

STEVEN YANG

☎ (647)469-6927 — ✉ steve.yang@mail.utoronto.ca — 🌐 Website — [in](#) linkedin

Technical Skills

- **Programming Languages:** Python, Java, C, JavaScript, HTML/CSS
- **Developer Tools:** Codex, Claude Code, Git/GitHub, Linux, VS Code, IntelliJ, Jupyter, LaTeX
- **Frameworks/Libraries:** NumPy, pandas, matplotlib, scikit-learn, SciPy, statsmodels

Education

University of Toronto **Sep. 2024 – May 2028 (Expected)**

- *Honours Bachelor of Science*, Data Science Specialist, Computer Science Major
- cGPA: 4.0/4.0
- Relevant Courseworks: Software Design, Software Tools and System Programming, Data Structures and Analysis, Data Science, Probability and Statistics, Linear Algebra, Multivariable Calculus

Experience

Research and Development Intern **May 2026 – August 2026 (Expected)**
NSFOCUS *Beijing, China*

- Built **MVP** components for a **cyber monitoring** platform, supporting threat-data collection and analysis.
- Prototyped a client-profile-driven **threat intelligence** workflow for structured event matching, deduplication, and delivery decision.
- Analyzed **AI Coding Agent**'s architecture and workflow in a research report.

Projects

NHANES Supplement Use Analysis **2026** *Python, pandas, NumPy, matplotlib, scikit-learn*

- Analyzed NHANES 2021–2023 health and nutrition datasets in a team of 2 to investigate relationships between demographic, dietary, and supplement-use patterns.
- Performed large-scale **data cleaning, preprocessing**, feature engineering across multiple linked survey datasets.
- Built multiple statistical and **machine learning** models to classify supplement users and evaluate predictive factors.
- Created **visualizations** and quantitative reports to communicate trends, model performance, and public health insights.

Question Bank Platform **2026** *Java, OpenAI API, Full-stack*

- Built a full-stack Java application to extract and structure math exam questions from PDFs into a searchable digital database.
- Integrated **OCR** and multimodal **OpenAI Vision models** to parse question text, diagrams, metadata, and page-level images..
- Designed async processing with confidence scoring, manual review flags, PDF segmentation, and structured JSON storage.

GroupFlow **2025** *Java, Swing, Clean Architecture, Git, JUnit, MongoDB*

- Designed and implemented a collaborative task and group management application using **Clean Architecture**, separating entities, use cases, and UI layers.
- Built core features including group creation, role-based permissions, task assignment, and progress tracking.
- Improved code maintainability and testability through interface-driven design and unit testing with **JUnit**.
- Collaborated in a team environment using **Git-based workflows**, pull requests, and code reviews.

Awards

- **Dean's List Scholar** 2025, 2026
University of Toronto
- **Woodsworth College Scholarship Fund** 2026
Woodsworth College, University of Toronto