

# ARON MOLNAR

Cambridge, UK [Arotte.github.io](https://github.com/Arotte) | [LinkedIn/molnaron](https://www.linkedin.com/in/molnaron) | [GitHub/Arotte](https://github.com/Arotte) | [molnar2on@gmail.com](mailto:molnar2on@gmail.com)

## SUMMARY

- Mid-Senior Research and Machine Learning Engineer with 6 years of experience in the industry.
- Publishing cutting-edge NLP/LLM research at top conferences (e.g. EMNLP).
- Strong software engineering background – built systems at large multi-national software companies, small start-ups and fast-paced scale-ups.

### Multi-faceted ML engineer and researcher with expertise in:

Natural Language Processing (NLP) – Machine Learning (ML) – Large Language Models (LLMs)  
– Training and Deploying LLMs – All aspects of the software development lifecycle

## EXPERIENCE

### R&D ENGINEER (LLMs)

Sep 2023 – Current

Luminance AI, Cambridge, UK

- Fine-tuning and deploying custom legal LLMs that compete with GPT-4 performance in various use cases.
- Creating LLM-based chatbots that power discourse over large repositories of legal documents. Discovering new ways to prompt, train and efficiently infer domain-adapted LLMs.

### RESEARCH ASSISTANT (Language Modeling)

Jun – Sep 2023

University of Aberdeen – The Context Lab, Aberdeen, UK

- Investigating the effects of local context during utterance generation in dialogue. First-author paper published in CoNLL.
- Running computational experiments that compare the inner mechanisms of LLMs and the human mind (Tech: Python, PyTorch, HPC).

### SOFTWARE ENGINEER

2022 – 2023 (1 year)

University of Aberdeen, Aberdeen, UK

- Constructed three full-stack systems that substantially automate various internal processes.
- Created an automation tool that saves hundreds of developer hours annually and two management tools for handling sensitive assets. (Tech: C#.)

### SOFTWARE ENGINEER

2022 – 2023 (1 year)

IT-Pro, Wales, UK

- Delivered four versatile full-stack systems, significantly improving businesses across various sectors. These include business management systems for the NHS, a defence industry supplier and a ski holiday business; a full-stack mobile app for a safety consultancy company. (Tech: Java, C#.)

### SOFTWARE DEVELOPER

2019 – 2021 (2 years)

HyperDAP, Aberdeen, UK (university spin-off startup)

- Designed and implemented a novel AI algorithm, enabling accurate prediction of potential oil well failures based on temperature sensor readings. (Tech: Python, Java, Kafka, Cassandra.)
- Developed a computer vision algorithm to detect indoor objects, integrating it into an app that generates household energy efficiency reports using temperature camera readings. (Tech: Java, Python.)

## SOFTWARE DEVELOPER

2019 (6 months)

CollMot Entertainment, Budapest, Hungary

- Delivered a robust API that enables seamless management of drone fleets (comprising hundreds of units) for captivating drone shows. (Tech: Python.)

## SOFTWARE DEVELOPER

2018 – 2019 (1 year)

Ericsson, Budapest, Hungary

- Contributed to the architectural refactoring of a telephony application server with millions of lines of legacy code responsible for routing and managing millions of daily phone calls nationwide, gaining invaluable knowledge in large-scale software engineering. (Tech: C/C++.)

## RESEARCH

I aim to expand my understanding of language models' sometimes unreasonably mind-blowing performance in dialogue settings. I have an engineering mindset, however, in my spare time, I conduct AI research at The Context Lab (TCL) as a research assistant, investigating LLM behaviour in dialogue settings.

- *Attribution and Alignment: Effects of Local Context Repetition on Utterance Production and Comprehension in Dialogue* published in CoNLL 2023 co-located with EMNLP in Singapore. Link to arXiv.
- In an ongoing project, we plan to *understand the extent to which LLMs behave similarly to humans when comprehending structure in language, and whether resulting activation and prediction patterns are similar*. Link to project description.

## EDUCATION

BSc Computing Science / University of Aberdeen, Aberdeen, UK

2019 – 2023 (4 years)

- Graduated with first-class honours.
- Authored a thesis on LLM's human-like linguistic patterns in dialogue, introducing three novel contributions: a pipeline-based experiment architecture, large-scale experiments, and a unique LLM interpretability algorithm. (Models: GPT-2, OPT, DialoGPT. Tech: Python, PyTorch, HPC.)
- Achieved a rare A1 grade in a security assignment involving password cracking, which the professor hadn't awarded in a decade, displaying adeptness in digital security and Python.
- Successfully led and tutored a seven-member team to develop a high-scale software system comprising backends and a mobile app (an intelligent kitchen assistant), winning the annual CS competition.

## TECH

Expertise in a large range of technologies that support a workflow from fast prototyping to robust deployment.

Python (5 years) – PyTorch (3 years) – HuggingFace (2 years) –  
HPC (1 year) – LLMs (2 years) – Pandas (3 years) – NumPy (3 years) –  
LangChain (<1 year) – TensorFlow (<1 year) –

SQL (5 years) – Kafka, Cassandra (1 year) –  
MSSQL, PostgreSQL, MongoDB (2-3 years) –

Java, JavaScript (4 years) – C#, C/C++ (2 years) –  
ASP.NET (1 year) – Node.js (2 years) –  
Flask, Django (2-3 years)

## LANGUAGES

English (proficient) – Hungarian (native)