

Autonomous Scientific Director and Research-Operations Architecture

STATUS: METHODS / RESEARCH-OPERATIONS CASE — PUBLIC PORTFOLIO DEPLOYED; EXTERNAL VALIDATION PENDING

Research question

Can a human-authorized, AI-assisted research operation select among heterogeneous scientific targets, produce independently checkable work, preserve corrections and negative results, and prepare and publicly deploy scoped research packages without crossing human-only identity, legal, authorship, license, email, or scholarly-submission boundaries? The architecture treats research quality as a control problem spanning target selection, evidence, verification, reproducibility, privacy, deployment, and handoff.

Contribution and current result

The versioned director workspace records a completed 3 / 3 core objective across three distinct campaigns: Graffiti3 finite-group refutations; semiorder zonotope bounds; and the Erdős #64 excess-degree theorem package. Each package has a separate claim ledger, manuscript/supplement, reproducibility archive, checksums, verification gates, and staging record. All three are now public GitHub preprint releases at tags `v1.0.0-preprint`; external specialist review remains pending and none is journal/arXiv submitted. The public portfolio and GitHub profile are live, and Erdős #409 separately has the public pointwise certificate release `v3.0.0-f104` proving exactly $F(400000287233629)=104$, hence $\sup_n F(n) \geq 104$. The system also preserved protected states: RH remains unsolved/in progress, with version 2 rejected at N06 and version 3 stopped at its frozen dependency-schema obstruction; Study 14's confirmatory core and held-out outcomes remain unchanged; and the public Erdős #409 scope remains exactly the $F=104$ pointwise certificate. Public GitHub and Pages deployment occurred only after explicit human authorization. No email was sent, no journal or arXiv submission was made, and no legal, authorship, affiliation, funding, conflict, or license attestation was supplied on the human's behalf.

Methods

The workflow combines portfolio-level target comparison, frozen campaign contracts, source registries, append-only state, proof reconstruction from definitions, exact computation, method-distinct replay, hostile audit, mutation testing, prior-art scoping, PDF inspection, checksum and zero-mutation archive checks, correction/supersession records, private staging, and versioned handoff bundles. Heavy computation and model-worker concurrency are bounded. Scientific validity, novelty, external review, publication readiness, and public authorization are tracked separately.

Zackary's role

Zackary is the human program owner and authorization boundary. He set the portfolio objective and constraints, supplied the research corpus, and retained sole control over identity, authorship, affiliation, conflicts, funding, licensing, submission, email, and public visibility. Recorded Study 14 episodes show one initial authorization and no post-freeze human route selection; the system's autonomous choices therefore remain distinguishable from Zackary's final legal and publication decisions.

AI and tool role

Codex and other AI systems performed bounded discovery, target selection, mathematical exploration, code generation, proof reconstruction, adversarial checking, documentation, and reversible private staging. Local tools included Git, Python, GAP, C, exact data tables, LaTeX/PDF tooling, hashes, and structured YAML/JSON ledgers. Multiple AI lanes provide internal cross-checking only; they are not external peer review.

Verification

The public deployment checkpoint is inspectable at portfolio commit `d3c9fa1ed680e6ccd7741d0e2ff2f24784e94fcb`, profile commit `692fcfe7ba3dc3537f5b148b5a661dc85538281c`, Graffiti3 commit `e7b96c80f8f275f22fb0908c4cae862c9dbdc67`, semiorder commit `84fff91062d04ed7cc98f3892ea4617eb70c690d`, clean Erdős #64 commit `dace23d4837fee2491fe8942d041b0dfffe33d1aa`, and Erdős #409 commit `ee62d68d16f187b808c814108fd53881c4e3e614`. The package-specific checks validate only their declared scopes. The methods architecture itself has not been compared in a randomized study, audited by an external research group, or shown to improve outcomes causally.

Exact nonclaims

The workflow is not claimed to be autonomous science without human responsibility, a replacement for peer review, or proof that ETI outperforms baselines. Three public GitHub preprint releases are not three peer-reviewed, accepted, journal-published, or arXiv-submitted papers. Naturalistic episodes are not causal efficacy evidence. No global open problem is claimed solved, and public deployment does not supply external validation.

Public artifacts

Public-safe outputs include the [live research portfolio](#), the three dedicated preprint repositories linked there, and the [Erdős #409 \$F=104\$ release](#). The full director workspace, private prompts and transcripts, Drive receipts, private-control histories, rejected or unreleased packets, and mixed-rights archives remain excluded.

Employer relevance

The case demonstrates agent orchestration, research program management, verification pipelines, data lineage, correction handling, privacy controls, reproducible packaging, and responsible escalation to a human decision-maker. It is relevant to AI research operations, research engineering, formal verification, technical program management, and safety-critical knowledge workflows.