Liliana Hotsko

lilianahotsko Liliana Hotsko lilianahotsko

Education

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

- **GPA**: 90/100 (Diploma Supplement)

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

- **Level:** A (Transcript)

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

- **Fully funded** research-based graduate program, **AI Lab**
- **Research Interests**: NLP, Multimodal systems, Efficient Adaptation (LoRA/PEFT)

Awards

International Masters Award of Excellence	2025-2027
MITACS Award	2025

Work Experience

University Of Waterloo, Student Researcher Intern — May 2025 - Aug 2025

Co-supervised by Pengyu Nie and Yuntian Deng. Designed a repository-aware code-generation **hypernetwork** that generates **LoRA** adapters for the frozen model on the fly.

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

I worked on designing applications for internal process automation. I was one of the first AI R&D Engineers at the company, therefore, I created the main expertise on **AI tools for automation**, which included **RAG systems**, **agentic systems**, and **user support streamlining**. The systems I developed were integrated by one of the biggest logistics companies in Europe.

Center for Responsible AI, Student Researcher Intern @ New York University — Sept 2023 – Dec 2023

I worked on Monotonic Graph Neural Networks(MGNNs) in cooperation with **New York University** and **Oxford University**. **Benchmarked** the proposed system against the datasets with logical expression simplification, family relations, and other domains. Proposed **algorithmic improvements** to the initial work.

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

I was working on **real-time embedded systems** and **on-device ML** models integration. **Area: dynamic motion planning.**(Additional details under NDA)

Publications

Program as Weights | Paper in Submission

Contributed to the development of a **neural compiler–interpreter framework** that translates high-level task descriptions into lightweight **neural programs** (discrete tokens and continuous vectors) executed locally. The highlight of the proposed coding paradigm is that it supports versionable and reproducible execution of fuzzy logic traditionally handled via LLM APIs.

LLM Anonymization Framework | GitHub | Publication

I created 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 with ChromaDB and MongoDB Atlas, Homomorphic Encryption, Supports Ukrainian and English languages, UI for testing and integration.

AR Safety Training Research @ ELEKS | Publication

Conducted research on **AR-based workplace safety training systems**, analyzing hazard-simulation workflows and industrial training requirements. Evaluated how AR can improve hazard recognition, and training efficiency for manufacturing and energy clients.

Selected Research Works

CheapInfer @ University of Waterloo | Preprint

Developed a hybrid LLM serving architecture that offloads initial token generation to CPUs via speculative decoding, reducing GPU contention. Integrated **CPU drafting and GPU verification** into the vLLM scheduler for speculative decoding, enabling earlier computation of draft sequences for queued requests.

Monotonic Graph Neural Networks (MGNNs) @ New York University & University of Oxford | GitHub | Preprint

Extracted explanatory rules from the trained Monotonic Graph Neural Networks and evaluated their quality. Investigated 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.

Hypernetwork for Repository-Level Code Completion @ University of Waterloo | In Progress

Building a **hypernetwork**-based **LoRA** generator that produces task- and repo-specific adapters directly from repository embeddings, allowing full-context conditioning and rapid on-demand specialization.

MacTell: OPEN Interpreter | GitHub | Preprint

Swift-powered project designed to integrate multiple functionalities into a macOS application. Using **NLP**, 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.

Voice Recognition in Low Computational Environment | GitHub

Explored **on-device ML** for **low-resource microcontrollers** by implementing a keyword-spotting model using Edge Impulse and transfer learning. Designed a lightweight audio feature extraction pipeline (FFT-based spectrograms) and optimized model quantization for embedded deployment. Demonstrated robust inference under extreme hardware constraints on a **PSoC** board. Project done in collaboration with **Infineon**.

Certificates & Licences

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

Volunteering

Google Developer Group - Technical Assistant at the workshops for Google Developer Fest.
Conference Reviewer - reviewer for ICLR Conference.

Skills

Machine Learning: NLP, Multimodal systems, Transformers, Computer Vision, LangChain, vector databases, RAG systems, multiagent systems, pandas, numpy, TensorfFlow, HuggingFace, PyTorch

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

Embedded systems: On-device real-time systems and microcontrollers, AR/VR/XR, ML integration into low-computational environments

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

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