Updated September 2026

Amirhossein Roshandel

roshandel2004@gmail.com Website LinkedIn Github

Research Interests

Large Language Models (LLMs), Information Retrieval (IR), Personalization & User Modeling, Conversational Recommender Systems (CRS), LLM Evaluation & Reliability, Faithfulness & Evidence Grounding, Trustworthy AI

Education

University of Tehran **Tehran, Iran**

M.Sc in Software Engineering *Sep 2023–Present*

- **Thesis**: Personalization and Diversification of Recommendations in Conversational Recommender Systems with a Focus on User Preferences
- **Supervisor**: Dr. Azadeh Shakery
- **Expected defense**: January 2027
- **GPA**: 17.12/20

Ferdowsi University of Mashhad **Mashhad, Iran**

B.Sc in Computer Engineering *Sep 2018–Jun 2023*

- **GPA**: 18.31/20

Publications

- **AuDiPE: Behavior-Grounded User Simulation via Audited Dialogue-based Profile Enrichment for Conversational Recommender Systems**
  Manuscript in Revision
  A. Roshandel*, E. Shahabi*, M. Sadeghi, A. Shakery (*equal contribution)

Research Experience

University of Tehran **Tehran, Iran**

Graduate Researcher — Advisor: Dr. Azadeh Shakery *Sep 2023–Present*

- **M.Sc. Thesis**: Personalization and diversification in conversational recommender systems — modeling user preferences from dialogue and studying the trade-off between recommendations aligned with known preferences and maintaining diversity over the course of a conversation.
- **AuDiPE**: Developed a behavior-grounded framework for LLM-based user simulation and dialogue-based profile enrichment, with a focus on improving the reliability of inferred user profiles (manuscript in revision).
- Contributed to the design and execution of the evaluation protocol, including human annotation with ordinal labels, inter-annotator agreement using quadratic-weighted κ, and controlled ablation studies of individual pipeline components.
- Implemented and evaluated multi-stage LLM inference pipelines in Python and conducted experiments using ReDial and MovieLens 25M.

Teaching Assistant

- **Advanced Data Mining (Winter 2025–Present, University of Tehran)**: Chief TA (Winter 2026–Present) assisting Dr. Azadeh Shakery. Designed assignments on OLAP, data warehousing, and frequent pattern mining (Apriori, FP-Growth).
- **Intelligent Information Retrieval (Fall 2024–Present, University of Tehran)**: Assisted Dr. Azadeh Shakery. Developed projects on statistical language models for IR, query likelihood retrieval, feedback via KL-divergence EM, and smoothing methods (Jelinek-Mercer, Dirichlet prior, absolute discounting).
- **Statistical Inference (Fall 2025–Present, University of Tehran)**: Assisted Dr. Abolghasemi. Covered MLE, Bayesian inference, and hypothesis testing.
- **Social Networks (Fall 2025–Present, University of Tehran)**: Assisted Dr. Asadpour. Focused on affiliation networks and community detection.
- **Earlier teaching (Spring 2022)**: Artificial Intelligence (Ferdowsi University of Mashhad, with Dr. Abrishami and Dr. Harati) — assignments on search algorithms, heuristics, and automated planning; Python Workshop (Iranian Geophysical Society) — Python programming for master's students.

Experience

NeshanMaps **Mashhad, Iran**

Software Engineer - Gamification Team *Mar 2025–Mar 2026*

- Architected and deployed a Check-The-Fact (CTF) validation system, enabling daily validation of one-tenth of the data without human intervention.
- Integrated LLMs into comment moderation, cutting manual review time by 60%.
- Refactored leaderboard logic to eliminate 30-minute calculation delays, enabling real-time dynamic scoring.

Engineering Manager - Search Team *Jul 2024–Mar 2025 (concurrent)*

- Led a team of 4 developers, overseeing backend design for the Search Team.
- Designed a bidding/allocation optimization for ad placement using ML, doubling ad revenue and enhancing personalization.
- Designed a robust suggestion system to improve user engagement.
- Built dashboards and visualizations of search and product metrics for non-technical stakeholders (sales/product teams).

Software Engineer - Search Team *Aug 2021–Mar 2025*

- Redesigned and refactored the geocoding system, achieving a 3x increase in scalability and maintainability while reducing query latency by 15% and enhancing accuracy by 20%.
- Optimized the reverse geocoding pipeline, achieving a 10% improvement in processing speed and a 12% boost in location matching precision.
- Improved search CTR from 57% to 64% through targeted optimizations and advanced indexing strategies.
- Worked on search infrastructure during a period in which the product grew from 10 million to over 30 million users, with more than 3 million daily active users.

Selected Projects

Core LLM Coursework

- **Prompt Engineering Suite — Llama 3 + Unsloth**
  Stack: Llama 3.2 3B-Instruct, Unsloth (4-bit), Transformers, TRL, Datasets
  Designed and evaluated Zero-shot, Few-shot, Chain-of-Thought, Role-Play, and Least-to-Most prompting on GSM8K; analyzed impacts on reasoning accuracy and interpretability.
- **RAG Pipeline — LangChain + Phi-4 Mini**
  Stack: LangChain, Sentence Transformers, FAISS, Transformers, Hugging Face Hub
  Built a hybrid Retrieval-Augmented Generation system (TF-IDF + dense) with structured JSON outputs; improved factual grounding and response consistency in open-domain QA.
- **Text-to-SQL Graph Router — LangGraph + Gemini**
  Stack: LangGraph, LangChain Google GenAI, Pydantic, Python-dotenv
  Implemented difficulty-aware routing and ReAct-style schema reasoning; reduced invalid queries and improved accuracy on multi-table SQL generation.

Spring AI Workshop (Hands-on)

- *Stack*: Spring Boot, OpenAI API, Java, LangChain4j
- Developed hands-on exercises for integrating LLMs in Spring apps, covering prompt engineering, RAG, and agent workflows.
- Bridged backend engineering with LLM-based pipelines for applied AI developers.

Skills

- **Machine Learning**: PyTorch, Scikit-learn, Hugging Face `evaluate`, 4-bit quantized fine-tuning (Unsloth, PEFT)
- **LLMs**: LangChain, Unsloth, HF Trainer, PEFT, Datasets, TRL
- **Programming Languages**: Python, Java
- **Retrieval**: Sentence Transformers, ColBERT, BM25
- **Databases**: ElasticSearch, Lucene, lanceDB, Redis, PostgreSQL
- **Infrastructure**: Linux, Docker Compose, GitLab, Prometheus, Grafana, Sentry
- **Web Technologies**: Spring Boot
- **Tools**: JIRA, Confluence
- **Additional**: Code Reviewing, Repository Maintenance, Technical Communication, Teamwork

Languages

Persian: Native, **English:** Fluent