Daniel Aarao Reis Arturi

Research Scientist and Research Engineer

New York, NY | +1 917-655-9244 | dar579@nyu.edu | LinkedIn | GitHub | FANNAgent code

RESEARCH PROFILE

Research scientist and engineer building reproducible benchmarks, agentic systems, and experimental infrastructure for open-ended problems. Current work spans AI for mathematics, mechanistic interpretability, and LLM-guided stopping policies for filtered vector search.

Research interests AI for mathematics, agentic systems, database systems, query optimization, vector search, mechanistic interpretability, and reproducible evaluation

RESEARCH EXPERIENCE

NYU Tandon School of Engineering, VIDA Center — May 2026-Present

Research Scientist | DARPA expMath | Lea website | Code | Program

- Helped develop Lea, an open-source platform for mathematician-controlled formalization that supports proof decomposition, mid-run intervention, trajectory inspection, manual correction, and auditable Lean 4 output.
- Worked on the agentic harness and its research infrastructure, integrating project-aware execution, configurable sub-agent roles, Lean checking, typed event streams, persistent proof history, and shared state across standalone and Overleaf clients.
- Built the React/Vite/TypeScript and FastAPI research interface and developed the Overleaf-to-Lean workflow, including dependency-aware formalization, counterexample handling, self-repair, theorem-specific review, and GitHub export.
- First-authored the Lea paper accepted to the Sixth Workshop on Mathematical Reasoning and AI at NeurIPS 2026; contributed related-work synthesis, research framing, architecture documentation, case-study analysis, and technical writing.

New York University, Courant Institute of Mathematical Sciences — Sep 2026-Present

Independent Study Researcher | Filtered Vector Search | Advisor: Prof. Jinyang Li

- Developed FANNAgent, an offline LLM-guided system for workload-specific early stopping in filtered approximate nearest-neighbor search; implemented bounded policies inside pgvector's HNSW search without query-time model calls.
- Built FANNBench across PostgreSQL/pgvector, Elasticsearch/Lucene, and Weaviate, with exact filtered top-*k* references and native instrumentation separating search, feature-collection, and policy-execution costs.
- Measured 59.7-62.0% lower mean latency than native pgvector for an agent-generated policy on 300 development query-filter cases, with 95% mean Recall@10 versus 99% for the baseline.

Algoverse AI Research — Apr 2025-Present

Research Intern | AI Safety and Mechanistic Interpretability | Workshop paper | Project | Code

- Characterized shared parameter subspaces and cross-task linear mode connectivity in LoRA-adapted Qwen2.5-7B and Llama 3.1-8B models; interpolations retained misaligned behavior while preserving more than 90% coherence.
- Compared harmful and matched benign fine-tunes across data-generation conditions, showing that shared parameter geometry can reflect template provenance rather than misalignment alone; tested robustness across LoRA ranks and model sizes.
- Evaluated post-hoc negation and orthonormal projection, finding better reasoning retention with projection in the tested mixed-data settings; contributed to the full-length manuscript under review at ICLR 2027.
- Joint first author of the NeurIPS 2025 workshop paper; a separate ongoing project studies how multi-thousand-turn conversations and persistent memory affect model persona stability.

McGill University and University of Passau — Aug 2025-Mar 2026

Research Intern | Graph Database Systems | Artifact | Earlier preprint

- Adapted Dewey and PrePost structural indexes to tree-shaped property-graph data and evaluated baseline traversal and indexed variants across Apache AGE, Kuzu, and Neo4j using 100,000-node synthetic trees and LDBC SNB workloads.
- Measured 6.7x to more than 300,000x microbenchmark speedups and 1.1x to 16x gains on three LDBC queries in AGE, while documenting mixed results in native graph engines and index-maintenance costs; released a Dockerized reproducibility artifact and co-authored the vision paper accepted to the EDBT 2027 Research Track.

McGill University, Distributed Information Systems Group — Jan-Jun 2026

Research Intern | Quantum Computing for Data Systems | Paper | Artifact

- Formulated fault-tolerant distributed data allocation through binary-slack, calibrated unbalanced-penalty, and domain-wall QUBO encodings, with exact ILP and brute-force oracles for feasibility and optimality checks.
- Evaluated D-Wave simulated annealing, hybrid CQM, Advantage hardware, and QAOA while separating QPU time from end-to-end cloud overhead. The study found no quantum advantage, mapped current scaling barriers, and produced a first-author workshop paper accepted at VLDB 2026 QC&DKM and delivered as an oral presentation.

PUBLICATIONS AND MANUSCRIPTS

Lea: An Agent Backbone for Mathematician-Controlled Formalization in Lean
Daniel Aarao Reis Arturi, Shaswat Patel, Kevin Gomez, Jaume de Dios Pont, Jorge Poco, Eduardo V. Teixeira, Sam Westrick, Claudio Silva, and Chinmay Hegde. Accepted to the Sixth Workshop on Mathematical Reasoning and AI (MATH-AI), NeurIPS 2026. First author. [Website] [Code]

Open-source Lean formalization backbone shared by LeaChat and LeaOverleaf, with a case study formalizing the Durfee square generating function under a mathematician's guidance.

Seeing the Trees for the Forest: Leveraging Tree-Shaped Substructures in Property Graphs
Accepted to the EDBT 2027 Research Track. **Co-author.** [Artifact] [Earlier preprint]

Implemented and benchmarked structural indexes across Apache AGE, Kuzu, and Neo4j and built the Dockerized reproducibility artifact. Selected for the 2026 McGill Science Undergraduate Research Showcase.

Shared Parametric Geometry and Data Provenance in Model Organisms for Emergent Misalignment
2026. Full-length manuscript under review at ICLR 2027.

Extends the NeurIPS 2025 workshop work with robustness analyses, matched benign controls, data-provenance experiments, and post-hoc interventions evaluated for alignment and reasoning retention.

Shared Parameter Subspaces and Cross-Task Linearity in Emergently Misaligned Behavior
Daniel Aarao Reis Arturi*, Eric Zhang*, Andrew Ansah, Kevin Zhu, Ashwinee Panda, and Aishwarya Balwani. Accepted to NeurIPS 2025 workshops. arXiv:2511.02022, 2025. Joint first author. [Project] [Code]

* Joint first authors with equal contributions; joint first authors listed alphabetically.

Evaluating a Quantum Solution: How hard can it be?
Daniel Aarao Reis Arturi, Emilie Fahber, Stefanie Scherzinger, and Bettina Kemme. Accepted at the 2nd Workshop on Quantum Computing and Data/Knowledge Management, VLDB 2026. First author and oral presenter. [Artifact]

Experience report and reproducibility artifact covering QUBO formulation, simulators, hybrid solvers, D-Wave Advantage hardware, QAOA, classical baselines, and practical scaling limits.

SELECTED RESEARCH SYSTEMS AND STUDIES

Profiling and Optimising Energy Efficiency in GFlowNet Molecular Generation

Python, PyTorch, CUDA, CodeCarbon, Slurm [Code] [Report]

- Instrumented a ~600M-parameter GraphTransformerGFN with timing-only, CodeCarbon, and CUDA-synchronized measurement modes at less than 0.2% overhead; found molecular sampling consumed 94.4% of step time while GPU utilization remained below 13%.
- Parallel sampling with four workers improved throughput about 1.9x and reduced total energy use by approximately 47%; automated Slurm sweeps covered batch sizes 64, 128, and 256 across three seeds and all measurement modes.

NYC Housing Research

Python, FastAPI, SQLite, retrieval and evaluation [Code]

- Built a local-first cited research system over official NYC legal sources, separating keyless citation search from optional semantic indexing and evidence-cited generation while preserving source versions, provenance, explicit coverage gaps, and per-resource provider consent.
- Developed retrieval and answer-evaluation infrastructure, citation-support checks, report-bound human-review workflows, explicit provider-cost approvals, and documented gates separating software regression evidence from substantive housing-law review.

EDUCATION

New York University, Courant Institute of Mathematical Sciences | M.S. Computer Science | Fall 2026-Present
McGill University | B.Sc. Honours Software Engineering | May 2026 | GPA 3.99/4.0 | First Class Honours
Selected coursework Graduate Distributed Systems, Graduate Applied ML, Graduate Computer Networks, Sustainability for ML, Algorithms, Operating Systems, Database Systems, Probability and Statistics

AWARDS AND RECOGNITION

NeurIPS 2026 MATH-AI Workshop Paper Acceptance | First Author
NeurIPS 2025 workshops UniReps - Oral Presentation + Honorable Mention | MechInterp - Spotlight | CogInterp - Poster Acceptance
VLDB 2026 QC&DKM Workshop Paper Acceptance | First-Author Oral Presentation
McGill Science Undergraduate Research Showcase (2026) | Wing Hing Chan Scholarship in Science | Algoverse Merit Scholarship (twice) | McGill Dean's Honour List

RESEARCH METHODS AND TECHNICAL SKILLS

Methods Experimental design, benchmark design, reproducible evaluation, filtered-ANN evaluation, query optimization, performance instrumentation, exact-oracle validation, mechanistic interpretability, parameter-subspace analysis, linear mode connectivity, data-provenance analysis, intervention evaluation, QUBO, statistical analysis
Research stack Python, PyTorch, Transformers, NumPy, pandas, scikit-learn, Jupyter, CUDA, Slurm, CodeCarbon, D-Wave Ocean, Qiskit Aer
Systems Lean 4, Mathlib, FastAPI, React, TypeScript, SQLite, PostgreSQL, pgvector, Elasticsearch/Lucene, Weaviate, Docker, Git, Apache AGE, Kuzu, Neo4j