

ARJUN DESIKAN

Email arjun_desikan@berkeley.edu GitHub github.com/airjan-airlines

LinkedIn [linkedin.com/in/arjun-builds](https://www.linkedin.com/in/arjun-builds)

EDUCATION

UNIVERSITY OF CALIFORNIA, BERKELEY

B.A. Data Science · GPA 4.0 · Berkeley, CA

RESEARCH & EXPERIENCE

5 POSITIONS

AI SAFETY & AGENT HARNESS RESEARCH FELLOW

JUNE 2026 – PRESENT

Theorem.dev; Secure Program Synthesis Fellowship, Apart Research · Remote

- Built an LLM-driven specification synthesis system spanning C/ACSL and Rust/Lean 4, integrating Frama-C/WP, Alt-Ergo, Charon/Aeneas, mutation-guided refinement, executable counterexamples, and kernel-checked proofs.
- Achieved a 0.976 mean mutation kill rate versus a 0.622 baseline across 189 Lean benchmarks while reducing prover calls 11-fold (19.9 vs. 224.8) and wall time 2.4-fold (71 vs. 168 seconds) with a five-gate validation oracle.
- Closed an end-to-end verification chain on the production Rust crate `gcd 2.3.0`: proved 11/11 specifications with 0 refuted and generated a 40-theorem, zero-sorry Lean proof comprising 8 soundness goals and 32 supporting lemmas.
- Delivered 10 certified C artifacts with Frama-C/WP and Alt-Ergo, discharging up to 27/27 proof obligations; imported 35 benchmark algorithms and rejected weak specifications through property-based and TCE-filtered mutation testing.

SCIENTIFIC MACHINE LEARNING RESEARCHER

MAY 2026 – PRESENT

Centrum IntelliPhysics; Johns Hopkins University · Remote

- Built PyTorch multimodal CNN-MLP latent-fusion and DeepONet baselines for visual-to-tactile acceleration prediction; in preliminary runs, the time-domain DeepONet achieved 1.11 test RMSE, 35% below the CNN-MLP baseline's 1.72.
 - Evaluated material generalization using exclusive texture splits; used Fourier encodings to encourage oscillatory predictions.
 - Implemented a Fourier DeepONet on normalized log-FFT magnitudes to address spectral bias; achieved 0.92 preliminary test RMSE (17% below time-domain DeepONet; 46% below CNN-MLP) and explained over 80% of lower-frequency X-axis variance.
-

QUANTITATIVE FINANCE RESEARCHER & SOFTWARE ENGINEER

SUMMER 2026

MIT Computer Science & Artificial Intelligence Laboratory · Cambridge, MA

- Trained 4.9M-param financial time series transformer encoder on 24.8M observations (499 tickers, 12 features, 4,149 daily bars).
- Designed temporal/cross-asset architecture extending temporal neighborhood encoding (TNC) with graph convolution layers and forward-aligned objectives over 120-day historical correlations or learned attention; submitted to ICLR 2027.
- Benchmarked 9 representations across 21 targets, 3 horizons, and 3 seeds on 37,500 embargoed windows; at 21 days, achieved +0.110 relative-volatility and +0.090 beta rank-IC lift over strongest baselines with under 0.015 seed variation.
- Developed production features and bug-fixes for Mantis (lab software) using React, TypeScript, FastAPI, Zustand, and Docker.
- Designed full-stack swarm intelligence simulator Mantis feature, modeling evolving agent representations using aligned UMAP.

STUDENT RESEARCHER (FIRST-AUTHOR)

SUMMER 2025

Harvard OpenBio Student Research Institute · Remote

- Analyzed GISTIC copy-number and RNA-seq data (3,527 TCGA tumors, 12 cancer types) using PCA, ANOVA, MLR, and PLS.
- Directly associated SCNA burden with transcriptomic effective rank across quartiles (1000+ bootstrap resamples); PLS linked intracellular transport and immune response to burden with R-squared 0.883; abstract published in STEM Fellowship Journal.

DATA SCIENCE INTERN

SUMMER 2025

Fort Bend County Health Department · Missouri City, TX

- Produced 2021 statistical mortality brief over county population and 5 demographic subgroups; used by county to inform policy.
- Built an end-to-end SQL/R mortality pipeline for age-adjusted rates, t-tests, chi-square and K-S tests, and ggplot2 visualizations.
- Forecast influenza, RSV, and COVID-19 cases (XGBoost, Prophet, LSTM) reducing RMSE by 50% versus ARIMA baseline.

PROJECTS

10 PROJECTS

INTER-AGENT MISALIGNMENT: PROBING & ACTIVATION STEERING

OCTOBER 2026 – PRESENT

Python, PyTorch, Qwen3

- Built a two-agent harness testing whether a misaligned coding sub-agent pulls its conductor LLM toward reward hacking; diff-of-means activation steering on Qwen3-8B moved shortcut adoption from 22% to 3% or 75% at full coherence.
- Linear directions read corner-cutting (0.935 AUROC), but same-wording controls showed they encode content and negation, not intent (reversed AUROC 0.12–0.39); building a GRPO + LoRA post-training environment to induce reward hacking.

AGENT LEDGER: SHARED MEMORY FOR AI AGENTS

OCTOBER 2026

Next.js, TypeScript, Supabase/Postgres, pgvector, MCP, Playwright

- Team benchmarked 24 tasks/10 sites: naive retrieval cut success from 21% to 13%. Calibrated 384-d MiniLM/pgvector with a strict similarity floor and 0.7 similarity/0.3 confidence ranking, retrieving lessons in 10/14 failure cases and none for controls.
- Built a Next.js run inspector with trace replay, baseline-vs-memory view, and keyless test backend; orchestrated a single-task live-site pilot where one lesson raised same-model success from 2/4 to 6/6, cutting failed actions 69% and task time 57%.
- Live-tested MCP Streamable HTTP with Claude agents; contributed to the Playwright loop and WebArena-verified scoring, using Wilson lower bounds for confidence and Postgres constraints/RLS for idempotent outcome recording.

GITXL: VERSION CONTROL FOR EXCEL

OCTOBER 2026

Electron, JavaScript, Git, OOXML

- Built a 5,300-line JavaScript/Electron app bringing Git to Excel: parses OOXML into cell-level diffs for inputs, formulas, recalculated values, and formatting, with commit/history/restore/revert/branch/stash/push/pull workflows.
- Engineered an OOXML normalizer to suppress save-only diffs: hashes styles by font, fill, border, and number format; expands shared formulas; resolves relationship IDs; and strips volatile metadata and chart caches so unchanged re-saves produce no diff.
- Implemented cell-keyed three-way Git merges with LCS/diff3 for ordered content; auto-merges disjoint edits, ignores formula-cache churn, and safely rejects unsupported files; 39 tests cover Git/workbook cases.

MONTE CARLO ICU RISK MODEL

OCTOBER 2026

Python, NumPy, pandas

- Modeled daily ICU states with an absorbing Markov chain fit to 14,000+ observed transitions across 186 hospitals; derived closed-form mortality and length-of-stay estimates, matching observed mortality (8.32% vs. 8.31%).
 - Built a vectorized Monte Carlo engine for VaR₉₉/expected shortfall; control variates cut estimator variance 2.6x, common random numbers cut scenario variance 2.5–3.7x, and ignoring parameter uncertainty understated VaR₉₉ by 12% at n=50.
 - Compared 7 models with patient-level walk-forward/cross-validation and Bayesian shrinkage, reducing out-of-sample log-loss 19% vs. no-change baseline; a bootstrap likelihood-ratio test rejected one-day memory (p=0.002).
-

GLITCH MARKET

SEPTEMBER 2026

Solidity, React, Foundry, IPFS, wagmi/viem

- Built a Solidity/React marketplace for speedrun glitches on Base Sepolia testnet with wagmi/viem wallet integration.
- Implemented AES-256-GCM/IPFS content delivery with on-chain commitments, time- and sales-based price decay, optimistic settlement, reputation stakes, and decay-gated jury disputes.
- Wrote 34 Foundry tests and live end-to-end checks covering pricing, escrow settlement, bonded disputes, and juror eligibility.

ECG CLASSIFICATION WITH A 1D CNN

APRIL 2026

Python, PyTorch

- Trained a PyTorch 1D CNN for ECG abnormality classification, achieving an AUC of 0.98 in first-pass modeling.
- Used Grad-CAM to explain ECG predictions clinically, identifying decision-driving ST-segment and QRS complex regions.
- Optimized model performance with Optuna hyperparameter search over dropout, batch sizes, and learning rate scheduling.

EQUITIES BACKTESTING DASHBOARD

FEBRUARY 2026

Python, Streamlit, pandas, NumPy

- Built backtester for EMA crossover and mean reversion; customizable transaction costs, EMA horizons, position sizing, etc.
- Integrated Monte Carlo simulations (block bootstrap, parametric) for risk analysis and stress-testing across any input ticker.

SKATE-TRACKER

NOVEMBER 2025 – DECEMBER 2025

Python, FastAPI, React Native, PostgreSQL

- Built a React Native app and website for skateboarding progression tracking and text-based coaching, with 40+ interconnected tricks visualized in a directed prerequisite graph.
- Implemented a RAG pipeline grounded in skateboarding trick advice from 50+ wiki webpages scraped with BeautifulSoup.
- Prototyped Google MediaPipe/VLM video analysis comparing uploaded tricks to professional examples at 5 technique keyframes.

SETOS.AI STARTUP PROTOTYPE & PITCH

JUNE 2025 – NOVEMBER 2025

Python, FastAPI, React, PostgreSQL

- Led a 3-person team building personalized learning pathways across 1,000,000+ research papers from natural language queries.
 - Designed a content-based recommender with SciBERT embeddings, pgvector similarity search, LLM-based query expansion, and the kneedle algorithm to auto-tune pathway length.
-

FORGEQB

FEBRUARY 2025 – PRESENT

Electron, Flask, SQLite, FastAPI, PostgreSQL

- Lead 4-dev team building an adaptive quizbowl study app (500+ users) across Electron, FastAPI, and PostgreSQL stack.
- Engineered NLP ingestion, topic clustering and production epsilon-greedy softmax recommender (160,000+ questions).
- Prototyped RL recommender (DQN) against production softmax system using simulated student across varied reward structures.

WRITING

FORGEQB: RECOMMENDING QUIZBOWL QUESTIONS

OCTOBER 2026

How we serve question recommendations for a trivia vertical where both timing and accuracy matter, when every question comes with only a broad category and a broad difficulty.

THOUGHTS ON AMERICA'S 250TH

SPRING 2026

An exploration of my relationship with the USA and its founding ideals.

DREAMS OR DELUSIONS

SPRING 2025

An exploration asking whether naive dreaming serves us better than uncomfortable truth.

SKILLS

LANGUAGES/TOOLS	Python, Java, R, SQL, JavaScript, TypeScript, Solidity, Git, Docker, Property-Based/Mutation Testing
MACHINE LEARNING	PyTorch, scikit-learn, XGBoost, CNNs, Transformers, PINNs, Neural Operators, GNNs, RL, LLMs, TNC
DATA/STATISTICS	pandas, NumPy, SciPy, Matplotlib, Optuna, Linear Regression, PCA, PLS, Monte Carlo, Block Bootstrap
SOFTWARE STACK	React, React Native, Zustand, FastAPI, Flask, Electron, Streamlit, PostgreSQL, SQLite, pgvector, Foundry
EVALUATION/TESTING	Linear Probing, Activation Patching, Grad-CAM, UMAP, Backtesting, t/Chi-Square/K-S/ADF Tests
HOBBIES/INTERESTS	Racquet sports, badminton, thrifting, fashion, art history, classic literature, philosophy