

IRINA SAPARINA

Edinburgh, United Kingdom ♦ I.Saparina@sms.ed.ac.uk ♦ saparina.github.io

PROFILE

- Member of Technical Staff at Cohere, working on agentic code generation. Ph.D. in Informatics from the University of Edinburgh, supervised by Prof. [Mirella Lapata](#).
- Expertise in code generation, semantic parsing, reasoning, ambiguity resolution, diverse generation. Developed test-time search techniques for code generation [8] and methods for handling ambiguity through explainable interpretations [5], reasoning about user intent [1], and diverse response generation [4].
- First author publications include NeurIPS Spotlight on ambiguity [6] and EACL Oral on language variation [7].

WORK EXPERIENCE

Cohere

Member of Technical Staff, CodeGen

May 2026 – present

Remote

Working on agentic code generation.

- Improving model performance on vibe coding (zero-to-one application building from natural language).
- Post-training and evaluation for code generation models.

Yandex Research

Research Assistant

Jul 2020 – Sep 2022

Moscow, Russia

Worked on semantic parsing and program synthesis with a focus on text-to-SQL:

- Developed test-time search techniques for generating better database queries guided by external criteria [8].
- Proposed training with intermediate question representations for translating questions into database queries. Achieved SOTA execution accuracy without costly expert annotation [9].

Samsung Laboratory at HSE University

Research Assistant

Jun 2018 – Jul 2020

Moscow, Russia

Worked on reinforcement learning for sequence-to-sequence models:

- Investigated imitation learning and cost-sensitive training to directly optimize task-specific metrics.
- Designed alternative learning objectives, improving performance in machine translation and code generation [10].

Laboratory of Programming Technologies, Moscow State University

Intern

Jun 2017 – Sep 2017

Moscow, Russia

Developed a Python tool for analyzing social media posts:

- Implemented multilingual keyword extraction using matrix factorization (SVD, ONMF).
- Improved ROUGE metric by 5% on real-world data.

EDUCATION

Ph.D. in Informatics, The University of Edinburgh

Advisor: [Mirella Lapata](#)

Oct 2022 – Sep 2026

Edinburgh, UK

Thesis: *From Pattern Matching to Intent Reasoning in Semantic Parsing*

Ph.D. student in Computer Science, HSE University (GPA: 9.5/10.0)

Advisor: [Anton Osokin](#)

Nov 2020 – Aug 2022

Moscow, Russia

Thesis: *Augmenting Semantic Parsing with Intermediate Representations and Search*

Master of Computer Science, HSE University (GPA: 9.5/10.0)

Advisor: [Anton Osokin](#)

Sep 2018 – Jun 2020

Moscow, Russia

Thesis: *Cost-sensitive training for autoregressive models*

TECHNICAL SKILLS

Languages: Python (expert), SQL, C/C++
Frameworks: PyTorch (expert), HuggingFace Transformers, vLLM
Tools & Infra: Git, Docker, Linux, Slurm, Kubernetes, L^AT_EX

AWARDS

Outstanding Reviewer at EMNLP 2024	2024
Scholarship of Advanced Doctoral Programme at HSE University	2020 - 2022
EMNLP Student Travel Grant	2021
NeurIPS Travel Grant	2019
Russian Academic Excellence Scholarship	2014-2020
Winner of SAS Data Analysis Hackathon	2017

TEACHING

Foundations of Natural Language Processing , University of Edinburgh	
Teaching Assistant	Spring 2025
Deep Learning , HSE University	
Lecturer	Spring 2022
Seminar Teacher	Fall 2020
Teaching Assistant	Spring and Fall 2019
Bayesian Methods of Machine Learning , HSE University and Moscow State University	
Teaching Assistant	Fall 2018
Summer School on Deep Learning and Bayesian Methods (Deep Bayes)	
Assistant	Summer 2018, Summer 2019

SERVICE

Reviewer: WiML @ NeurIPS 2019; EMNLP 2023; ACL ARR 2024 Feb (ACL 2024), Jun (EMNLP 2024), Feb 2025 (ACL 2025), Jan 2026 (ACL 2026); NeurIPS 2025; COLM 2026.
Mentor: Women in Machine Learning 2024-2025.

PUBLICATIONS

- [1] **Irina Saparina** and Mirella Lapata, “[Reasoning about Intent for Ambiguous Requests](#),” in COLM, 2026.
- [2] Miao Li, **Irina Saparina**, Alexander Gurung, and Mirella Lapata, “[Long-Context Reasoning Through Proxy-Based Chain-of-Thought Tuning](#),” in ACL, 2026.
- [3] Miao Li, Alexander Gurung, **Irina Saparina**, and Mirella Lapata, “[Who Gets Cited Most? Benchmarking Long-Context Language Models on Scientific Articles](#),” Under review, 2025.
- [4] Sergey Troshin, **Irina Saparina**, Antske Fokkens, and Vlad Niculae, “[Asking a Language Model for Diverse Responses](#),” in Workshop on Uncertainty-Aware NLP at EMNLP, 2025.
- [5] **Irina Saparina** and Mirella Lapata, “[Disambiguate First, Parse Later: Generating Interpretations for Ambiguity Resolution in Semantic Parsing](#),” in Findings of ACL, 2025.
- [6] **Irina Saparina** and Mirella Lapata, “[AMBROSIA: A Benchmark for Parsing Ambiguous Questions into Database Queries](#),” in NeurIPS Datasets and Benchmarks Track (Spotlight), 2024.

- [7] **Irina Sapparina** and Mirella Lapata, “[Improving Generalization in Semantic Parsing by Increasing Natural Language Variation](#),” in EACL (Oral), 2024.
- [8] Anton Osokin, **Irina Sapparina**, and Ramil Yarullin, “[Searching for Better Database Queries](#),” in Findings of EACL, 2023.
- [9] **Irina Sapparina** and Anton Osokin, “[SPARQLing Database Queries from Intermediate Question Decompositions](#),” in EMNLP, 2021.
- [10] **Irina Sapparina** and Anton Osokin, “[Cost-Sensitive Training for Autoregressive Models](#),” in Women in Machine Learning Workshop at NeurIPS, 2019.