2026.

	HAND-CODED '25	CLAIMS GRAPH '26
PEOPLE RECALLED	124	209
CLUSTERING	0.0975	0.0388 still star-shaped
TRIANGLES	57	51
MEMORIES MEET THROUGH	people P · P · P 33	institutions I · P · P 23
CENTRAL TECHNOLOGY	IBM 360/20 · TUSTEP	IBM 7094 · FORTRAN

Busa & Zampolli — never interviewed, still anchoring the remembered field.

35.7 % Wikidata-anchored · 25 hearsay · 169 probable — every cell of this table is queryable, with provenance.

QUALITY CENSUS

Reliable, not exhaustive

1,429 claims. Every one read back against the transcript by hand.

warranted — the evidence says what the claim says	91.3 %
over-specified — more concrete than the evidence supports	8.5 %
unsupported — no basis in the text	2 claims

WHAT THE 8.5 % ACTUALLY IS — INFLATION, NOT INVENTION

> *…the usual 1980s route of having a Sinclair Spectrum and Commodore 64…*
>
> — GABRIEL EGAN

→ Egan — usesTechnology → Sinclair Spectrum

owning is not using — the direction is right, the verb says one notch too much.

Selective by design — we extract what the evidence anchors, not everything a reader would notice.

CONTRIBUTIONS

What this reframes

01

Categories as hypotheses

interpretive commitments, held lightly and revised — not truth claims

02

Quality with contestation

bottom-up induction, bounded extraction, provenance — traceable, and still plural

03

Extraction as evaluation

the LLM as an instrument that tests the schema — not a consumer of it

And a lower wall between KOS designers and the communities whose knowledge is organised.

RESOURCES · OPEN SOURCE

Open code, open ontology

THE ONTOLOGY

medorah_ontology

github.com/articoder/medorah_ontology

THE WORKBENCH & TOOLKITS

medorah_nlp

github.com/articoder/medorah_nlp

MeDoraH

idohl.org/medorah/about