

# Jake Maly

647-394-2726 | [jake.maly@mail.utoronto.ca](mailto:jake.maly@mail.utoronto.ca) | [linkedin.com/in/jakemaly](https://linkedin.com/in/jakemaly) | [github.com/jakemaly](https://github.com/jakemaly)

## EDUCATION

---

### University of Toronto

Toronto, ON

*Bachelor of Applied Science, Engineering + Artificial Intelligence with PEY Co-op*

*Sep. 2025 – April 2029*

- GPA: 3.6/4.0; x2 Dean's Honours List; Elliot M. Wilson Scholarship (\$5,000)

## TECHNICAL SKILLS

---

**Languages:** Python, SQL, JavaScript, Dart, HTML/CSS

**Frameworks & Libraries:** FastAPI, React, Flutter, Streamlit, LangChain, LangGraph, Hugging Face, DeepEval, RAGAS, Pydantic, pandas, NumPy, Plotly, OpenCV, MediaPipe

**Tools & Technologies:** Git/GitHub, Jupyter, Azure Cosmos DB, SQL, MCP, llama.cpp, Docker, Qdrant, OpenRouter, OpenTelemetry, Jira, LLM APIs (Claude, GPT-4o, Gemini)

## EXPERIENCE

---

### Agentic AI Research Engineer

May 2026 – Present

*University of Toronto*

*Toronto, ON*

- Orchestrating LangChain 3-agent (prosecution, defense, judge) RAG-based pipeline predicting Canadian criminal appeals verdicts across 10,000 cases under Prof. Chi-Guhn Lee and DORL Lab
- Improving verdict prediction accuracy by designing DeepEval + RAGAS evaluation harness
- Building observability layer and dashboard to reduce token cost and latency across 1000s of eval iterations

### Software Engineer, Freelance

Nov. 2025 – Present

*Reframe Marketing*

- Built an automated AI investment analysis pipeline evaluating early-stage startup founders, processing resumes, pitch decks, and financials up to 50 pages to reduce manual review time from 2+ hours to under 3 minutes per startup
- Architected a multi-provider LLM system (Claude, GPT-4o, OpenRouter) with custom prompting, vision capabilities, and regex-based extraction to generate confidence-scored evaluation reports across 15 key founder metrics, delivered via a Streamlit/Plotly dashboard with CSV/Markdown export

*Applied Optimal*

- Shipped features and bug fixes across 15+ endpoints/screens in a Flutter/FastAPI/PostgreSQL multi-tenant club management platform, working within sprint cycles, PRDs, and Jira-tracked code review workflows
- Integrated role-based access control into authenticated backend APIs using Dio interceptors and Supabase JWT handling, connecting frontend UI surfaces to backend routes across the platform

*Blue Marble Academy*

- Built and shipped a primary engineering role on a production AI education platform (React/FastAPI/Azure Cosmos DB) serving Ontario students, architecting a share-link system with role-based permission tiers (owner/viewer/editor) for shareable previews and authenticated cloning across 6 AI-generated study artifact types
- Designed and integrated quantitative question generation into the AI quiz feature using custom prompt engineering, while resolving 6 production bugs across authentication, generation, and review workflows

*EduTutor (Present)*

- Building AI-powered smart quiz creations for Ontario high school teachers

## PROJECTS

---

### MediVue | *Agentic ER Monitoring System*

- Built a serial Research → Diagnostic → Dispatcher LangChain pipeline over Gemini for multi-stage clinical reasoning, with Railtracks session tracing for agent observability
- Designed a RAG memory layer using Moorcheh semantic similarity search over patient-scoped intake documents, with a hot in-process store for zero-latency context during live assessments
- Implemented ElevenLabs/Gemini voice kiosk for structured symptom extraction at intake; MediaPipe/OpenCV deterioration detector feeds visual observations as grounded context into the agent pipeline
- Exposed a FastMCP sidecar as an external agentic control plane for submitting intakes, triggering assessments, and inspecting reasoning traces