

James Han

jameshan.cs@gmail.com | 1-647-570-1883 | Canadian Citizen | J-1 Visa Eligible
linkedin.com/in/jameshan27 | github.com/lxyhan | jameshan.net

EDUCATION

University of Toronto

Toronto, ON

B.Sc. Computer Science and Statistics, Focus in Machine Learning | In-Major GPA: 3.8

Sep. 2023 – May 2027

- UofT/Trinity College Academic Grants (\$6500), Leonard McLaughlin Scholarship, RBC Tech Scholars Award (\$25400)
- Linear Algebra (100%, 1st/350), Data Structures and Algorithms (A+), Computational Biology (A+)

EXPERIENCE

Shopify

Toronto, ON

Incoming Engineering Intern

Mozilla Firefox

Toronto, ON

Machine Learning Engineering Intern, Privacy

Oct 2025 – Present

- Published research and deploying **ONNX model** doing **live per-request inference** in Firefox’s anti-tracking engine serving **250M+ users**, predicting transfer size of blocked trackers at request-block time. XGBoost trained on **348M requests** across **4,592 domains**, reducing MAE **20.9%** over lookup baselines (**46.4%** on scripts).
- Diagnosed and patched **2 high-priority security vulnerabilities** in Gecko’s **C++ engine**, reverse-engineering **Rice decoding, I/O serialization** through IPDL protocol definitions and content process sandboxing. Traced bit-level data propagation across IPC boundaries to isolate exposure vectors; fixes shipped to **500M+ daily sessions**.
- Designed and ran **full-scale model selection experiment**: benchmarked **7 architectures** (XGBoost, LightGBM, ridge, quantile regression, conformal prediction) on **3.5M requests** with **domain-level train/test splits** preventing data leakage. Conducted **feature ablation studies** across **19 engineered features**, applied **SHAP** for interpretability, and used **Optuna** for hyperparameter search to select production model shipped to **30M+ daily active users**.

Anthropic

Toronto, ON

Claude Ambassador

Sep 2025 – Present

- Built and delivered technical workshops on **RAG pipelines, LLM fine-tuning, and model evaluation frameworks** to **1500+ students**; led AI hackathon (**200+ developers**) with live demos of **semantic search** and **retrieval** systems.

RESEARCH, PROJECTS, AND AWARDS

Predicting the Cost of Blocked Tracker Requests from Pre-Response Features | *Tim Huang (Sr. Staff Engineer)*

- First-authored paper on **counterfactual cost prediction** for blocked tracker requests in Firefox (**250M DAU**). Trained on **3.49M requests** across **3,723 domains**; benchmarked **XGBoost, LightGBM, quantile regression, conformal prediction** with **Tweedie loss, CNN-learned URL embeddings, and 80-dim feature ablation**. Best model reduces MAE **47.5%** over lookup baselines, outperforms exact-path memorization, and generalizes on **3-month temporal holdout (6.0% median error weekly)**. Deployed as **ONNX** for **live per-request inference** in Firefox’s blocking path.

Shopify Merchant Intent Classifier | *PyTorch, Hugging Face, PEFT, LoRA, scikit-learn*

- Fine-tuned DistilBERT on synthetic merchant query dataset for real-time intent classification across billing, fulfillment, and catalog domains; applied semantic embeddings for query rewriting, PEFT/LoRA for efficient fine-tuning, and per-class F1 evaluation against zero-shot baseline.

1st Place Overall, Google x Hack the Future | *XGBoost, PyTorch* | 1st/260 teams, presented at Google HQ

1st Place Overall, UofTHacks | *PyTorch, ONNX, OpenCV* | 1st/330 hackers

2x Winner, Hack the North | *OpenCV, LangChain, PyTorch*; Finalist, AWS ML Case Competition | *Bedrock*

ATHLETICS AND LEADERSHIP

Triathlon Canada National Team - Incoming, representing Canada at the **2027 World Championships**

Vice President, UofT Blueprint - Led 4 engineering teams building pro-bono software for non-profits

Director, UofT AI - Organizing **ML workshops**, research talks, and hackathons for **2,350+ students**

TECHNICAL SKILLS

Languages: Python, Rust, C++, Bash, Typescript

ML/DL Frameworks: PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, XGBoost, LangChain

ML Infrastructure: MLflow, W&B, ONNX, model evaluation pipelines, A/B testing, Optuna, conformal prediction

Data & Experimentation: Pandas, NumPy, Spark, SHAP, feature engineering, hypothesis testing