

MARK ASSEJEV

+45 60569383 | mr.assejev@gmail.com | www.markererer.github.io

EDUCATION

IT UNIVERSITY OF COPENHAGEN

Copenhagen, Denmark

Bachelor of Science (BSc), Data Science

2022-2025

• Relevant Coursework: Algorithms, Machine Learning, Databases, Large-Scale Data Analysis, Data Visualisation

PROFESSIONAL EXPERIENCE

Simplesoft OÜ

Remote

Junior Data Scientist

September 2024 - August 2025

• Designed and maintained a daily analytics pipeline exporting event-based .csv data from Google Analytics into BigQuery, supporting Looker Studio visualizations and business performance tracking.

• Developed SQL scripts to extract prompts and behavioral segments from RevenueCat’s export DB, enhancing product-led insights.

• Ensured GDPR compliance across all reporting pipelines and external data integrations.

• Created and advised on dynamic dashboards in Looker Studio, improving app engagement monitoring across >10 key user interaction events.

JustDev OÜ

Remote

Junior Data Scientist

December 2024 - September 2025

• Developed components of an AI-driven news pipeline aggregating and parsing global RSS feeds (>150/day), enabling near real-time geopolitical risk scoring.

• Co-developed a real-time threat alerting system based on NLP-derived military, political, and economic signals.

• Contributed to backend data architecture and validation framework using custom scoring rules and historical conflict datasets for system benchmarking.

Southwestern Advantage

Alberta, Canada

Sales Representative

June 2024 - September 2024

• Generated over \$26,000 in total sales revenue through direct sales, using a door-to-door approach.

• Demonstrated entrepreneurial mindset, time management, and performance under pressure.

Invisible Technologies Inc.

Remote

Data Operator

Aug 2023 - May 2024

• Executed cleaning and validation processes using custom-built algorithms to ensure data accuracy and integrity.

SELECTED PROJECTS

Bachelor Thesis: A Critical Evaluation of the LLM2Vec Method — PyTorch, Hugging Face, HPC, Slurm

Conducted a controlled empirical study comparing decoder-only and encoder-only LLMs on NER and POS tagging tasks. Extended LLM2Vec to support DeBERTa probing and UD-EWT datasets; trained on HPC clusters (A100). Surpassed original baselines by implementing MaChAmp’s dynamic learning and gradual unfreezing strategies.

Large-Scale Yelp Review Analysis — Apache Spark, scikit-learn, Pandas

Used Spark to analyze 6M+ Yelp reviews, identified influencer impact and authenticity language, and built logistic and linear regression models to predict review scores.

MLOps Pipeline with Go and Python — Dagger, GitHub Actions, MLflow

Transformed a monolithic Python ML project into a modular pipeline with automated testing, model training, and artifact management. Integrated CI/CD using GitHub Actions and container-based execution.

Comparison of DistilBERT to LSTM using CheckList — PyTorch, NumPy, NLP

Compared a fine-tuned DistilBERT transformer model with a custom LSTM on sentiment classification tasks using the CheckList evaluation framework. Analyzed model robustness to negation, paraphrasing, and named entity changes, revealing gaps in generalization and syntactic sensitivity.

ADDITIONAL INFORMATION

Languages: English (fluent), Estonian (fluent), Russian (fluent), Danish (beginner)