

# Sylvester Wong

sylvester@theSylvester.ca

## Agentic AI Engineer | Full-Stack Developer

[theSylvester.ca](https://thesylvester.ca) · [linkedin.com/in/sylvester-wong-41552256](https://linkedin.com/in/sylvester-wong-41552256)

---

### Professional Summary:

Full-stack AI engineer who builds products end-to-end, from RAG and context engineering to agent orchestration, developer tooling, and user-facing interfaces. Built and shipped Crispy, a multi-vendor agent harness for Claude Code and Codex with persistent memory, remote control, and extensible tooling for real-world AI workflows.

### Technical Proficiencies:

- **LLM & AI:** OpenAI, Claude, Gemini, Hugging Face, LangChain, LlamaParse, Prompt Engineering, Context Engineering
- **Agentic Systems:** Claude Code, Codex, Agent SDK, MCP, CLIs, Multi-Agent Orchestration, IPC/JSON-RPC
- **RAG & Databases:** LlamaIndex, Qdrant, Neo4j, PostgreSQL, SQLite/FTS5, Redis, MongoDB, DynamoDB
- **Cloud & DevOps:** AWS, Serverless Architectures, Docker, CI/CD, Microservices, REST APIs
- **Full-Stack Engineer:** React, Next.js, Node.js, NestJS, BullMQ, FastAPI, Pydantic, Zod, Jest
- **Programming Languages:** Python, TypeScript, JavaScript

### Professional Experience:

Crispy, Toronto, Canada

#### Founder & Product Engineer

01/2025-Present

- Built a multi-vendor agent harness orchestrating Claude Code and OpenAI Codex through a unified adapter contract, delivered across five surfaces — VS Code extension, browser UI, CLI daemon, Windows desktop app via Tauri, and as a headless Discord bot with remote agent control from mobile.
- Engineered session orchestration multiplexing live Claude Code and Codex sessions across multiple subscribers, with resume, fork, and cross-vendor fork flows that let users branch any session into a different AI vendor mid-conversation.
- Implemented cross-session agent memory with SQLite FTS5 hybrid BM25 search, local semantic embeddings via llama.cpp, and RRF-based keyword/semantic fusion, plus a backfill pipeline ingesting local Claude Code and Codex transcripts — fully local, no external API calls.
- Designed a parallel code review system pitting Claude against Codex on the same target with independent false-positive review — each model consistently surfaces defects the other misses, reducing single-model blind spots and hallucinations.

**Claro Customs AI**, Toronto, Canada (Remote)

## **Senior Software Developer - AI**

12/2024-12/2025

- Architected a conversational AI with full feature parity to the web frontend of a Canada/US customs platform—document submission, compliance queries, client inquiries, and workflow triggers all through natural-language chat and email, integrated across the full backend stack.
- Increased document ingestion accuracy from 70% to 95% by migrating OCR to LlamaParse and Gemini 2.5 Pro with structured outputs for unclassified multi-document PDF segmentation.
- Built a snapshot-based replay testing framework with an HTML diff dashboard, reducing e2e test cycles from 15 minutes to under 5 minutes by capturing events, jobs, and database state to bypass expensive LLM/OCR calls during development.
- Introduced agentic development practices across a 5-person team—context-isolated task handoffs, self-testing sub-agents, and automated review validation—cutting review resolution time by 50% and establishing a Research → Plan → Delegate → Review methodology as the team's default workflow.

**PromptCore**, Guelph, Canada (Remote)

## **Prompt Engineer**

09/2024-12/2024

- Built a Human-in-the-Loop evaluation system (Next.js, FastAPI) with a custom workflow runner for real-time LLM output scoring, replacing manual prompt-tuning workflows and Microsoft Prompt Flow.

**LighthouseAI**, United States (Remote)

## **Advanced Products Developer**

05/2023-09/2024

- Boosted GPT-3.5-turbo performance on document classification tasks to GPT-4 levels (from 70% to 90%+) by leveraging agentic workflows, combining advanced, multi-step prompting with proprietary logic.
- Developed a modular, multi-vendor Python LLM framework as a lean and flexible alternative to LangChain.
- Built AI-driven lead enrichment agents that automated lookups across Apollo.io, LinkedIn, web search, and Google News to surface sales intelligence for incoming leads.

**TheSylvester Creations**, Toronto, Canada

## **Freelance Software Developer / Game Developer**

12/2021-05/2023

- Contributed a UX enhancement PR merged into Visual Studio Code's mainline release.

**Calabria Roofing Inc.**, Toronto, Canada

## **Software Developer & IT Lead**

12/2016-03/2021

- Built an ERP application automating project tracking, inventory management, and financial reporting.

## **Education**

**Computer Programming and Analysis**, Seneca College

Toronto, Canada

2003