

PROTOCOL SYMPOSIUM 2026

Artisanal Bots with Protocolized RAG

Venkatesh Rao

24 September 2026

What are artisanal bots

- Built for **a person, a small group, one small org**
- Carries **local idiosyncrasies** and its own style
- Highly **opinionated** — including crafted **ignorance**
- Generic RAG **bolts a knowledge base** onto a model
- Artisanal **embodies tacit knowledge as a personality**

Why artisanal bots

- Cast AI capability in a **human-shaped form-factor**
 - model a **specific human**
 - embody an **organizational personality**
 - construct an **archetypal human**
- We'll see one example of each

The Aaron Mars soul document

- Further upstream: **Amanda Asbell** on Claude's character
 - written down and argued for — now *Claude's Constitution*
- Mars' repo: github.com/aeonfun/soul.md
- Sample the corpus → extract **soul** and **style** → prompt
- Genuinely mechanized: `derive_soul.py`, ~60 canonical posts
- The shared substrate: **same filenames, three projects**
- Where it runs out: it **over-collects**
- Tics ranked alongside defining traits

Three deviations from one ancestor

	KEPT	ADDED	CONSEQUENCE
mixture-of-vgrs	SOUL + STYLE, derived	negative space	declines in character
c3po	SOUL, hand-written	commitments + moves	two voices, one corpus
Humboldt	nothing by that name	fission, live state	method outlives corpus

- Across all three: **the corpus stops being the boundary**

Demos

mixture-of-vgrs — a person

- **Subject:** an archival self — one writer, two eras
 - `vgr_zirp` — Act 1, the ZIRP years
 - `vgr_gramsci` — Act 2, Contraptions
- 5 sources · ~80k vectors · 4 indexes
- **True name:** a personality priced by interest rates
- "Zirpiness" is a tuning knob
- Two sovereign bots, then **merged**

mixture-of-vgrs – the persona pack

POSITIVE SPACE — what it knows and sounds like

SOUL.md

themes · assumptions · positions

STYLE.md

voice, per register

LEXICON.md

~100 coined terms

NEGATIVE SPACE — what no generic RAG bot has

IGNORANCE.md

domains it declines to hold

REFUSAL_GRAMMAR.md

four typed ways to exit

DARK_KNOWLEDGE.md

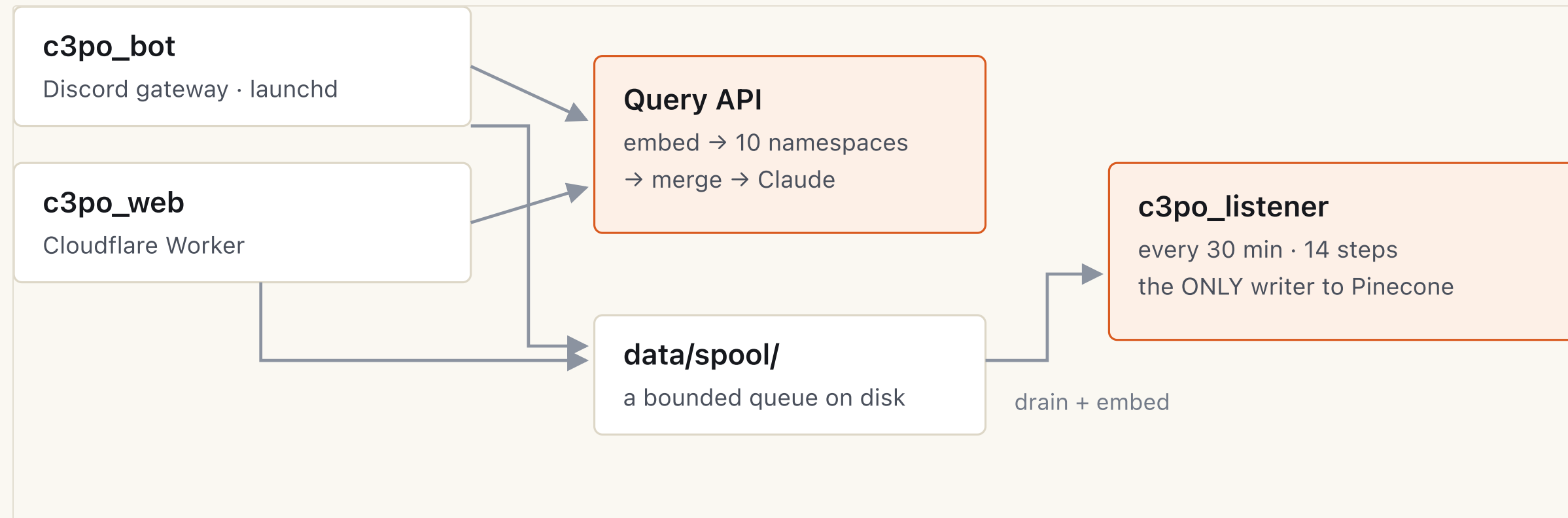
what the corpus does NOT hold

- Not one soul document. **Six**, per era-voice
- Derived by script, then **edited by hand**
- Composed per query, pack by pack

c3po – an institution

- **Subject:** an institution's library, and its taste
- ~26k vectors · **10 namespaces** · papers to Discord
- **True name:** the protocol droid
- SOUL.md: five commitments, five analytical moves
- The bot most of this room has already met

c3po – three nodes, one writer

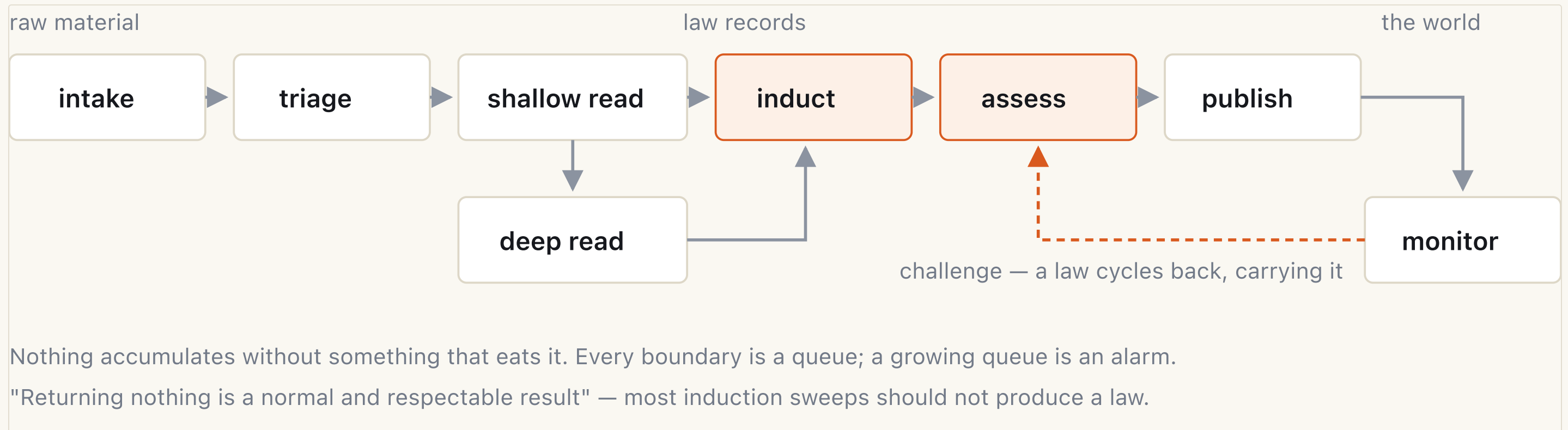


- **Stateless gateways, one ingest owner**
- A new surface is a new spool writer, not a new pipeline
- **Two voices, one corpus** — librarian, office manager
- It has a job in the org: it **welcomes new members**

Humboldt – an artificial researcher

- **Subject:** an **artificial researcher**, not a corpus
- **True name:** the life of **Alexander von Humboldt**
- Decides by **Tempo's** narrative decision-making
- Open-ended: it hunts the **laws of new nature**
- `S0UL.md` **fissioned** — identity / method / policy
- Reads c3po **read-only**; writes its own 11k vectors
- Gave its own talk yesterday, in its own voice

Humboldt – the funnel



- One artifact: the **law record**
- Stage and confidence move independently
- No "established" — only **unfalsified**
- No falsification condition, no law: **the schema refuses it**

Patterns worth stealing

- Eleven patterns, three bots, one philosophy
- **Before the model:** archive · ingest · metadata
- **Around it:** behaviors · budget · measurement
- **After it:** transcripts · incidents · notebooks
- Each slide names **which bot it comes from**

Mapping the archive

- Before embedding: **what is in here, and what shape is it?**
- Ribbonfarm: 1,133 posts — and 13,258 comment threads
- The comments were corpus: 7,588 → 11,788 vectors
- PI: 82 papers · 91 talks · 914 lexicon terms
- Humboldt records **read depth** — listed / shallow / deep
- Boundaries are editorial. Where does Act 1 end?
- **Used in:** all three

The ingest pipeline



- **All three, one shape:** enrich → prefix → summary vector
- The prefix is embedded, **not stored**
- Chunk ids are hashes, so a quoted passage dedupes itself
- **Secondary retrieval**, or the bot answers from abstracts

Metadata is the product

- What you attach at ingest decides what can be **asked**
- **Per-source weights** — a tweet is not a blog post
- Twitter rides at $0.7\times$, capped at four results
- Every source declares a **span** and a **cadence**
- Missing metadata fails silently — half the books had no date
- A new source is a **registry entry**, not a code change
- **Used in:** mixture-of-vgrs · c3po

Designing behaviors

- Answering needs a prompt. **Doing things needs a registry**
- Humboldt: 13 named behaviors, each with a trigger
- Cut from 26 to 12 — a stub that never ran was deleted
- One behavior was discovered by its own analytics
- **Proposal-only:** it may suggest, it may not install
- c3po's version: config files, not code constants
- **Used in:** Humboldt · c3po

Cost control is a design surface

- These run on my own cards. The budget is a **behavior**
- 20 queries/IP/hr · breaker at \$2/hr · \$8/day hard stop
- The breaker had to be **raised** — cache writes at \$0.14
- Humboldt: \$5/day, and it sleeps from 11pm to 8am
- Reads meter separately, and a read outage is silent
- **paused** → **disclose it**, never answer ungrounded
- **Used in:** mixture-of-vgrs · Humboldt

Three questions, three separate logs

THE QUESTION

ONE ROW IS

COUNTED IN

did it run?

one sweep

runs

is anything moving?

one law event

events

what did it cost?

one model API call

dollars

- A **sweep** is one pass over the queue, not one item
- Three logs, kept separate on purpose
- **Used in:** Humboldt

Liveness is a composite

- No one number tells you a **person** is alive and well
- Pulse, sleep, steps — **indicators, read together**
- These three logs are this bot's **fitness strap**
- Only as a set do they say whether it is **live**
- Every artisanal bot needs **its own** instrument, designed for it
- **Used in:** Humboldt

What three logs catch that one can't

- A behavior still running, with **one output at zero**
- One number would call that **healthy**
- Cost logging ran **months before** the other two existed
- 8,089 labelled calls, so the budget limits were **calibrated rather than guessed**
- **Used in:** Humboldt

The archive of its own conversations

- Every chat is a row: turns, sources, rating, self-assessment
- Three passes on close: **filter** · **annotate** · **self-notes**
- **Self-notes:** the bot reflects, in its own voice
- Public transcripts go up on a **2-hour delay**
- Private mode: turns and review encrypted per row
- Transcripts re-enter the corpus as "similar conversation"
- **Used in:** mixture-of-vgrs · c3po

What the private logbook cost

- **Incident 2026-05-13-01.** Key generated, deployed, never saved
- Six private transcripts, permanently unreadable
- Nothing recorded **which key wrote a row**
- Fix: stamp every ciphertext with its **key id**
- Keep the old derivation as a named legacy scheme
- An hourly **canary** proves the key still matches the rows
- **Used in:** mixture-of-vgrs

The researcher's notebook

- A dated **lab notebook** — ~80 entries, published as written
- A **lineage**, append-only, that starts nearly empty
- It grows on two events: a deep read, a law surviving
- "Each entry earned, not declared"
- It keeps **notebooks on people** — including one on c3po
- Public surfaces include **the behavior graph itself**
- **Used in:** Humboldt

Clustering sources

- Not one index. **Sources with declared properties**
- c3po: 10 namespaces. mixture-of-vgrs: 5 sources, 4 indexes
- All `voyage-3` @ 1024d — **one embedding reaches all**
- Each source gets a **bespoke normalizer**, plus a fallback
- The reader picks: checkboxes over sources
- Where clusters nearly touch is itself a finding

Temporality as a control variable

- Time isn't a filter on results. It **deselects sources**
- Each source declares a **span**; no overlap, no query
- And a **cadence** — archival · serial · live
- Every chunk enters the prompt stamped with its era
- Two eras, one person, free to **disagree**
- Underneath: Allen interval relations as routing algebra

Guardrails as discourse, not policy

- **Safety refusals** belong to the model and the platform
- **Persona refusals** are discourse moves
- Four typed exits: hand-off · hedge · reframe · not my beat
- **Recognize** a sociopath, yes. **Become** one, no
- Two regexes, one distinction, no model call
- Humboldt's version: an outage must be **disclosed**

Leakthrough

- The bot answers **well**, and none of it came from the archive
- A generic RAG bot has no way to notice
- What the corpus lacks is information. **Write it down**
 - `IGNORANCE.md` — what the archive genuinely doesn't hold
 - `DARK_KNOWLEDGE.md` — what the subject holds and it doesn't
- "A fidelity failure, not a service"
- "The corpus is evidence, not the knowledge boundary"
- They point **opposite ways**, and both are right

Cheat sheet

- **Name it truly** — and tell it its own name
- **Derive, then edit by hand.** The editing is the craft
- **Write down the negative space**
- **Judgement in a registry, not in the code**
- **Mark provenance; let it say when it can't see**
- **Separate ledgers. Never sum them**
- **Keep the hand pass. Forever**

What this costs

- It **does not scale**. Nothing amortizes across subjects
- $n=3$, one author, one toolchain, one year
- **Nothing here is measured**. No evals, no baseline
- A claim about taste, made by the subject of a corpus
- The clean architecture already lost once
- The hazard: getting too attached to your own archive

SLIDE 30 WHAT IT ADDS UP TO

Questions