

# Amirhossein Layegh

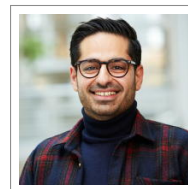
KTH Royal Institute of Technology  
Stockholm, Sweden

✉ [amlk@kth.se](mailto:amlk@kth.se)

🌐 [amirlayegh.github.io](https://amirlayegh.github.io)

in [amirhosseinlayegh](#)

🔗 [AmirLayegh](#)



## Research Interests

Scalable Machine Learning (ML) and Deep Learning (DL), Information Retrieval, RAG Techniques, Large Language Models, Natural Language Processing, Model Optimization and Fine-tuning

## Education

2021 – now **PhD in Information and Communication Technology**, KTH Royal Institute of Technology, Sweden

**Thesis Title:** Natural Language to Executable DAGs: A General Approach for Workflow Orchestration in Big Data and ML Systems

**Advisers:** Prof. Mihhail Matskin and Dr. Amir H. Payberah

*Research is dedicated to advancing novel retrieval and information extraction systems that leverage LLMs, with a emphasis on:*

- Investigating AI agentic workflows for code generation
- Designing prompting strategies that leverage external knowledge bases for zero-shot scenarios
- Exploring approaches to construct knowledge graphs from code bases
- Building domain knowledge graphs and integrating them into GraphRAG pipelines for retrieval-augmented generation
- Creating fairness-aware data augmentation techniques using knowledge graphs

2018 – 2019 **MSc in Big Data Science**, Queen Mary University of London (QMUL), UK

**Thesis Title:** Implementation of a Recommendation System for a retail store based on basket analysis.

**Adviser:** Dr. Arman Khouzani

*Developed a recommendation system for Koolbitz, a London-based retailer, using the Apriori algorithm for basket analysis. The system identifies patterns in purchase data to suggest products to customers, enhancing sales and customer experience*

2012 – 2017 **BSc in Software Engineering**, Ferdowsi University of Mashhad (FUM), Iran

**Adviser:** Dr. Mohsen Kahani

## Industrial Experience

2025 – now **Technical Content Consultant**, Superlinked, Inc., Remote

**Project: Content Creation for AI and ML Community Platform.** Engaged under a consultancy agreement to research, implement and publish advanced retrieval-systems content for Superlinked's VectorHub platform, [LINK](#)

- Authored a benchmark article “*Airbnb Search Benchmarking – Comparison of Retrieval Techniques*”
- Built a reproducible Python pipeline that compares BM25, dense-vector search, hybrid fusion, ColBERT, cross-encoder reranking and Superlinked's mixture-of-encoders on Stockholm Airbnb dataset
- Open-sourced notebooks, visualisations and code, enabling developers to reproduce results and adapt the methodology to their own structured-data search problems

- 2020 – 2021 **Data Scientist**, *Neshan Maps*, Iran  
Utilized deep neural networks trained on large-scale user datasets to develop speed estimation models for analyzing traffic flow on individual road segments. Implemented an inference application deployed as a Kubernetes pod to dynamically estimate speed segments, improving traffic forecasting, navigation accuracy, and overall user satisfaction
- 2019 – 2020 **Data Analyst**, *Koolbitz Ltd.*, UK  
Developed and implemented machine learning algorithms to analyze large sales datasets, identifying key patterns and trends for product recommendations. This significantly improved targeted marketing strategies and customer engagement
- 2017 – 2018 **Software Developer**, *Ranir*, Iran  
Engaged as an Oracle Application Development Framework (ADF) developer, focusing on building Java-based enterprise applications

---

## Honours & Awards

- Jun 2025 **Winner – Google Cloud AI Hackathon (Firebase Studio Enterprise Vibe Coding)**, *Google Cloud Sweden Region Launch 2025*, Stockholm, Sweden  
Built “AI Procurement Assistant” – a full-stack app that lets retail buyers run natural language semantic searches over supplier catalogues, powered by Superlinked embeddings, Qdrant vector DB, a FastAPI backend on Cloud Run, and frontend generated by Firebase studio, [LINK](#)

---

## Publications

- INTERSECTIONRE: Mitigating Intersectional Bias in Relation Extraction Through Coverage-Driven Augmentation, CIKM 2025, under-review
- REA: Refine-Estimate-Answer Prompting for Zero-Shot Relation Extraction, Presented as full paper at NLDB 2024, [LINK](#)  
*Developed a novel multi-stage prompting framework for improved information extraction in zero-shot scenarios*
- Wiki-based Prompts for Enhancing Relation Extraction using Language Models, Presented full paper at SAC 2024, [LINK](#)  
*Created a knowledge-enhanced prompting approach leveraging external knowledge bases for improved extraction*
- ContrastNER: Contrastive-based Prompt Tuning for Few-shot NER, presented as full paper at COMPSAC 2023, [LINK](#)  
*Introduced a novel contrastive learning framework for few-shot entity extraction*
- Datacloudsl: Textual and Visual Presentation of Big Data Pipelines (2022), [LINK](#)
- A survey of big data pipeline orchestration tools from the perspective of the datacloud project (2021), [LINK](#)

---

## Skills

- **Programming Languages & Frameworks:** Python, PyTorch, HuggingFace Transformers, Streamlit, FastAPI
- **Large Language Models:** Prompt Engineering, RAG, Vector Databases, Parameter Efficient Fine-tuning (LoRA, Adapter), LangChain, LangGraph, Fine-tuning, Zero-shot/Few-shot Learning, Contrastive Learning
- **MLOps & Infrastructure:** GCP, SageMaker, ZenML, Neo4j, Git, Docker, KubeFlow
- **Deep Learning:** Neural Networks, Transformers, Foundation Models