




# Irina Bigoulaeva

 <https://ibigoula.github.io/> |  [irina.bigoulaeva@tu-darmstadt.de](mailto:irina.bigoulaeva@tu-darmstadt.de) |  Google Scholar

## About

---

Ph.D candidate at the Ubiquitous Knowledge Processing Lab at the Technische Universität Darmstadt in Germany, specializing in explainability of LLMs with a focus on post-training methods and mechanistic interpretability. Additionally has 3+ years experience with teaching and science communication to the public.

## Education

---

**Ph.D Candidate in Natural Language Processing** Oct 2021 - Present

*Ubiquitous Knowledge Processing Lab*

- **Supervisor:** Prof. Dr. Iryna Gurevych
- **Topic:** Explaining the acquisition of complex abilities in LLMs

**Master of Science in Computational Linguistics** Oct 2018 - Oct 2021

*Ludwig-Maximilians-Universität München*

- Minor in **Computer Science**
- Coursework: ML/Deep learning, PHPC, Knowledge Discovery in Databases
- **Supervisors:** Dr. Alexander Fraser, Dr. Viktor Hangya
- **Topic:** Cross-Lingual Transfer Learning for Hate Speech Detection

**Bachelor of Arts in Linguistics and Philosophy** Aug 2014 - May 2018

*University of Florida, USA*

- *Summa cum laude*
- Minor in **Mathematics**
- Coursework: Theoretical linguistics, syntax, philosophy of mind

## Work Experience

---

**Working Student** Jul 2021 - Nov 2021

*Centrum für Informations- und Sprachverarbeitung (CIS), LMU München*

- Trained and evaluated Transformer-based hate speech detection models. Improved performance in low-resource settings using cross-lingual transfer learning. Published a workshop paper at EACL 2021.

## Teaching Experience

---

**M.Sc Thesis Supervision** Mar 2024 - Present

*Ubiquitous Knowledge Processing Lab*

- **Topic:** *Code Pretraining for Improving State-Tracking Performance in Large Language Models*
- Main supervisor. Planned and developed the research topic.

**Seminar "Understanding LLMs"** Oct 2024 - Feb 2025

*Ubiquitous Knowledge Processing Lab*

- **Course topic:** Overview of Large Language Models and interpretability methods.
- Independently planned the curriculum for and taught a seminar of 30 students.

**Tutorial "INCEpTION: Efficient Text Annotation"** Jan 2023

*Université de Neuchâtel*

- Planned and led a 2-hour hybrid tutorial about text annotation using the INCEpTION annotation platform developed by the UKP Lab. The session had ca. 20 participants.

## M.Sc Thesis Supervision

Jun 2022 - Dec 2022

*Ubiquitous Knowledge Processing Lab*

- Topic: *Exploring Data Biases in Document-Level Natural Language Inference Datasets*
- Main supervisor. Planned and developed the research topic.

## Tutorial “Annotation and Modeling of Textual Data: Concepts and Tools”

March 2022

*Zürcher Hochschule für Angewandte Wissenschaften*

- Planned and led a two-day online tutorial about data annotation using the INCEpTION annotation platform developed by the UKP Lab. Each session was 4 hours long, with ca. 10 participants.

## Selected Publications

### The Inherent Limits of Pretrained LLMs: The Unexpected Convergence of Instruction Tuning and In-Context Learning Capabilities

Preprint, 2025

Irina Bigoulaeva, Harish Tayyar Madabushi, Iryna Gurevych

### Are Emergent Abilities in Large Language Models Just In-Context Learning?

ACL 2024

Sheng Lu\*, Irina Bigoulaeva\*, Rachneet Sachdeva, Harish Tayyar Madabushi, Iryna Gurevych

\* Equal first-author contribution

### Cross-Lingual Transfer Learning for Hate Speech Detection

EACL Workshop 2021

Irina Bigoulaeva, Viktor Hangya, Alexander Fraser

## Industry Experience

### Collaboration with Nexplora: Artificial Intelligence in Construction (AICO)

Oct 2021 - Dec 2024

- Researched on creating an LLM-based system for legal NLP.
- Helped integrate the INCEpTION annotation platform into a larger data annotation pipeline. Independently developed conversion scripts from the INCEpTION-native format to the CSV format.
- Researched on developing LLM-based chatbots in accordance with the company’s specific needs.
- Contributed a literature search and a project outcome report for the final whitepaper at the end of the project.

## Talks and Events

### Article: Deepseek-Modelle auf dem Prüfstand (en. “DeepSeek Models Tested”)

May 2025

- Co-wrote an article in German for the university newspaper. Conducted experiments with the DeepSeek reasoning models, which highlighted that these models possess fundamentally similar weaknesses to base and instruction-tuned LLMs, despite their complex post-training procedures.

### Invited Talk: European Kidney Summer School 2023 (EUKISS)

Jul 2023

- Gave an invited talk at the session “Bioinformatics, advanced image analysis, and AI”, which was attended mostly by specialists in medicine.
- **Topic:** The limits and possibilities of ChatGPT and GPT-4 for the medical field.

## Technical Skills

**Programming:** Python,  $\LaTeX$ , C++ (basic)

**Frameworks:** Huggingface, PyTorch, NNSight, TransformerLens, vLLM, Slurm

## Languages

**English:** native

**German:** C2