Kevin Yao

📞 647-918-0799 | ✉ yao.kevinl@gmail.com | 🔗 linkedin.com/in/kevin-yao-6a40a2240/ | ⌗ github.com/TheAwakenedBob

EDUCATION

University of Ottawa	Sep 2022 – Present
Honour's Bachelor of Computer Science	*Expected Graduation Date: Dec 2026*

EXPERIENCE

Software Developer | *Python, Tesseract, OpenCV* — Sep 2025 – Dec 2025
Environment and Climate Change Canada — *Ottawa, ON*

- Built an **AI-driven OCR automation system** using **Python**, **Tesseract OCR**, and **OpenCV** to extract structured data from scanned forms in PDFs with over **95% accuracy**
- Optimized document **denoising** by contrasting to improve OCR recognition on poor scans by **30%**
- Engineered a Python **fullstack** pipeline to **automate** repetitive workflows at scale, achieving a **45–55%** efficiency improvement and **100% accuracy** rate

Data Scientist | *R* — Jan 2025 – April 2025
Statistics Canada — *Ottawa, ON*

- Designed and implemented **reusable data wrangling** pipelines in **R** for annual statistic reports on company revenue streams with **dplyr** and **tidyr**
- Reduced processing time by over **80%** using **optimized** data pipelines which standardized **25,000+** rows of raw tabular data for the 2022 and 2023 fiscal year

Data Analyst | *Python, pandas, PyTorch* — May 2024 – Dec 2024
SickKids — *Toronto, ON*

- Collaborated on developing a **natural language processing (NLP) machine learning pipeline** using **Python**, **PyTorch** and **pandas** to automatically fill missing key information for over **20,000** CHIRPP injury cases and classify **30,000+** patients into **50** injury categories with **98% accuracy**

PROJECTS

Paideia | *Next.js, TypeScript, Python, Docker, Gemini AI* — Sep 2025 – Present

- Built an **AI quiz generation platform** with **Gemini + Python FastAPI** microservices on **GCP Cloud Run**, containerized with **Docker** and secured via **Firebase Auth**; achieved **sub-250 ms** median workflow latency
- Developed an adaptive **prompt-chaining** pipeline (LLM reasoning, semantic checks, feedback refinement) to generate quizzes with **1:1 fidelity** to uploaded MCQ exams, preserving complex symbols/formatting
- Engineered a **pdfplumber**-based ingestion pipeline with **column-aware segmentation** and layout heuristics to output structured **JSON**; reached **99% parsing accuracy** on multi-column PDFs/slides
- Implemented **spaced repetition**, real-time analytics, and cache-optimized regeneration using **Firestore batched writes + sessionStorage** for replay without reuploading content

AI Mario Speedrunner | *PyTorch, Stable-Baselines3, Neural Network, Machine Learning* — Feb 2026

- Trained a **deep Reinforcement Learning Proximal policy optimization** agent with a **CNN** policy to speedrun **Super Mario Bros** from pixels + frame stacks, optimizing rightward progress and completion time
- Built a **Gymnasium** pipeline on **gym-super-mario-bros** (discrete action space, checkpointing, deterministic eval, **1080p** milestone videos); trained **5000+ iterations (2M+ timesteps)** with consistent clears and typical finishes at **340–345** time remaining
- Stabilized PPO by tuning **KL/target-KL**, entropy, clipping, LR, batch size, and epochs; achieved steady training (**clip_fraction 0.08–0.10**, **explained_variance 0.92–0.97**) while improving returns

TECHNICAL SKILLS

Languages: Python, Java, SQL, JavaScript, HTML/CSS, C++, Typescript
Libraries/Tools: Git/GitHub, pandas, NumPy, Excel, Firebase, React, PyTorch, MongoDB, Node.js, Express, PyTorch, Azure, Jupyter, UML, R, RStudio, Tesseract, Typescript, Docker, Google Cloud Platform, Gemini AI, Next.js

HOBBIES

Volleyball, Basketball, Weightlifting, Gaming