

Neil Tripathi

San Francisco, CA | (929) 476-5255 | neiltripathi93@gmail.com
github.com/neilt93 | neiltripathi.com | linkedin.com/in/neil-tripathi-001b50300

Research engineer and founder building evaluation infrastructure, simulation benchmarks, and sim-to-real systems for embodied AI.

EDUCATION

New York University

B.A. Computer Science, Minor: Mathematics, GPA: 3.77/4.00

– Dean's List (2023–2024, 2024–2025). Coursework: ML, AI, NLP, algorithms, operating systems, cryptography, databases.

St Olave's Grammar School

A levels: Mathematics, Further Mathematics, Computer Science, Physics (A* in each)

Sep 2022 – Dec 2025

New York, NY

Sep 2016 – Jun 2022

London, UK

RESEARCH & PUBLICATIONS

Benchmarking Commonsense Visual Reasoning for Vision-Language Models

2025 – 2026

Co-author with Prof. Ernest Davis (NYU); arXiv cs.AI, Mar 2026

- Designed a visibility-reasoning benchmark (100 families, 300 headline cells, 2×2 XOR design) with automated graders and scoring.
- Evaluated 9 VLMs across three tiers (flagship, prior-gen, open-source); surfaced a 26% abstention rate in GPT-5 as key finding.

Adversarial Robustness of 3D U-Net Radiation Therapy Dose Prediction

2025

Co-author w/ Prof. Birjoo Vaishnav (U. Maryland) & Prof. Weixian Liao (Towson); SERA 2026 (IEEE/ACIS), submitted; ASTRO 2026 submission under review

- Evaluated adversarial (FGSM/PGD) and clinically plausible CT perturbations on the OpenKBP dataset using DVH metrics; identified resolution downsampling as the only perturbation with immediate clinical impact.

EXPERIENCE

Cybernetic Labs

Co-Founder

Mar 2026 – Present

San Francisco, CA

- Designed and shipped a VLM-as-Judge evaluation system with dual-judge mode and Cohen's kappa agreement metrics (86/86 tests passing).
- Trained and evaluated a fine-tuned SmolVLA vision-language-action policy on the SO-101 arm for pick-and-place and cube-stacking tasks.
- Built PackingBench, a benchmark for evaluating embodied / VLA manipulation policies on warehouse packing tasks.
- Conducting mechanistic interpretability research on vision-language-action (VLA) policies to probe and explain model internals.
- Extended the sim-to-real pipeline with 2-camera SmolVLA deployment, joint-space policy transfer, and a batch sim episode generator.
- Contributed to the FastAPI/SQLAlchemy backend and Vite/React evaluation dashboard; collaborated with the CEO on the pre-seed investor deck and technical roadmap.

Sentient Labs

AI Product & Research Intern

May 2025 – Aug 2025

New York City, NY (Hybrid)

- Built an internal evaluation suite of 100+ LLM prompts with automated scoring for tool-call correctness and safety-violation rate.
- Improved feedback-to-ticket latency by 20% via regression harness and iteration notes; shipped 3 tool integrations with engineers.
- Built a real-time crypto signals dashboard (2–3k lines backend TS, 1k lines frontend React) with latency under 2s for exchange data.
- Prototyped 'Dobby Mode' agent (300-line Streamlit demo) using Dolphin Mistral; enforced JSON schema to reduce hallucinations.

New York University

Teaching Assistant, Database Design

Jan 2025 – May 2025

New York City, NY

- Led recitations on SQL, MongoDB, Flask/Jinja, and schema design; supported 50+ students via office hours and grading.

4mainstreet.ai

ML and Full-Stack Intern

May 2024 – Aug 2024

Fremont, CA (Remote)

- Built data pipelines for customer voice + text (4,000 records); integrated outputs into a Flutter front-end.
- Built an evaluation harness to benchmark Llama 3 8B, Gemini, and GPT-class models on 4,000 labeled examples; computed precision/recall/F1 and surfaced failure modes to guide model selection.

Scroll.io

Software Engineering Intern (Blockchain and Data)

May 2023 – Aug 2023

New York City, NY (Remote)

- Built a Discord analytics tool for a 20K+ community; automated ingestion into BigQuery and cut manual reporting by 5 hours/week.

PROJECTS

Physical.AI: Data-Driven Marketplace for Physical Ads

2025

Co-Founder

San Francisco, CA

- Built Zillow-style discovery experience for physical ads on a map; achieved top 6 at YC W25 hackathon finals out of 100+ teams.

Connectome-Driven Locomotion Controller

Mar 2026 – Present

Independent Research

- Simulated the MANC ventral nerve cord (13,101 neurons, 1.9M synapses) in Brian2 with FlyGym v2; built a proprioceptive encoder from BANC sensory pathways and a structural descent optimiser for connectome-guided gradient projection.
- Demonstrated emergent odor avoidance from connectome wiring alone (no engineered navigation logic) and discovered a 2-neuron hygrosensory bottleneck (DNb05).
- Deploying to a Waveshare HexArth hexapod (18-DOF Dynamixel servos) with serial protocol bridge for closed-loop sim-to-real transfer.

Prediction Market Forecasting Agent

Mar 2026 – Present

Independent Project

- Built a local-first forecasting agent with LLM-driven sub-question decomposition, devil's advocate debate, and LightGBM calibration for automated Kalshi trading.
- Backtested across 229 markets with Qwen 14B: 80% directional accuracy, Brier score 0.0033; dynamic ECE thresholds for position sizing.

TECHNICAL SKILLS

Languages

- Python, TypeScript, SQL, Java, Rust, C

Machine Learning

- PyTorch, TensorFlow, MuJoCo, FlyGym, Brian2, diffusion models, world models, PPO, LLM fine-tuning, evaluation pipelines, agent tool use

Data & Cloud

- Postgres, MongoDB, BigQuery, Supabase, AWS, RunPod, Docker, Next.js, FastAPI, Vercel

LANGUAGES

- English (fluent), Hindi (fluent), Spanish (intermediate), French (basic)

OTHER INTERESTS

- Competitive gaming (usually top 2,000 globally in *Clash Royale*, Grand Champion III in *Rocket League*, peaked at rank 598 in *Marvel Rivals*), the gym, and sports like football, rugby, badminton, table tennis, and fives.