

# Haolin Ye

<https://github.com/HaoolinYe/>

Email : [haolin.ye@mail.mcgill.ca](mailto:haolin.ye@mail.mcgill.ca)

Mobile : +1-514-567-6198

## EDUCATION

---

- **McGill University** Montréal, CA  
*Bachelor of Honours Software Engineering; GPA: 3.85* *Aug. 2021 – May. 2024(expected)*
- **Marianopolis College** Montréal, CA  
*Pure and Applied Science; CEGEP diploma* *Aug. 2019 – May. 2021*

## EXPERIENCE

---

- **Mila AI Institute** Montréal, CA  
*Undergraduate Researcher - Machine Learning* *May 2022 - Present*
  - **OCaml-BERT**: I helped to develop OCaml-Bert who is a variant of BERT developed by Google and fine-tuned to capture type errors made by human programmers when coding in OCaml. With this fine-tuning method, one can easily obtain similar models for other programming languages. The paper proposes this model is published at <https://conf.researchr.org/home/aplas-2022>, a conference for Programming Languages and Systems.
  - **Data Collection**: I designed a Python web scraping script that downloaded 350k OCaml programs from public GitHub repositories for model training purposes. Data that are obviously noisy are ignored while web crawling.
  - **Data Processing**: I wrote Python scripts to further process and filter the collected data. A Python parser parsed the selected programs into proper sequences of tokens that can be fed to models.
  - **Develop VM**: To compete for our work with other data-driven approaches, I developed our project on the same virtual machine as NATE, a project uses supervised learning to capture type errors, in Python and Haskell.
  - **Evaluation Metric Design**: Designed a systematic and rigorous way of quantifying the error diagnosis ability of a data-driven model.
  - **Dev accessibility and verifiability**: I added quick bash commands which correspond to models with details already specified for people who want to reproduce and verify our experiments.
  - **Data Visualization**: I plotted graphs based on the experimental results for analyzing and comparing the performance of different models, performance under different thresholds, and evaluation metrics.
  - **Academic Writing**: I involved in the academic writing for this paper.
- **ENG WEEK Hackathon** Montréal, CA  
*Winner Team Leader* *March 2021*
  - **McGill Study Mate Finder**: This project is a prototype for an online real-time friend finder to reduce students' anxiety and depression caused by the COVID-19 lockdown. It matches mind-liked people according to their surveys and the information provided. The platform supports online chatting.
  - **Web Design**: I designed web pages for the chatting room, rooms' lobby, friends matching, and survey using CSS, HTML, and JavaScript.
  - **Data Flow**: I developed a program in JavaScript that sends a data structure containing each user's survey results to the backend server as the input of the matching algorithm.
  - **Leadership**: Delegating tasks, communicating goals, managing meetings.
  - **Algorithm Design**: I participated in designing the survey and matching algorithm, which can cluster the students with similar mindsets.
- **Marianopolis College** Montréal, CA  
*Calculus Tutor* *January 2020 - May 2020*
  - **Weekly Tutorial**: Help to clarify students' confusion and provide them with extra practice problems.

## PROJECTS

---

- **Food Service Industry (*Java OOP*)**: An Object-Oriented-Design project written in Java which simulates a mini food service industry.
- **Pandemic Visualization**: Pandemic visualization for COVID-19 using bash commands to analyze the pandemic at a certain region over time.

## GENERAL INFORMATION

---

- **Programming Languages**: Python, Java, OCaml, C++, C, Javascript **Languages**: English Chinese