Hassan Sepehr

hassan@sepehrcode.com · github.com/hsahmad112

Education

Master of Science (MSc), Computer Science *September 2024 – July 2026*
Department of People and Technology, Roskilde University, Denmark

Thesis: Adapting the MATES data-influence framework for register-cohesive segment-level selection in Pythia continued pretraining.
Grade: 10 out of 12

- Implemented continued LLM pretraining pipelines on HPC infrastructure, adapting TurkuNLP's CORE register-segmentation pipeline and the MATES data-influence framework; conducted ablation studies across Pythia-70M model variants with Pareto cost-performance analysis.

Bachelor in Humanities and Technology (BSc) *August 2016 – February 2021*
Department of People and Technology, Roskilde University, Denmark

Specialization: Computer Science and Informatics

Thesis: LocaLive - developed a platform interfacing actors in the live-music environment
Grade: 12 out of 12

Research & Academic Experience

Student Research Assistant *August 2025 – May 2026*
Digital Media Lab, Roskilde University, Roskilde, Denmark

- Designed and implemented a secure inference environment using SSH remote port forwarding within an isolated institutional network, enabling GDPR-compliant access to locally hosted LLMs for researchers working with sensitive interview data (LAIQA).
- Collaborated with interdisciplinary researchers to translate complex research questions into transparent, reproducible computational methodologies.
- Structured and processed unstructured social media data for cross-platform comparative analysis (AlterUse and Altea).

Publications

LAIQA: Local Artificial Intelligence for Qualitative Analysis *In preparation, 2026*
Co-author

- Describes an open-source, locally hosted LLM toolkit for qualitative coding of sensitive interview data, built around a two-stage codebook construction and application workflow with transparent probability reporting, developed at Digital Media Lab, Roskilde University.

Technical Competencies

Systems & Infrastructure: HPC clusters (Slurm for job scheduling and monitoring), Docker, comfortable in Linux environments, Git version control, data pipeline engineering.

Programming Languages: Python, SQL (PostgreSQL).

Machine Learning & Frameworks: PyTorch, HuggingFace, Ollama, CUDA, scikit-learn, data-influence frameworks.

Theoretical Foundations: Transformer-based architectures, loss functions and optimizers, solid grasp of statistical methods.

Other Professional Experience

Freelance data engineer *October 2026*

Software Engineering Intern *September 2025 – December 2025*
Ampleo, Roskilde, Denmark

- Completed the migration of client-facing software from .NET 7 to .NET 10 utilizing Blazor, Razor syntax, and MVC, meeting company and client requirements. Part of Master's education, undertaken simultaneously with work at Digital Media Lab.

Junior Data Specialist *September 2024 – August 2025*
Daquma, Copenhagen, Denmark

- Developed and deployed RAG pipelines over proprietary pharmaceutical documents within isolated Azure organizational tenants, maintaining GDPR-compliant data boundaries between client and internal environments.
- Built a generalized Python ETL library from a project-specific tool.

Technical Support Specialist *May 2022 – August 2024*
Foundever, Turku, Finland

- Delivered technical support and enhanced skills in server infrastructure.

IT Supporter *August 2020 – February 2021*
21-5, Hørsholm, Denmark

- Onsite user IT support, managed Office 365 and Active Directory, and optimised on/offboarding processes using Power Automate.

Language Proficiency

- **English:** Fluent (TOEFL iBT 2024 score: 110/120)
- **Finnish:** Beginner
- **Danish:** Native