

Niloy Roy

www.royniloy.com

Toronto, ON • its.royniloy@gmail.com • www.linkedin.com/in/r-niloy • <https://github.com/R-Niloy>

SUMMARY

Machine Learning Engineer specializing in agentic AI systems that run in production. Shipped a client-facing conversational assistant that made content publishing roughly 8x faster, founded an agentic AI restaurant-operations platform, and built LLM agent platforms for clinical research that produced three peer-reviewed publications.

EDUCATION

Toronto Metropolitan University

September 2018 – April 2024

Bachelor of Engineering, Computer Engineering (Software Engineering Stream)

- Academic Excellence: Dean's List Recipient 2021–2022, 2023–2024.

PROFESSIONAL EXPERIENCE

Software Engineer (AI)

December 2025 – Present

L Squared Digital Signage, Toronto

- **Built** an intelligent chat assistant giving clients end-to-end natural-language control of the digital signage platform, orchestrated with LangGraph, MCP servers, and third-party LLM inference, with Redis for in-memory state management and Grafana for real-time monitoring.
- **Orchestrated** custom ML classifiers and data pipelines to solve person/entity classification and re-identification challenges, enabling intelligent agent reasoning.
- **Developed** context-aware agentic systems with Postgres and pgvector memory management that persist and retrieve contextual information through vectorized reranking queries, enabling autonomous agents to deliver real-time insights with stateful multi-turn reasoning.

Founder & Lead Developer

September 2025 – Present

Dishpit - AI Restaurant Automation Platform, Toronto

- **Founded and shipped Dishpit**, an agentic AI restaurant-operations platform (React, Vite, TypeScript, Supabase, Tailwind CSS, shadcn/ui), owning agent architecture, product, and deployment end to end.
- **Engineered** a multi-agent system with per-domain agents for scheduling, inventory, and menu optimization, each grounded in live operational context and coordinated through a shared state layer to reason across turns.

AI Solution Leader

March 2025 – September 2025

Opifiny Corp., Toronto

- **Led** the end-to-end design and deployment of a production-grade document explainer using C# and Google's Gemini Pro API for advanced clinical data extraction and summarization.
- **Architected and deployed** a company-wide **agentic RAG chatbot** on AWS using Anthropic's LLM and MySQL-native vector search, providing instant internal knowledge base support.

Automation and Generative AI Consultant (Contract)

December 2024 – March 2025

Freedom Mobile, Toronto

- **Implemented** automated testing frameworks using TypeScript and Selenium, achieving a 70% reduction in manual testing effort.
- **Deployed** an AWS-based ETL pipeline integrating AWS Textract and NLP technologies for automated text extraction.

ETS Master Data and Material Management Co-Op

September 2022 – August 2023

Sanofi, Toronto

- **Spearheaded** an integrated inventory management solution connecting Rockwell Automation with 3D warehouse models and Power Automate, reducing overhead costs by 10%.
- **Developed** comprehensive data pipelines using Python and SQL to ingest data from 8 distinct sources, utilizing R for statistical modeling and trend prediction.
- **Built** interactive Power BI dashboards capturing 30+ KPIs to drive stakeholder decision-making while coordinating cross-functional teams using Agile methodologies.

RESEARCH & OPEN-SOURCE ENGINEERING

Machine Learning Engineer (Research)

September 2024 – Present

Interventional Psychiatry Program, St. Michael's Hospital - Unity Health, Toronto

- **Engineered** the Meta-Analysis Assistant (MAA), an LLM-powered LangGraph agent that automates data extraction and statistical computation (in R) for systematic reviews and meta-analyses.
- **Built** the STACER Agentic System (SAS), an LLM-powered simulation platform pairing a virtual patient agent with an automated rater agent to deliver STACER-aligned feedback for psychiatry residency training.
- **Built front-end interfaces for multiple AI initiatives and deployed and maintained the resulting agentic AI applications on AWS.**

Keenan Summer Research Fellow

May 2024 – August 2024

St. Michael's Hospital - Unity Health, Toronto

- **Accelerated** research progress through the application of machine learning techniques to medical imaging, particularly MRI.
- **Implemented** NLP frameworks for analyzing therapeutic outcomes, utilizing BERT and transformers.
- **Conducted** rigorous quality assessments and analyzed functional MRI datasets, ensuring 100% accuracy and reliability.
- **Authored** literature reviews and coordinated the preparation of research proposals, significantly contributing to two publications.

SKILLS

- **ML & AI:** PyTorch, TensorFlow, Scikit-Learn, LangGraph, RAG, Vector Search, Reranking, MCP, Transformers, LLMs, NLP, Computer Vision
- **MLOps & Infrastructure:** Kubernetes, CI/CD (GitHub Actions, Jenkins), AWS, Git, Model Deployment & Monitoring
- **Programming & Data:** Python, SQL, C#, Java, R, Pandas, NumPy, ETL, PostgreSQL, pgvector, Redis, NoSQL
- **Web & Product:** React, TypeScript, Vite, Supabase, Tailwind CSS, REST APIs
- **Developer Tools & Methodologies:** Playwright, Selenium, JIRA, Agile/Scrum, Docker
- **Languages:** English (Fluent), Bengali (Native).

SELECTED PROJECT EXPERIENCE

Avatar-Based STACER Simulated Patient

Tools: Python, MuseTalk, LangGraph

- **Extended** the STACER Agentic System with an embodied virtual patient, streaming audio tokens through MuseTalk (open-source Python package) to generate realistic 2D facial expressions and lip-sync in step with the patient agent's speech.

MRI Literature Quality Assessment for sgACC Review Checklist

Tools: Excel, MRI Literature

- **Conducted** a comprehensive quality assessment of 42 MRI studies focused on the **Subgenual Anterior Cingulate Cortex (sgACC) neuroimaging**, using a standardized checklist, ensuring reliability and validity of MRI methodologies in Neuroimaging research.
- **Collaborated** with a cross-functional research team to refine the assessment framework, integrating feedback mechanisms that enhanced the quality assurance process and improved overall review times by 25%.

Estimating Quality in Therapeutic Conversations Using NLP

Tools: Python, BERT, Scikit-learn

- **Architected and implemented** an NLP-based machine learning system achieving 94% accuracy in therapeutic quality prediction. **Developed** custom sentiment analysis models using BERT and transformer architectures.
- **Engineered** data preprocessing pipeline handling 1,500+ chat transcripts and **implemented** SMOTE for handling imbalanced dataset challenges.

CERTIFICATIONS

- **AWS:** AWS Certified Cloud Practitioner (2025)
- **Bootcamp:** Build Multi-Agent Applications - A Bootcamp - LangGraph, Cursor, n8n (2025)

PUBLICATIONS

- [Enhancing Psychiatry Training Using an Agentic AI Simulated Consultation Tool: Prospective Cohort Study](#)
JMIR Medical Education 12, e88580 (2026).
- [Subgenual anterior cingulate cortex in major depression: a systematic review and meta-analysis of MRI studies with patients and healthy controls](#)
Translational Psychiatry 16, 225 (2026).
- [Meta-Analysis of Magnetic Resonance Imaging Studies with Unmedicated Patients with Depression and Healthy Controls Reveals Subregions of the Subgenual Anterior Cingulate Cortex with Distinct Structural and Functional Properties](#)
Biological Psychiatry, Volume 97, Issue 9, S199