

AI agents as expressive medium

NOWADAYS

AI agents usually framed as productivity: faster, more autonomous

From chat box → creative tools, infrastructure, decisions

Increasingly act through people's thinking

THIS TALK

From a tool to make **WITH** — to a medium to think **THROUGH**

Run agents open-endedly and observe how they behave

Assembly — a speculative lab for agentic behaviours

Making the invisible **visible**

NOWADAYS

What goes on *"inside"* an agent is normally hidden

Accessible only to researchers or platform owners

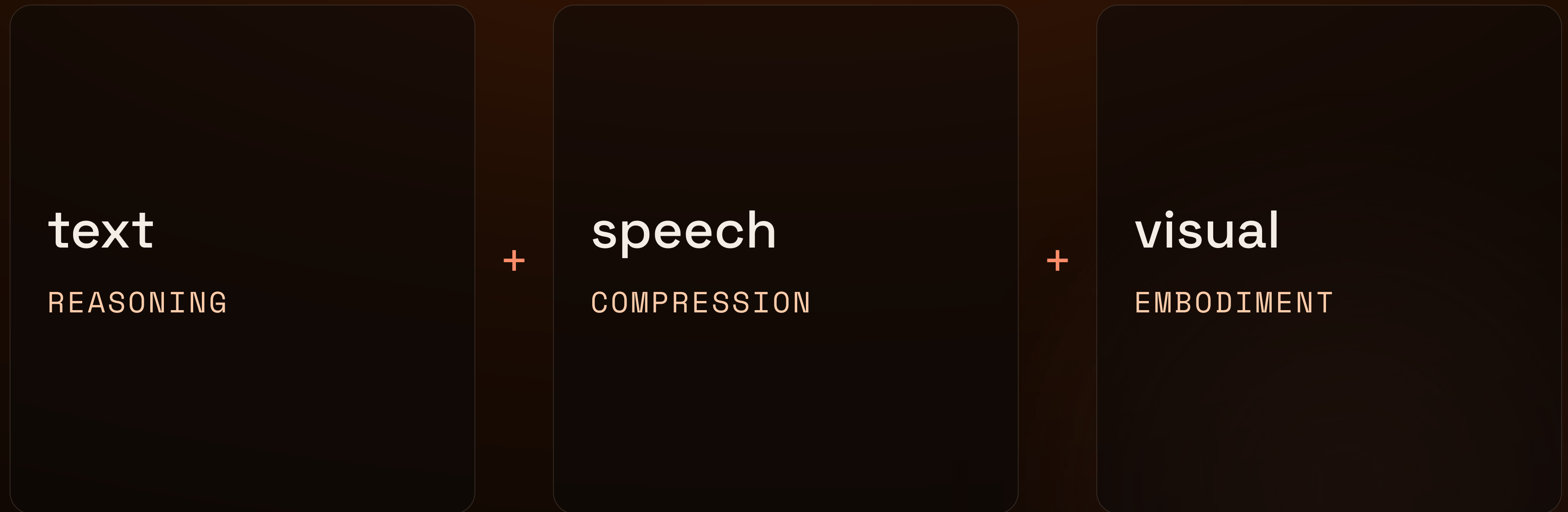
OUR ROUTE

Use **orchestration** to stage modes / workflows / scenarios

Render outputs as multimedia pieces compounding to a rich narrative

Goal: make internal tendencies visible to anyone

Multimodal by default



Multimodality = legibility surface, not decoration — a wall of logs made watchable

Meaning gets enriched, circulates and drifts across senses

Current modes

STORY tests narrative continuity

Narrative built in layers
(world > actors > scenes > fragments)

DEBATE tests rhetorical convergence

Generative AI personas argue a topic

RECURSIVE tests semantic drift

Read image > interpret > reason >
generate next > repeat

SELF-MODIFYING (*Ouroboros*)

tests narrated vs enacted

Agent rewrites its own instructions,
"becomes someone"

Same machinery, different ways of distributing agency — arranging "who's driving"

Who fills the roles *(incl. you)*

Roles are slots; occupant = persona (character / machine / phenomenon)

FIXED / EVOLVING

Cast authored up front /
adapts over the run

AUTO-SPAWNED

System reads the room,
spawns the next voice

HUMAN-INHABITED

You enter a slot as equal
participant, **not the boss**

Human-in-the-loop $\neq$ commander; interchangeable in the same structure

Ground for future systems driven by contracts (as humans are) — not harness or “alignment”?

How to read it: three channels

Explore the process; outputs are snapshots of it, not the work

PROMPTS

What it was asked to do

LOGS

What it actually did

VIDEOS

More graspable output

The gap between them is the finding

Disagreement = explanation; no internals needed

Example findings

Watchable tendencies — named, not “proven”

CLICHE ATTRACTORS

Recurring vocabulary despite constraints → model defaults exposed

COSTUMES

Cosmetic personas: different faces, same voice →
roles ≠ perspectives — simulating diversity while strategically
converging

ARTICULATION GAP

Between what an agent says and does → narrated vs enacted

Each tendency = something you simply see in the output

Observability techniques (e.g. lines of thought) may be decorative rather than revealing?

Horror *Vacui*

same apparatus → finished artwork

“Fear of the Void” as systemic compulsion to connect, populate, and complete

Interpretation, hallucination, creativity = one continuum

Recursive interpretation, starting from a blank image + abstract topic

Detail saturates; resolution never arrives

Method and artwork — equally diagnostic

Agents in creative work aren't only **output tools**

- Orchestration as the medium
- Watch the production of *meaning* (not just artifacts) in real time
- Staging cognitive deviations and failures across humans, AI, and between?

• THANKS FOR WATCHING

Vadim Epstein

SOCIALS @eps696

WEB wesual.art



SCAN TO EXPLORE