David Huu Pham

Canadian Citizen | TN1 and US J-1 Eligible
Website: dhpham.com
LinkedIn: dhpham-software
Email: davidpham.tech@gmail.com
GitHub: dayvidpham

Education

Master of Science, Computing Science — Artificial Intelligence May 2025 – Present
Simon Fraser University *Burnaby, BC*
- Thesis: modeling data-consent ecosystems for data-related collective action using large-scale agent-based simulation

Bachelor of Science, Computer Science & Statistics Double Major — AI Concentration Jun 2024
Simon Fraser University *Burnaby, BC*

Publications

D. H. Pham, et al. *What if We All Just Submit the Transcripts? Provenance and Auditability for AI-Native Research Processes*. NeurIPS 2026 Workshop on AI-Native Academia: Authorship, Peer Review, and Conference Governance under AI.

P. Zhao, **D. H. Pham** (co-first author), N. Vincent. *Mechanism Plausibility in Generative Agent-Based Modeling*. FAccT 2026.

J. Tan, N. Vincent, K. Elkins, M. Sahlgren, J. Low, **D. Pham**, et al. *If open source is to win, it must go public*. ICML 2026.

Experience

Founder Jan 2026 – Present
Peasant Labs *Vancouver, BC*
- Created **open-source, self-hostable** cloud platform on Railway for users to **redact and share** Claude Code, Codex, and Cursor transcripts, enabling **transparency and observability** of agentic engineering and open research processes
- Built a **multithreaded ETL pipeline** that ingests, indexes, and analyzes **10+ GB** of agent sessions into local SQLite databases in **5 minutes**, several times faster than other solutions

Research Assistant Sept 2024 – Apr 2025
Dr. Nicholas Vincent, PaDComp Research Group, Simon Fraser University *Burnaby, BC*
- Built massively parallel, fully-reproducible Python simulation of users in social media platform, making it **1,000x faster** and able to model decisions of **160M users** by using NumPy, Polars, GPU-accelerated cuDF, Ray, Apache Arrow, and Apache Parquet
- Implemented supercomputer integration to allow **efficient scaling** of simulation runs as Slurm jobs, running on Canada's **fastest academic supercomputer** Fir, fully utilizing an entire node with 192 CPU cores and 768 GiB RAM

Software Engineer Intern Sept 2023 – Apr 2024
Mastercard *Vancouver, BC*
- Built a Java CRUD API infront of Cassandra DB, reaching **90%+ test coverage** with integration and unit tests
- Containerized legacy monolith, deploying the **first AWS ECS microservice in the org** using TypeScript CDK and CloudFormation for maintainable infra; complete with Jenkins CI/CD pipeline for automated test, image build, and ECS deploys

Full-Stack Engineer Intern Jan 2021 – Aug 2021
Solaires Entreprises Inc. *Victoria, BC*
- Architected and built **entire full-stack distributed application** with an Node.js/TypeScript backend API and web dashboard with Google OAuth2 login, letting users control IoT devices over MQTT, storing events and sensor history with PostgreSQL
- Built **company's first** GitLab CI/CD pipeline to allow maintainable and fast agile development while retaining stability, complete with an automated test suite and Docker images standardizing dev, test, and prod envs

Software Head Developer Aug 2024 – Present
SFU Robot Soccer Club *Burnaby, BC*
- Led C++ development of team's **robot control software**, integrating BehaviourTree library with custom event-driven architecture for **easily configurable** strategies, and designed **conflict-free, multi-agent** path planning using D* Lite and CBS
- Implemented new MuJoCo C++ simulator for **faster, realistic physics**, used in Python **RL training pipeline** through pybind11

Projects

TraceBench
AI benchmark to measure effect of past coding agent transcripts on coding performance within a given repo, across three harnesses

Technical Skills

Languages: Python, R, SQL, C/C++, Java, TypeScript, MATLAB, LaTeX
ML & Statistics: PyTorch, scikit-learn, NumPy, Apache Spark, Apache Hadoop, tidyverse, ggplot2, matplotlib
Simulation & Robotics: Isaac Lab, RSL-RL, MuJoCo, grSim, Gymnasium
Engineering: Docker, Apptainer, AWS (EC2, ECS), Nix/NixOS, GitLab CI/CD, SQLite, PostgreSQL
Research: scaling laws, agent-based modeling, recommender systems, computer vision, statistics