

Paul He

☎ 416 659 5646

✉ paul.he@mail.utoronto.ca

🌐 [linkedin](#)

🔄 GitHub

🌐 Academic Website

Education

University of Toronto

Honours Bachelor of Science in Computer Science

September 2021 - May 2025

Toronto, Canada

- **Relevant Coursework:** Introduction to Machine Learning, Computational Linguistics

Swiss Federal Institute of Technology (ETH Zürich)

Exchange Year

September 2023 - August 2024

Zürich, Switzerland

- EESTEC Generative AI hackathon winner
- **Relevant Coursework:** Natural Language Processing, Probabilistic Artificial Intelligence, Machine Perception, Large Language Models, Advanced Formal Language Theory

Publications

Pointwise Mutual Information as a Performance Gauge for Retrieval-Augmented Generation

Tianyu Liu*, Jirui Qi*, **Paul He**[†], Arianna Bisazza, Mrinmaya Sachan, Ryan Cotterell

NAACL 2025 Main

Experience

Student Researcher – Causal Formalization and Multilingual LLMs

Jan 2025 – Present

Supervisor: Prof. Dr. Zhijing Jin

Institution: Vector Institute

- Developing a **novel evaluation framework** to assess LLMs' ability to formalize mathematical expressions in the causal domain.
- Enhancing interpretability of multilingual LLMs by **modifying attention layers**, applying causal mediation analysis, and experimenting with perturbation techniques.

Student Researcher – Computational Linguistics

Sep 2024 – Present

Supervisor: Prof. Dr. Gerald Penn **Institution:** University of Toronto, Computational Linguistics Group

- Developed the **first public combinatory categorial grammar (CCG) parser**, tested on the Parallel Meaning Bank.
- Investigating self-attention limitations in transformers and **designing inference-time optimizations** to address them.

Student Researcher – Large Language Models

Jun 2024 – Sep 2024

Supervisors: Tianyu Liu, Prof. Dr. Ryan Cotterell **Institution:** Rycolab, ETH Zürich

- Explored attribution methods using Jensen-Shannon divergence and Fisher information to approximate LLM attention focus.
- Designed and implemented a novel mechanism which **improved open-source LLM accuracy by 3%** on NQ-Open and ELI5 QA tasks.

Projects

Word Segmentation and Part-of-Speech Tagging with Transformer | Python, PyTorch

October 2023 – January 2024

- Designed a **new decoder embedding mechanism**, allowing the decoder to better capture subtle linguistic nuances and dependencies between characters and tags.
- Implemented a **transformer** model using **PyTorch** with beam-search for generation.
- Achieved leading-edge results with validation accuracy of **96.29%** for CWS, **93.37%** for POS tagging, and **91.99%** for joint CWS and POS tagging after just 10 epochs.

Mental Health Support Bot | Python, BeautifulSoup, FastAPI

October 2023

- Led the development of a simple mental health chatbot web app tailored for Computer Science students, assisting team members with debugging.
- Performed **web scraping** to obtain data for **few-shot prompting**.
- Implemented a real-time chat feature using FastAPI to host the chat platform and model on a server, allowing multiple simultaneous chats with the bot and recall of previous conversations.

Technical Skills

Programming Languages: Python, C, Java, R, SQL

Frameworks/Libraries: PyTorch, scikit-learn, NumPy, FastAPI, Flask

Tools/Platforms: Git, Docker, Jupyter, GitHub

Databases: PostgreSQL, psycopg2