Insoo Son

ON, Canada
github.com/insooeric | linkedin.com/in/insooson | insooeric.son@gmail.com | +1 (226) 224-3028 | Portfolio

Education

Western University — London, ON
Bachelor's Degree in Computer Science — *Apr 2028 (Expected)*

Fanshawe College — London, ON
Advanced Diploma in Computer Programming and Analysis — *Sep 2021 – Dec 2024*

- Dean's Honor Roll **- 3.97 GPA**

Technical Skills

Languages: Python, HTML, CSS, Java, JavaScript, SASS, TypeScript, C, C++, C#
Frameworks: .NET Core, ReactJS, Vue.js, Node.js, Express.js, REST APIs, Pandas, PyTorch, NumPy, LoRA, matplotlib, Redis, MongoDB, MySQL
Developer Tools: AWS, Git, GitHub, Docker, GCP, Visual Studio, Jira, Linux, Vercel, OnRender, Firebase

Experience

AI & Backend Software Engineer Intern — Jan 2026 – Apr 2026 (Internship)
SpotOn – Startup Company — *Remote (USA)*

- Collaborate with backend engineers to **integrate LLM-assisted features** into server-side systems (prompting, integrations, dataset processing, reliability checks).
- Develop dataset pipelines ingesting **3+ backend APIs** into model-ready data with logging, validation, and failure-safe error handling.

Extracurricular

Misinformation AI Developer — Sep 2025 – Present
Western AI Club — *Western University, London, ON*

- Processed **5 datasets (300,000 rows)** into a unified 3-label format; established **80/10/10** splits.
- Authored Python preprocessing scripts to convert raw sources into training-ready datasets (IDs + masks).
- Implemented a **RoBERTa** pipeline to classify **3 labels** and retrieve **5 supporting references** per claim.

Projects

Custom Large Language Model | *Python, Jupyter Notebooks, PyTorch, LLM* | Code — Sep 2025

- Designed a 136M-parameter text-generation model (layers=12, heads=12, vocab=60k, ctx=512) in PyTorch.
- Achieved **80.5% faster** training steps and **45% higher throughput** by tuning optimizers and compute settings, while reducing training perplexity.
- Applied LoRA to update 4.7M parameters (3.46% of total), cutting memory usage by 40%.
- Developed a streaming data loader to train on **2.12B tokens** under limited-memory constraints.

Number Recognition | *Python, Jupyter Notebooks, Multi-Layer Perceptron, HDF5* | Code — May 2025

- Implemented an MLP digit recognizer from scratch (**575k params; 784→512→256→128→64→10**) without TensorFlow.
- Trained using a Kaggle digit dataset, MNIST, (**48,000 training images; 28,000 test images**) and exported the trained model to HDF5.

Gomoku AI | *ReactJS, .NET Core, Docker, JavaScript, C#, API design* | Code — Jan 2025

- Developed a Gomoku AI web app supporting **2 rule sets**, win detection, and robust invalid-move handling.
- Applied advanced game-search algorithms **Monte Carlo Tree Search and Minimax with Alpha-Beta pruning** to let the AI plan several moves ahead and provide challenging gameplay for human users.
- MCTS evaluation: 30 AI-vs-player games (**2 losses**) and 100 AI-vs-AI games (**97% draw rate**).

Stemma | *ReactJS, .NET Core, Docker, Redis, TypeScript, C#, data validation* | Code — Apr 2025

- Engineered a web service that generates live GitHub-stat badges for developers.
- Introduced secure login with session management to protect user accounts and data.
- Optimized API performance using parallelization, **cutting latency by 50%** and allowing thousands of badge renders per day under load.
- Stored images on Google Cloud Platform for persistent hosting and automatic updates.
- Deployed a containerized service with 12 protected endpoints and production error handling (**0% backend errors observed so far**).