

# Edward Tran

Vancouver | giaphutran012@gmail.com | [Linkedin.com/in/edwardtran123](https://www.linkedin.com/in/edwardtran123) | [Github.com/giaphutran12](https://github.com/giaphutran12) | [Giaphutran12.github.io/Online-portfolio2/](https://giaphutran12.github.io/Online-portfolio2/)

## EDUCATION

### Douglas College

**Expected Graduation: Dec 2025**

Diploma in Computer Science | CS SFU Transfer @2026

## SKILLS

**Languages** **Proficient** JavaScript(2yrs) · C++(3yrs) **Intermediate** Python(3yrs) · Java(1yr) **Beginner** SQL(0.5yr) · PHP(0.5yr)

**Software** NextJS · Tailwind · TypeScript · Postgres · Prisma ORM · Bootstrap · Git · Unreal Engine 5 ·

## EXPERIENCE

### Software Engineer Resident

**July 2025 - Present**

Headstarter

*Vancouver, BC, Canada*

- Built 3+ AI engineering and fullstack projects in a fast-paced software team environment
- Developed 3+ apps in TypeScript and NextJS on Vercel with production and dev environment
- Implemented llm-chaining, hyperparameter tuning, fine tuning on 5+ LLM models controlling for latency & accuracy
- Coached by Google Machine Learning, Google Kubernetes, Two Sigma, Tesla, Figma and Citadel Engineers
- Created 120+ commits on github with 7-day deadlines getting a Career Capital increase of 40% from start date

### Software Engineer Intern

**Apr 2024 - Dec 2024**

Douglas College

*New Westminster, BC, Canada*

- Built and maintained White Board ITS, a **work day management web app** used by 100+ IT staffs in Douglas College
- Deployed to production 10+ new features, reducing time spent in delegating organization tasks by **75%** for 100+ daily users
- Provided real-time information/announcement to 100+ staff in Douglas College's IT department, reducing time spent communicating and figuring out what to do for the day by 35%
- Leveraged PHP and AJAX calls to MySQL to provide accurate and secure user's CRUD, implemented IP-based secured log in

## PROJECTS

### AI Customer Support Platform | Open Source (~60 hours) - [GitHub Link](#)

**August 2025**

- Built an AI using NextJS to provide **real-time** customer support for businesses via text, chat, and calls, with escalating option
- Monorepo, built 3 separated apps with shared packages, saved 30% of computing resources
- Created end-to-end error-tracking dashboard using Sentry + Convex for real-time alerts, replay session, and end-to-end trace logs, ensuring full information during development and production, reduced error tracking time by 36%
- Increased business confidence by 40% by allowing agents **White Labelling** using AWS Secrets Manager and BYOK

### AI Website Generator | Published Web app & Open Source (~50 hours) - [GitHub Link](#) - [Website Link](#)

**July 2025**

- Used Inngest, E2B, Prisma, Postgres Neon, NextJS to create a context-aware AI website generator, handling up to **1000 daily active users** and 50,000 projects, averaging 10ms per network request, process up to 200 concurrent background jobs
- Integrated TRPC and TanStack React Query for end-to-end type safety and 3-layer security guard
- Leveraged effective prompt engineering combined with Anthropic API to generate code with 32% more accuracy

### AI Customer Support Agent | Open Source (~30 hours) - [GitHub Link](#)

**July 2025**

- Used Firecrawl, Pinecone, VAPI, Gemini to build a context-aware AI customer support for a fintech company
- Used Firecrawl API to dynamically scrape and preprocess data, increasing data effectiveness by 55%
- Leveraged Gemini Embedding and Pinecone to vectorize scraped data, reduced resource use by 80% with semantic search
- Built a voice-enabled webcall experience using VAPI for real-time speech-to-text and text-to-speech interaction

### AI LLM Router | Open Source (~30 hours) - [GitHub Link](#)

**Aug 2025**

- Built a web app similar to ChatGPT which **routed user's prompt** to the best AI model based on cost, latency, response quality
- Used gpt-oss-20b to categorize the the user prompt into 5+ categories, then routed to the best-fit LLM (GPT5, Claude 4.1, Gwen Coder 3, Gemini Flash 2.5) for the task, reducing cost by 68%, latency by 35%, and response quality by 25%

## Activities

Medium Blogs: [How to build an AI website generator](#), [GPT-5: Discussing AI Capabilities](#), [Cursor is great!](#), [How to stand out in the AI era](#)