Kidist Amde Mekonnen

PhD Researcher in Information Retrieval, Generative AI & Recommender Systems

kidistamde · kidist-amde · Google Scholar · k.a.mekonnen@uva.nl · Amsterdam, NL

SUMMARY

Third-year PhD researcher at the **University of Amsterdam IRLab** working on generative information retrieval, neural retrieval, recommender systems, and multilingual NLP. **Incoming PhD Machine Learning Research Intern at Booking.com** working on scalable and generalisable ID embedding learning. Published at **SIGIR 2025**, **SIGIR 2026**, **ACL Findings 2025**, and **MeLLM @ ACL 2026**. Experienced in building end-to-end retrieval and recommendation research pipelines using **PyTorch**, **Hugging Face**, **FAISS**, **Docker**, and **Slurm**. Interested in generative retrieval, LLMs for retrieval, representation learning, multilingual search, and robust retrieval systems.

RESEARCH EXPERIENCE

PhD Machine Learning Research Intern — Aug 2026 – Oct 2026 (Upcoming)

Booking.com — Amsterdam, NL

- Incoming research intern working on **scalable and generalisable ID embedding learning** for large-scale recommendation systems.
- Research focus on learning and evaluating identifier representations for scalable retrieval and recommendation.

PhD Researcher — Oct 2023 – Present

University of Amsterdam, IRLab — Amsterdam, NL

- Developed methods for **continual generative retrieval** over dynamic corpora, mitigating catastrophic forgetting during sequential index/model updates; accepted as a **SIGIR 2026 full paper**.
- Reproduced and stress-tested **look-ahead decoding** for generative retrieval under query variation, typographical noise, and cross-lingual shift; accepted in the **SIGIR 2026 Reproducibility Track**.
- Designed **DDRO**, a lightweight pairwise relevance optimization objective for generative retrieval, improving **MRR@10 by 7.4% on MS MARCO** and **19.9% on Natural Questions**.
- Led Amharic passage retrieval benchmarking, including dataset construction, model training, and evaluation; improved dense retrieval **MRR@10 by 5.01%** over strong multilingual baselines.
- Built training and evaluation pipelines for retrieval experiments, including ablations, error analysis, re-ranking, and reproducibility checks.

Data Science Research Intern — Apr 2022 – Oct 2022

Nanovery Limited — Newcastle upon Tyne, UK

- Implemented and evaluated **GAN-based** models for synthetic DNA strand generation for nanorobot probe design.
- Developed predictive models for forecasting sensor behavior across laboratory environments and communicated results to research stakeholders.

TECHNICAL SKILLS

- **Languages:** Python, Java, C++, R
- **ML/IR:** PyTorch, TensorFlow, Hugging Face, FAISS, Elasticsearch, dense retrieval, sparse retrieval, learning-to-rank, recommender systems, LLMs
- **Systems:** Docker, Slurm, Git, Linux/Unix, reproducible experimentation
- **Methods:** ablation studies, error analysis, benchmarking, reproducible research

EDUCATION

PhD Candidate, Information Retrieval — Oct 2023 – Present

University of Amsterdam — Amsterdam, NL

MSc in Data Science — 2020 – 2023

University of Trento — Trento, Italy

MSc in Machine Intelligence — 2019 – 2020

AIMS–AMMI — Kigali, Rwanda

BSc in Computer Science — 2013 – 2017

University of Gondar — Gondar, Ethiopia

SELECTED PUBLICATIONS

- **Mekonnen, K.A.**, Tang, Y., and de Rijke, M. *A Parametric Memory Head for Continual Generative Retrieval*. **SIGIR 2026, Full Paper Track**. doi
- **Mekonnen, K.A.**, Li, Y., Tang, Y., Lupart, S., and de Rijke, M. *Lost in Decoding? Reproducing and Stress-Testing the Look-Ahead Prior in Generative Retrieval*. **SIGIR 2026, Reproducibility Track**. doi
- **Mekonnen, K.A.**, Tang, Y., and de Rijke, M. *Lightweight and Direct Document Relevance Optimization for Generative Information Retrieval*. **SIGIR 2025, Full Paper Track**. doi
- **Mekonnen, K.A.**, Alemneh, E., and de Rijke, M. *Optimized Text Embeddings & Benchmarks for Amharic Passage Retrieval*. **ACL Findings 2025**. link
- Alemneh, Y.W., **Mekonnen, K.A.**, and de Rijke, M. *The Multilingual Curse at the Retrieval Layer: Evidence from Amharic*. **MeLLM @ ACL 2026**. ACL Anthology

SELECTED PROJECTS

Semantic Identifier Search for Recommendation — 2025 – Present

Research Project

- Investigated semantic IDs, trie-constrained generation, and centroid-guided beam search for sequential recommendation.
- Built FAISS-based retrieval and evaluation pipelines for recommendation experiments.

Spatiotemporal Mobility Analytics — 2023

MSc Project

- Built a spatiotemporal analytics pipeline for GPS and time-diary data, including trajectory cleaning, temporal aggregation, subgroup analysis, and interactive visualization. Code

Birth-rate Estimation Pipeline — 2023

MSc Project

- Developed an ML pipeline using Twitter and Eurostat data, including data collection, Flask-based labeling, BERT classification, and regression-based birth-rate estimation. Code

TEACHING & SERVICE

Teaching Assistant — 2024 – Present

University of Amsterdam

- TA for **Information Retrieval** and **Recommender Systems**; supported labs, assignments, projects, and exam grading.

Assistant Lecturer — 2017 – 2019

University of Gondar

- Lectured computer science courses and supervised labs, assignments, and undergraduate projects.

Academic Service & Mentoring — 2019 – Present

- Reviewer for **SIGIR** and **ECIR**; received **ECIR best reviewer recognition**.
- Mentor: **ELLIS ESSIR'24** , **AddisCoder** , and **MeetMentors.org** .

SELECTED ADDITIONAL WORK

- Lupart, S., **Mekonnen, K.A.**, Abbasiantaeb, Z., and Aliannejadi, M. *uva-irlab-conv at SemEval-2026 Task 8: Multi-Turn RAG with Learned Sparse Retrieval and Listwise Reranking*. **SemEval 2026**.
- **Mekonnen, K.A.** *Adv-KD: Adversarial Knowledge Distillation for Faster Diffusion Sampling*. arXiv.
- **Mekonnen, K.A.** *Balanced Face Dataset: Guiding StyleGAN for Labeled Synthetic Face Image Dataset for Underrepresented Group*. arXiv.