

# A fleet that watched itself work

produced by a Noogram agent fleet · [noogram.org](https://noogram.org)

ENS · CSD coding-agents seminar · 2026-06-18

# Emmanuel Sérié

**Applied mathematics, then quantitative research** — physics PhD, then years inside quantitative research.

**Teaching and research today** — École Polytechnique, CMAP, the MaQI master.

*One question today: can a fleet of agents do real mathematics — and catch itself when wrong?*



# Noogram — an instrument built in the open



**A research programme, not a product** — federative agentic AI, human-amplified, never replaced.

**Live today** — the public core is shipping, not promised.

✓ core: [Cosmon](#) · [AGPL-3.0](#)

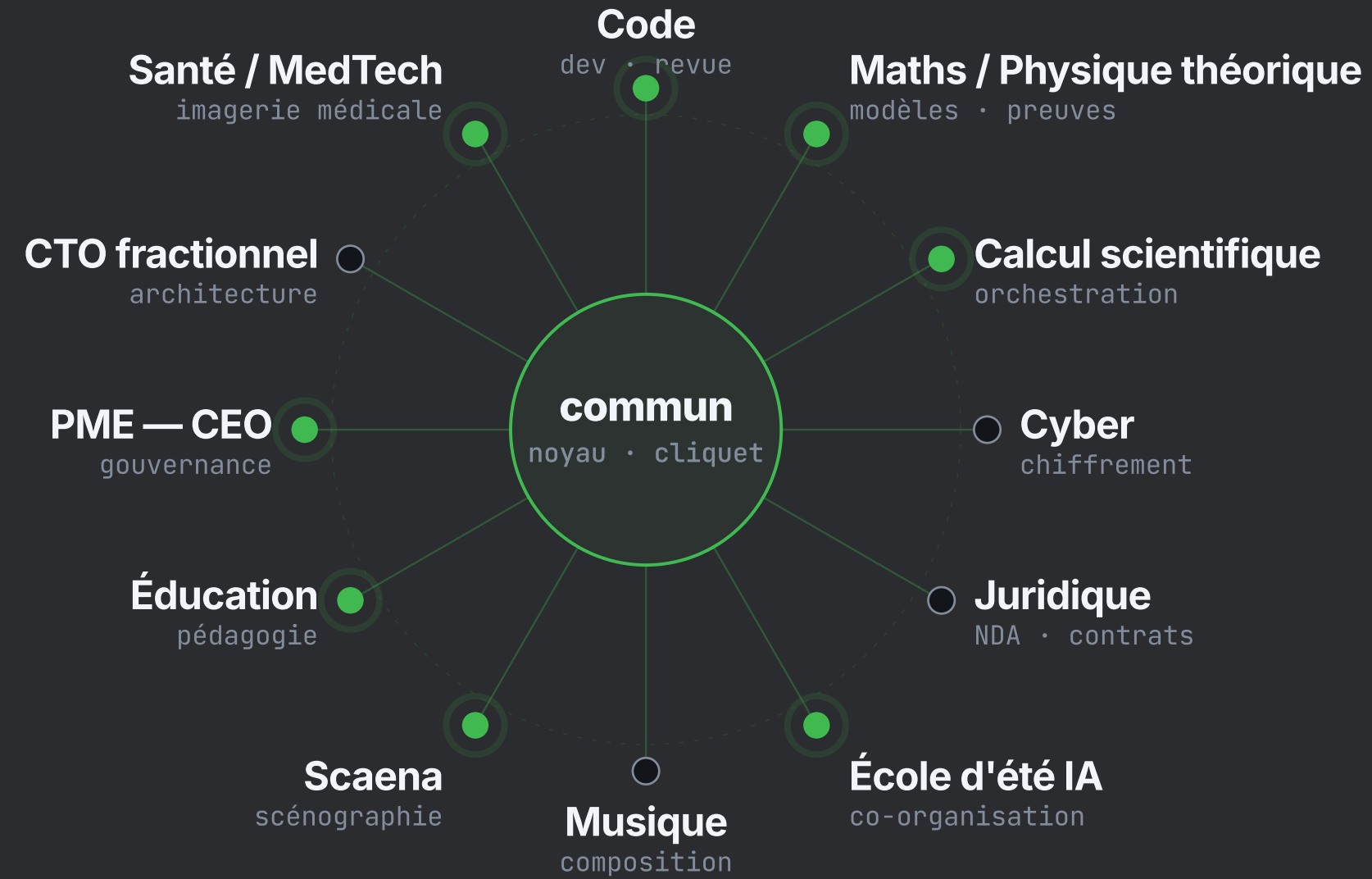
✓ engine: [OxyMake](#) · [public](#)

✓ [godel-kernel](#)

✓ [noogram.org](#) · [live](#)

# One instrument, many crafts

commun · noyau · cliquet *at the center* — *every spoke is a different craft run on the same open core. The green dots are live public artifacts: click one to open it.*



# Cosmon — the engine, under your control

**Four principles, formally pinned** — self-reference, transport-and-cognition, intentions-not-ownership, minimum-action; TLA+-validated.

**Built by the method it runs** — agents propose, a chief decides.

✓ commits: 10.7k

✓ tests: 6.8k

✓ Rust: ~490k lines

✓ agent co-authored: 1.6k

*La Formule 1, pas le pilote — one mission it just ran.*

# The question, verbatim

**One line in a to-do file** — a researcher dropped a half-formed maths question into `todo.md`: flow matching versus optimal transport, on Gaussians.

**We changed nothing** — that line, untouched, is everything the fleet was given.

# Not one agent — a fleet

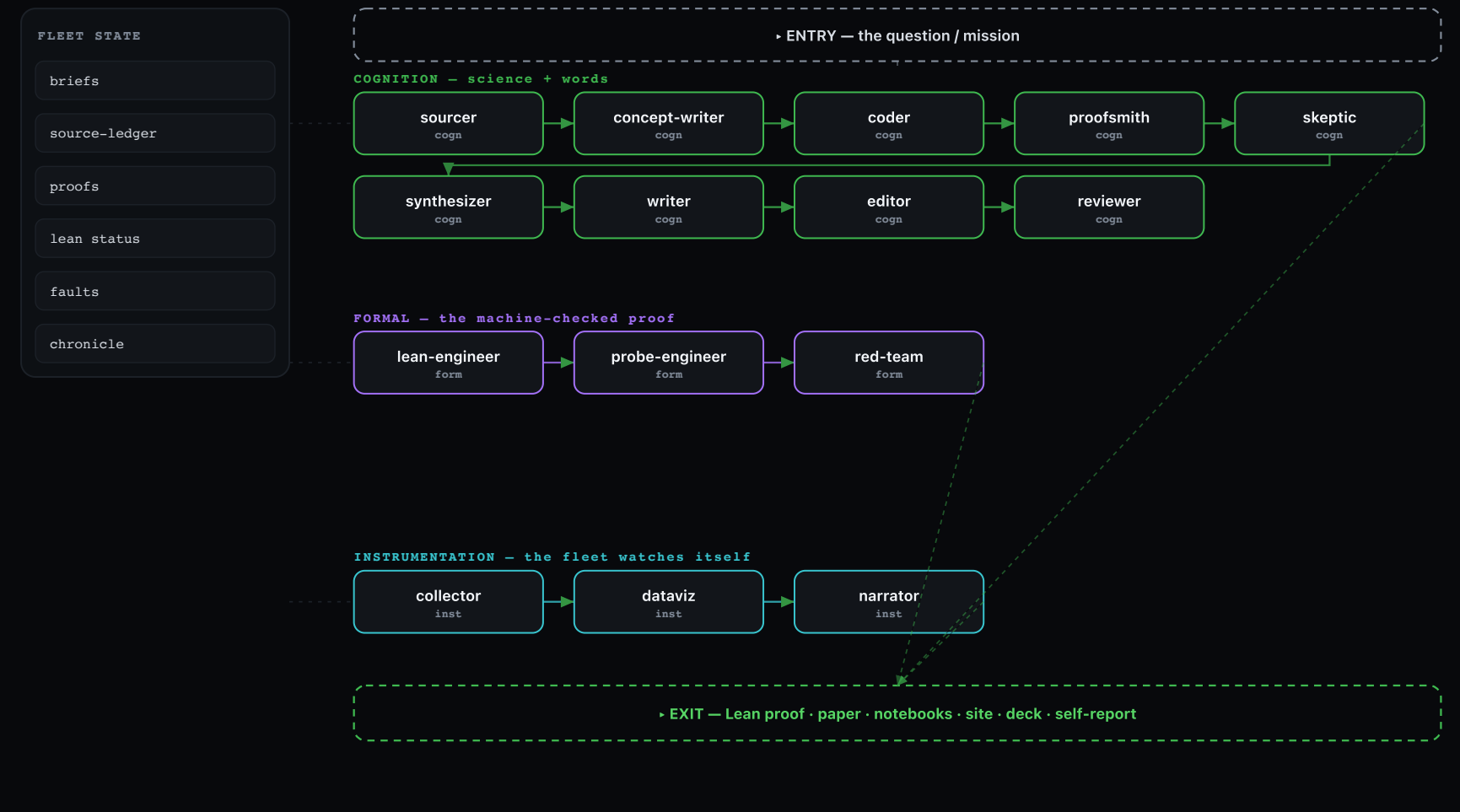
**One chatbot is one researcher** — fast, fluent, alone in the room with its own mistakes.

**This is a research group** — many agents split into branches, with a referee built in: a sub-fleet whose only job is to attack the others' work.

# Who does what — the role graph

## Fleet architecture

three sub-fleets · a question in, the artifacts out · hover any role



**Three branches, one referee** — a question in, the artifacts out; the **red-team** role exists only to attack the others' work. [zoom / more graphs](#) ↗

# The plan lays itself down



**17 tasks, written in one breath** — the opening science plan, before a single worker moves. Four branches: notebooks, paper, proof, self-report.

# It rewrites itself mid-run



**It found its own mistake and rewrote the plan** — 5 repair tasks it never planned punch in, wired *backward* into a branch it thought was done. [zoom / more graphs ↗](#)

# The result, stated formally

**Schedule-independent; commutation = symmetry = the Bures-Wasserstein OT map:**

$$[\Sigma_0, \Sigma_1] = 0 \iff T_1 = T_1^\top \iff T_1 = T_{\text{OT}} \iff T_1 = \Sigma_1^{1/2} \Sigma_0^{-1/2}$$

# It is machine-checked

**No charisma gets past the kernel** — a machine read the proof and agreed. This is the part that does not get to bluff.

✓ lake build: `exit 0`

✓ sorries: `0`

✓ project axioms: `0`

✓ Lean: `v4.29`

# The verifier caught us — we fixed it for real

**A draft of the hard direction was wrong** — an invalid step, plus a gap in dimension three and up. The adversarial sub-fleet flagged both.

**We closed it, unconditionally** — on the straight-line schedule a closed form forces commutation. No prose was patched.

# See it yourself

**flow-matching-gaussians.pages.dev** — the proof, the notebooks, the paper: all live, all generated by the fleet.

**The trust is not in the agents** — it is in the gate they could not argue with.

✓ adversarial corpus: 9 / 9

✓ sweep: 9000 pairs pass

✓ paper · 2 notebooks · Lean proof

# Appendix — backup slides

*Everything past this point is past the end — reachable during Q&A with the arrow keys or slide overview.*

# A • The converse — why symmetry forces commutation

**Orthogonal factorisation** — the flow map splits as  $\Phi_t = \Sigma_t^{1/2} O_t \Sigma_0^{-1/2}$  with  $O_t \in SO(d)$  and  $O_0 = I$ .

**Time-ordering is the obstruction** —  $T_1$  is symmetric iff the residual rotation  $O_1 = I$ ; off the commuting locus the time-ordered (Magnus) exponential leaves a real rotation behind.

**On the affine schedule it is algebraic** —  $T_1 = (\Sigma_1 \Sigma_0^{-1})^{1/2}$ , so symmetry forces  $\Sigma_1 \Sigma_0^{-1}$  symmetric, i.e.  $[\Sigma_0, \Sigma_1] = 0$ .

# B • The Lean formalisation

**Mathlib, kernel-checked** — the commuting case is a machine-verified anchor; the shipped source blob matches `kernel-provenance.log`.

**Author is not the scorer** — a nine-category adversarial corpus probes the proof: eight must be rejected, one must build.

✓ lake build: `exit 0`

✓ sorries: `0`

✓ project axioms: `0`

✓ corpus: `9 / 9`

# C • Three schedules, one map

**Linear, variance-preserving, cosine** — three interpolation schedules  $(a, b)$  between the same endpoints.

**The map does not care** — Proposition S:  $T_1$  is identical across all three (zero curvature of the connection  $\omega = \frac{1}{2} d\Sigma \Sigma^{-1}$ ).

**3-Dirac mixture vs OT** — the notebooks plot flow-matching trajectories against the Bures geodesic; they diverge exactly off the commuting locus.

# D • The Bures–Wasserstein OT map

**OT between Gaussians is closed-form** — the  $W_2$ -optimal map is

$$T_{\text{OT}} = \Sigma_0^{-1/2} \left( \Sigma_0^{1/2} \Sigma_1 \Sigma_0^{1/2} \right)^{1/2} \Sigma_0^{-1/2}.$$

**It is the unique symmetric PD solution** — of the Bures equation

$$T \Sigma_0 T^\top = \Sigma_1.$$

**That is why symmetry = OT** — symmetry of  $T_1$  singles out exactly this map.

# E • How a mission actually runs

**A DAG of molecules** — the fleet writes its own task graph, workers tackle the nodes, and a referee sub-fleet attacks the output.

**Formally pinned** — TLA+ proves no worker advances alone and a chief always decides.

**The cost** — this proof ran as 59 tasks across 78 worker sessions over hours, all reproducible from the recorded event log.

# F • What it cost

≈ **\$596** (≈ **€549**) — ≈ 548M tokens, 5,309 turns, 78 worker sessions on

`claude-opus-4-8`, summed from transcript `usage` at published rates.

Units: a *task* is one molecule; a *worker session* is one Claude Code session (more than tasks, since sub-agents add sessions). The science DAG was 22 tasks (17 plan + 5 repairs); the whole run was 59 tasks across 4 polymerisations.

**95% is cache reads** — caching turned a ≈ \$2,597 input bill into \$260, a ≈ \$2,337 saving.

**Honest sources** — `cs_ensemble` prints live cost, `events.jsonl` logs transitions; `costs.csv` is a bytes proxy, transcripts are the truth.

# G · Sister work — the same engine, other problems

**One instrument, many missions** — this proof is one of several federated works on the shared Noogram core; each a different problem, the same fleet.

**exp-families-stability** — Gaussian-convolution-stable exponential families: same notebooks + paper + Lean model as this galaxy, a sibling proof on the same rails.

**The rest is on the map** — the constellation links every live cousin; noogram.org is the index.

✓ cousin: [exp-families.noogram-labs.dev](https://exp-families.noogram-labs.dev) · live

✓ [noogram.org](https://noogram.org)