

Donya Rooein, PhD

(+39) 338 189 7214 | Milan, Italy | donya.rooein@gmail.com | linkedin.com/donya-rooein/ |
Languages: Farsi (native), English (C2), Italian (B1)

SUMMARY

Research Scientist and NLP Engineer with a scientific and entrepreneurial mindset and a diverse technological background. Over 7 years of experience in Data Science and Natural language Processing, and more than 10 years of programming experience in Python. Proven ability in leading data-driven research projects, collaborating within interdisciplinary teams, and managing projects efficiently. Seeking to utilize data to solve real-world problems with measurable impact.

SKILLS & EXPERTISE

- Data Science
- Machine Learning
- Natural Language Processing
- Software Engineering
- Data Engineering
- LLM Evaluations
- Entrepreneurship
- Project Management
- Collaboration

EXPERIENCE

Postdoctoral Researcher

Bocconi University (MilaNLP Lab)

Feb. 2023 – Now

- Designed, implemented, and optimized **IR/RAG evaluation pipelines** for multilingual **bias & fairness** across **3 dimensions** (cultural, gender, linguistic).
- Improved **LLM reasoning** and tutoring quality via prompt/program synthesis and retrieval grounding; defined KPIs (step-level accuracy, solution completeness, rubric-based scoring) and ran controlled evaluations.
- Co-authored **2** evaluation/robustness frameworks—*Co-DETECT* and *MSTS*—for edge-case discovery, multimodal safety testing, and annotator-disagreement modeling (probabilistic label aggregation, confidence-weighted scoring).
- Benchmarked LLMs over different NLP tasks such as: text simplification, summarization, sentiment analysis, emotion detection, hate speech, toxicity detection.

Visiting Researcher

ETH Zurich (AI Center)

Nov. 2024 – Feb.2025

Zurich, Switzerland

- Benchmarked **reasoning** and **pedagogical** capabilities of LLMs; built an evaluation harness for chain-of-thought fidelity, calibration error, and tutor feedback specificity with offline + human-in-the-loop testing.
- Developed **user-evaluation pipelines** (survey instrumentation, rubric-based grading, A/B comparisons) to validate adaptive/personalized learning; synthesized results into decision memos for model iteration.
- Collaborated in multidisciplinary teams to integrate **retrieval-augmented** tutoring prototypes into **production-ready** demos, coordinating data pipelines, prompts, and guardrails.

PhD Visiting Researcher

ETH Zurich (Center of Digital Health Interventions)

Oct. 2021 – May.2022

Zurich, Switzerland

- Researched and evaluated **digital-health chatbots** with emphasis on safety, appropriateness, and task success; defined acceptance criteria and created reusable test suites.
- Implemented a **data-driven chatbot** for online courses; integrated retrieval over course materials to ground answers and instrumented logs for **real-time** monitoring and **batch** analytics.

IBM Intern

IBM Research as a Service

June 2019– Sep.2019

Rome, Italy

- Developed and deployed conversational AI applications by integrating IBM Watson Assistant with complementary services (NLU, Discovery, Tone Analyzer) on the IBM Cloud platform.
- Designed and implemented chatbot workflows using Node-RED, leveraging Watson nodes for rapid prototyping and connecting with backend services.

Data Scientist

NazarBazaar Startup (acquired with Research On Mobile)

Nov. 2017 – Sep.2018

Tehran, Iran

- Generated and prioritized actionable solutions using an “80/20” approach to maximize business impact.
- Implemented innovative survey mechanisms to gather reliable data efficiently and affordably for businesses.
- Modeled user behavior in online surveys to uncover insights and guide strategic decision-making.

EDUCATION

| | |
|---|---------------------------------------|
| Politecnico di Milano <i>Ph.D. Computer Sciences</i> | Nov. 2018 – Oct 2022 Milan, Italy |
| University of Tehran <i>M.Sc. Information Technology Engg.</i> | Sep. 2013 – Sep. 2016 Tehran, Iran |
| Shiraz University of Technology <i>B.Sc. Information Technology Engg.</i> | Sep. 2008 – Sep. 2012 Shiraz, Iran |

SELECTED PUBLICATIONS

| | |
|---|-------------|
| PATS: Personality-Aware Teaching Strategies with Large Language Model Tutors | EACL 2026 |
| • D Rooein , S Chowdhury, M Eremeeva, Y Qin, D Nozza, M Sachan, D Hovy | |
| Can Reasoning Help Large Language Models Capture Human Annotator Disagreement? | EACL 2026 |
| • J Ni, Y Fan, V Zouhar, D Rooein , A Hoyle, M Sachan, M Leippold, D Hovy, E Ash | |
| Biased Tales: Cultural and Topic Bias in Generating Children’s Stories | EMNLP 2025 |
| • D Rooein , V Zouhar, D Nozza, D Hovy | |
| Teacher Demonstrations in a BabyLM’s Zone of Proximal Development for Contingent Multi-Turn Interaction | EMNLP 2025 |
| • S Salhan, H Gu, D Rooein , D Galvan-Sosa, G Gaudeau, A Caines, Z Yuan, P Buttery | |
| Measuring Gender Bias in Language Models in Farsi | ACL 2025 |
| • H Saffari, M Shafiei, D Rooein , D Nozza — <i>Proceedings of the 6th Workshop on Gender Bias in NLP</i> | |
| Co-DETECT: Collaborative Discovery of Edge Cases in Text Classification | EMNLP 2025 |
| • C Xiong, J Ni, Y Fan, V Zouhar, D Rooein , L Calvo-Bartolomé, A Hoyle, et al. — <i>System Demonstration</i> | |
| Can Large Language Models Capture Human Annotator Disagreements? | arXive 2025 |
| • J Ni, Y Fan, V Zouhar, D Rooein , A Hoyle, M Sachan, M Leippold, D Hovy — <i>under review ARR July</i> | |
| Exploring Subjective Tasks in Farsi: A Survey Analysis and Evaluation of Language Models | arXiv 2025 |
| • D Rooein , FM Plaza-del-Arco, D Nozza, D Hovy - under review ARR July | |
| Educators’ Perceptions of Large Language Models as Tutors | ACL 2025 |
| • SP Chowdhury, TJ Zhang, D Rooein , D Hovy, T Käser, M Sachan — <i>20th Workshop on Innovative Use of NLP for Building Educational Applications</i> | |
| Multilingual Performance Biases of Large Language Models in Education | ACL 2025 |
| • V Gupta, SP Chowdhury, V Zouhar, D Rooein , M Sachan — <i>20th Workshop on Innovative Use of NLP for Building Educational Applications</i> | |
| Can I Introduce My Boyfriend to My Grandmother? Evaluating LLMs on Iranian Social Norm Classification | NAACL 2024 |
| • H Saffari, M Shafiei, D Rooein , F Pierri, D Nozza — <i>findings NAACL</i> | |
| aCHAT-WF: Generating conversational agents for teaching business process models? Evaluating LLMs on Iranian Social Norm Classification | SoSym 2022 |
| • D Rooein , D Bianchini, F Leotta, M Mecella, P Paolini, B Pernici — <i>Journal of Software and Systems Modeling</i> | |
| Data-driven EDU chatbots | WWW 2019 |
| • D Rooein — <i>world wide web conference</i> | |