

NATALIYA STEPANOVA

Email: n.p.stepanova@sms.ed.ac.uk

Personal Webpage: <https://stepanat.github.io/>

LinkedIn: <https://www.linkedin.com/in/nataliya-stepanova/>

EDUCATION

PhD Candidate in Natural Language Processing (*Distinction*) Sept. 2025 – Present

University of Edinburgh, Scotland

Topic: how language models encode and represent generalizations in language, especially generalizations about people (stereotypes) and in scientific reporting (over-exaggerating of claims)

MSc Speech and Language Processing (*Distinction*) Sept. 2021 – Aug. 2022

University of Edinburgh, Scotland

BSc Mathematics, BSc Computer Science (*Magna Cum Laude, GPA 3.98*) Aug. 2017 – May 2021

University of Maryland, College Park, USA

INDUSTRY EXPERIENCE

Software Development Engineer (SDE II, L5) Oct. 2022 – May 2025

Amazon (Alexa Text-to-Speech), Cambridge, England

- Worked on deploying Large Language and Speech models into production: optimized computation efficiency, latency, and throughput for real-time speech synthesis; designed algorithms for maintaining output speech quality; implemented tools for the automatic evaluation of synthetic voices (worked in Python and C++)
- Promoted from SDE I (L4) to SDE II (L5) in July 2024 (1 year 9 months after starting)

Software Engineering (Intern) June - Aug. 2020 and 2021

Capital One Financial Corporation, USA

- Worked on an ML-based SlackBot to automate question answering in Slack support channels (June-Aug. 2021)
- Designed algorithm using NLP (sentiment extraction) to process agent/customer call and chat interactions to create a “smart note” describing the status of the customer’s refinancing application (June-Aug. 2020)

SELECTED PUBLICATIONS (see [Google Scholar profile](#))

Stepanova, N., Titov, I., Allaway, E., Ross, B. (2026). The Missing Minimal Pair: Stereotype Evaluation in LLMs. <https://arxiv.org/pdf/2610.08747> (*under review*)

Diao, C., **Stepanova, N.**, Allaway, E. (2026). How broad is that claim? Mapping Generalisation in NLP Research. <https://arxiv.org/pdf/2609.14770> (*will appear at EMNLP 2026*)

Cilleruelo, G., **Stepanova, N.**, Allaway, E., & Birch, A. (2026). Concepts in terms of other Concepts: a method for the analysis of collective word use. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 48). <https://escholarship.org/content/qt2606z9wx/qt2606z9wx.pdf>

Stepanova, N., Ross, Björn. (2023). Temporal Generalizability in Multimodal Misinformation Detection. Proceedings of the EMNLP 2023 Workshop on Generalisability and Benchmarking in NLP (GenBench). <https://aclanthology.org/2023.genbench-1.6.pdf>

AWARDS

Marshall Scholarship (*Marshall Aid Commemoration Commission, London, England*) Sept. 2021 – Sept. 2022

Granted to up to 50 Americans per year to pursue graduate study in the UK (see [news report](#))

Banneker/Key Scholarship (*University of Maryland, College Park, Maryland, USA*) Aug. 2017 – May 2021

Merit-based full scholarship tenable for up to 4 years

Critical Language Scholarship, Korean (*United States Department of State, USA*) June 2019 – Aug. 2019

Fully funded 8-week intensive and immersive Korean language program based in South Korea

TECHNICAL TOOLS

- **Programming Languages:** Python, Java, C++, C, OCaml, Ruby, R, SQL
- **Natural Language Processing:** Large Language Models (LLM) optimization, Topic Modeling, Sentiment Extraction
- **Speech Technologies:** Automatic Speech Recognition (ASR), Text-to-Speech (TTS)
- **Machine Learning (ML):** Transformer, (bi-)LSTM, RNN, CNN, DNN, SVMs, K-means
- **Computer Science Fundamentals:** Data Structures, Complexity and Optimization, Algorithms
- **Other:** AWS, Creating/Using APIs (Postman), Databases (PostgreSQL), Agile Methodologies

LANGUAGES

- **English** (Native, Fluent)
- **Russian** (Native, Fluent)
- **French** (Advanced) – received International Baccalaureate (IB) French bilingual diploma in 2017
- **Korean** (Advanced-low) – determined by the ACTFL Oral Proficiency Interview in August 2019