

Valeriya Zelenkova

PHD RESEARCHER

☎ +49 176 20994053 | ✉ vzelenkova03@gmail.com | 🌐 valeriya-zelenkova

Researcher with a cognitive science background, working on interpretability across biological and artificial neural networks. Currently completing a PhD on visual cortex circuits and contributing to LLM interpretability research through independent projects.

AI Safety Engagement

Supervised Program for Alignment Research (SPAR)

Remote

MENTEE (MENTOR: GABRIELE SARTI)

Feb. 2026 - May 2026

- Mechanistic interpretability project on how implicit user attributes are internally represented and retrieved across multi-turn conversations.
- Created a dataset of 7500 controlled multi-turn conversations on Llama-3.3-70B-Instruct.
- Combined linear probes on residual-stream activations with activation patching (via insight on NDIF) and LLM-as-judge behavioral evaluation.
- Co-authored final research report.
- Was selected to deliver an oral presentation at SPAR Demo Day.

BlueDot Impact

Remote

TECHNICAL AI SAFETY COURSE

Jan. 2026 - Feb. 2026

- Completed cohort-based course on technical AI safety foundations, threat models, and current alignment research directions.

European Network for AI Safety (ENAIIS)

Remote

AI SAFETY COLLAB

Aug. 2025 - Oct. 2025

- Completed 9-module course (~50 hours) covering AI capabilities, risks, interpretability, evaluation, and governance.
- Studied interpretability methods including sparse autoencoders, steering vectors, and dictionary learning.

Experience

University of Bremen, Cognitive Neurophysiology Group

Bremen, Germany

PHD RESEARCHER

Nov. 2021 - present

- Apply advanced computational methods to large-scale electrophysiology recordings: signal processing, statistical analysis of high-dimensional time-series data, and time-shift dependent gamma phase coherence analysis between visual areas V1 and V4.
- Work with shared experimental datasets and analysis pipelines across multiple labs in an EU-wide research consortium.

Stepik

Moscow, Russia

COURSE MANAGER AND TUTOR

July 2019 - Feb. 2021

- Ran an online Python course across 3 cohorts of 10-15 students each: oversaw student progress, scheduled classes, and handled course operations.

HSE University, Center for Language and Brain

Moscow, Russia

RESEARCH ASSISTANT

Jan. 2015 - May 2020

- Designed and managed human subjects research studies with healthy participants and patients with language disorders.
- Co-authored scientific publications and presented research at international conferences.

Selected Publications

- Malyutina S., **Zelenkova V.** (2020). Verb argument structure effects in aphasia are different at single-word versus sentence level. *Aphasiology*, 34(4), 431-457.
- Zyryanov A., **Zelenkova V.**, Malyutina S., et al. (2020). The Contributions of the Arcuate Fasciculus Segments to Language Processing: Evidence from Brain Tumor Patients. *The Russian Journal of Cognitive Science*, 6(1), 25-37.
- Malyutina S., **Zelenkova V.**, Buivolova O., Oosterhuis E.J., Zmanovsky N., Feurra M. (2018). Modulating the interhemispheric balance in healthy participants with transcranial direct current stimulation: No significant effects on word or sentence processing. *Brain and Language*, 186, 60-66.

Education

University of Trento and SISSA

M.Sc. IN COGNITIVE SCIENCE (GRADE: 110 CUM LAUDE)

Trento | Trieste, Italy

Sept. 2019 - Oct. 2021

- Coursework in computational methods for neuroscience, fMRI and electrophysiological data analysis.
- Thesis: "Coding of Visual Textures Defined by Local Multipoint Correlations in Rat Visual Cortex".

National Research University Higher School of Economics

B.A. IN COMPUTATIONAL LINGUISTICS (GPA: 9.22 / 10), MINOR IN MACHINE LEARNING AND DATA SCIENCE

Moscow, Russia

Sep. 2014 - June 2018

- ML coursework: supervised learning, modern ML methods, kernel methods, statistical analysis.
- Technical foundations: algorithms, databases, linear algebra, calculus, probability theory.

Additional Research Experience

ATLAS Neuro

VISITING RESEARCHER

Freiburg, Germany

Aug. 2023

- Tested a novel electrode insertion system; learned about probe fabrication in the cleanroom.

University of Fribourg, Dept. of Medicine, Neuro- and Movement Sciences

VISITING RESEARCHER

Fribourg, Switzerland

Sept. 2022 - Nov. 2022

- Analyzed brain recordings from optogenetic and micro-stimulation experiments.

Skills

Programming: Python · R · Matlab

ML / Data Science: numpy · pandas · scikit-learn · scipy · matplotlib

Interpretability: nnsight · linear probing · activation patching · LLM-as-judge evaluation

Methods: Signal processing · statistical analysis of high-dimensional data · experimental design · scientific writing

Tools: Jupyter Notebooks · Google Colab · LaTeX · git · GitHub

Languages: Russian (native) · English (C2) · German (C1) · Italian (A2)

Awards and Recognition

2021-2024: Marie Skłodowska-Curie Actions, Early Career Researcher Scholarship (competitive EU research funding)

2019-2021: International School for Advanced Studies (SISSA) and University of Trento, Excellence Scholarship (Italy)

2019: Young Faculty Support Program, Excellence Program for Young Researchers (Moscow, Russia)

References

Gabriele Sarti

SPAR MENTOR

gabriele.sarti996@gmail.com

Northeastern University