

Neil Tripathi

San Francisco, CA | (929) 476-5255 | neiltripathi93@gmail.com
github.com/neilt93 | website-mauve-nine-60.vercel.app | linkedin.com/in/neil-tripathi-001b50300

EDUCATION

New York University

Sep 2022 – Dec 2025

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

New York, NY

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

St Olave's Grammar School

Sep 2016 – Jun 2022

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

London, UK

RESEARCH & PUBLICATIONS

Benchmarking Commonsense Visual Reasoning for Vision-Language Models

2025

Co-author with Prof. Ernest Davis (NYU); manuscript in preparation

- Built a diagnostic benchmark (100 base + 100 counterfactual images; 100 questions + 100 flips) and automatic graders.
- Evaluated six VLMs (ChatGPT, Claude, LLaVA, etc.) and analysed hallucination/abstention behavior.

Legal Text Classification with BERT and LegalBERT

2024

Co-author; Natural Language Processing course project (NYU)

- Fine-tuned BERT-family models on legal corpora; compared against classical baselines with reproducible training and reporting.

EXPERIENCE

Sentient Labs

May 2025 – Aug 2025

AI Product & Research Intern

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

Jan 2025 – May 2025

Teaching Assistant, Database Design

New York City, NY

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

4mainstreet.ai

May 2024 – Aug 2024

ML and Full-Stack Intern

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

May 2023 – Aug 2023

Software Engineering Intern (Blockchain and Data)

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

Jan 2025 – Present

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.
- Integrated live *attention score* using Caltrans PeMS traffic data to estimate OTS volume, dwell time, and impact by time of day.
- Backend ingests ad placements, maps to nearby sensors, computes metrics against historical baselines, and surfaces ticker-style scores.

Sympli, Voice-First Health Companion App

Aug 2025 – Present

Technical Co-Founder

New York, NY

- Built voice intake → clinician-style note pipeline (ASR + summarization) with latency under 1 minute; added end-to-end instrumentation.
- Deployed HIPAA-aware AWS stack (TypeScript + Postgres + self-hosted LLM); shipped an alpha to 20 users (100+ sessions).

1:72 Scale HVAC Assembly Part Counter

Jun 2025 – Present

Computer Vision Developer (Contract)

New York City, NY

- Trained an object detector across 14 part classes; achieved 95% accuracy on a 1,200-image test set; delivered an iOS app for part counting.
- Improved robustness to lighting variance and motion blur; fixed 12 failure modes and optimised latency to under 5s on iPhone 16 Pro.

Transformer Language Model from Scratch

2025

Personal Project

- Implemented a Transformer LM in PyTorch; trained on an anonymised corpus (2.6M tokens) and reached validation perplexity 15.

TAP-Score (Push-T OOD Detection)

Jan 2026 – Present

Personal Project

- Replicated Diffusion Policy on Push-T; built TAP-Score to rate (obs, action-chunk) pairs using demo positives + shuffled/noisy negatives.
- Tested OOD detection under brightness/noise/occlusion; reporting AUROC and clean false-alarm rate with score traces.

TECHNICAL SKILLS

Languages

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

Machine Learning

- PyTorch, TensorFlow, object detection, data augmentation, LLM fine-tuning, evaluation pipelines, prompt design, agent tool use

Data & Cloud

- Postgres, MongoDB, BigQuery, Supabase, AWS, Docker, REST APIs

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.