Liliana Hotsko

lilianahotsko   in Liliana Hotsko   ⌂ lilianahotsko

Education

BS	**Ukrainian Catholic University**, Computer Science	Sept 2021 – May 2025

- **GPA**: 90/100 (Diploma Supplement ↗)
- **Thesis:** Comparative Analysis and Framework Development of Anonymization Techniques for Large Language Models.

University of Tartu, Computer Science — Sept 2024 – Feb 2025

- **Erasmus+** student exchange program.
- **Level:** A (Transcript ↗)

DePaul University, Computer Science — March 2022 – June 2022

- **Course:** Topics in AI
- **Course Work:** Family psychologist implementation methods.

MS **University of Waterloo**, Computer Science — Sept 2025 – Sept 2027

- Master of Mathematics, David R. Cheriton School of Computer Science
- Research Topics: NLP, Model training optimization methods

Certificates & Licences

IBM: Generative AI Engineer	Credential ↗
IELTS: English Proficiency	**Level: C1 (8)**

Experience

Waterloo University, Visiting Researcher — Waterloo, ON, Canada / May 2025 – Aug 2025

- Compared training techniques like Full-finetuning, LoRA and Hypernetwork for LoRA weights generation.
- Invastigated the memory requirements on GPU for these training methods.
- Tuned Qwen2.5-1.5B model for repository-based task-specific usage.

Eleks, Research & Development Engineer — Aug 2023 – Apr 2025

- Conducted research on LLMs and agentic systems in enterprise environment.
- Continuously created, tested and integrated applications for process optimizations.

Center for Responsible AI, Research Fellow — Sept 2023 – Dec 2023

- Tandon School of Engineering at the University of New York.
- **Project:** Benchmarking: Monotonic Graph Neural Networks (MGNNs)
- Supervised by David Tena Cucala (University of Oxford)

Student Research and Development Group, ML and Robotics Specialist — Sept 2023 – April 2024

- I was working on a project related to real-time embedded systems with integrating visual models. (Additional details under NDA)

Selected Works

LLM Anonymization Framework — GitHub ↗ / Thesis Document ↗

- Bachelor thesis in Computer Science.

- This project provides a modular framework for anonymizing sensitive data before sending it to external large language models (LLMs) such as GPT, Claude, or Gemini. It supports customizable anonymization pipelines, retrieval-augmented generation (RAG) workflows, and evaluation tools for measuring masking quality and retrieval performance.

  Features:
- Interactive anonymization pipeline
- RAG retrieval with ChromaDB and MongoDB Atlas
- Homomorphic Encryption
- Supports Ukrainian and English languages
- Open-source modular design
- UI for testing and integration.

MacTell: OPEN Interpreter

GitHub ↗
Preprint ↗

- Swift-powered project designed to integrate multiple functionalities into a macOS application. Using natural language processing, Mactell interprets user commands into executable actions, significantly improving user interaction within OS and productivity.
- Key functionalities include command-line interface (CLI) management, web browsing, code execution, media playback, and other OS-accessible tasks, all presented through an intuitive, unified interface.

Benchmarking: Monotonic Graph Neural Networks (MGNNs)

GitHub ↗
Report ↗

- Extracted explanatory rules from the trained Monotonic Graph Neural Networks and evaluated their quality.
- Invastigated how adjusting the parameters of the MGNN and the rule extraction algorithm impacted the quality of the extracted rules, ensuring optimal model performance and interpretability.

Delta/Terminal: Tool for analyzing business data with environmental factors

GitHub ↗

- Dashboard for businesses affected by the environmental changes.
- Developed functionality provides ETL process using Apache Airflow, Apache Spark, Apache Iceberg, Hadoop, DBT, and Postgres; and supplies users with secure environment to manage the data using Streamlit, NGNIX, and Google Auth.
- The poster won the 2nd place at the project competitions.

LLM Zoom to Salesforce Processor

GitHub ↗

- In the sales process, any meeting is a valuable source of information about the customer. Sales executives usually must fill this information into CRM manually. I developed an application that analyses data(.json) and transcript(.webvtt) of the Zoom meeting with LLMs taking the specific needs of the organization into account and updating contacts on Salesforce.

Voice Recognition in a Low Computational Environment

GitHub ↗

- Voice recognition in a low computational environment. Embedded ML project using PSoC board and Edge Impulse framework.
- Technologies: Real-time systems, Sound processing, FFT transform, microcontrollers, Transfer learning
- The project was supervised by Infineon.

Technologies

Languages: Python, C, C++, Java, R

Machine Learning: NLP, Computer Vision, Forecasting, pandas, numpy, TensorfFow, HuggingFace, PyTorch

Generative AI: Transformers, LangChain, vector databases, RAG systems, Fine-tuning, agent systems, VertexAI

Data Engineering: Data Bases (SQL, NoSQL, MongoDB, Neo4j), Docker, Apache Airflow, Apache Iceberg, data visualization

General: Python, C++, Git, Google Cloud, Azure, Flask, API, microservice architecture, Docker, Unix/Linux, distributed and parallel systems, embedded systems

Soft Skills: Project management, supervising, presenting, communication with team and customers, requirements management