Elena Labzina, PhD

Research Engineer / Applied ML Scientist

LLM Model Behavior · Evaluation · Production ML

Zürich, Switzerland
elenalabzina@wustl.edu
elenalabzina.net
GitHub
LinkedIn
Swiss C permit · no sponsorship needed

Profile

Research engineer and applied scientist with experience in production ML at Microsoft and large-scale evaluation at Google. Expertise spans transformer fine-tuning, experimentation, and causal inference.

Currently leading causal **AI research and architecture** at Kanvaz AI Labs while independently investigating **how AI interactions affect human reasoning, judgment, and agency**. This work includes *When AI Agrees*, a preregistered randomized study of LLM sycophancy, and open-source evaluation tools. PhD in Political Science and MA in Statistics from Washington University in St. Louis, with earlier degrees in applied mathematics, computer science, and economics.

Research & Engineering Experience

Independent Research & ML Engineering — *Zürich, Switzerland*

LLM Behavior, Human–AI Interaction & Evaluation (part-time) — 2026 – Present

- Designed and preregistered **When AI Agrees**, a randomized human study investigating whether sycophantic LLM interactions influence reasoning and how their effects vary across users (OSF, under embargo).
- Built **Controllable Sycophancy Evals**, a public Python framework for synthetic users with controlled reasoning structures while holding topic, belief, and confidence fixed; separates reasoning plans, language realization, constraint checks, and stage-blind evaluation.
- Calibrated the research workflow with anonymized human sentence-completion responses; implemented six-dimensional behavioral scoring and completed matched-condition held-out probes, documenting limitations and linguistic confounds.
- Public research code, methodology, and validation notes.

Kanvaz AI Labs — *Zürich, Switzerland*

Head of Causal AI/ML (part-time) — 2026 – Present

- Lead architectural and methodological design for enterprise causal AI, using **structural causal models, interventions, and counterfactual reasoning** for supply-chain decisions.
- Specify identification assumptions, constraints, and evaluation criteria for auditable LLM-assisted decision workflows.

Google — *Zürich, Switzerland*

Research Data Scientist, Ads Metrics — 2025 – 2026

- Designed measurement and evaluation methodology for large-scale production ranking and targeting systems: metric validity, sampling, cohort construction, controlled comparisons, and model-launch analyses.
- Used experimentation and causal inference to examine **heterogeneous treatment effects** across users and advertisers, distinguishing changes from measurement bias.
- Built reproducible **Python and SQL** pipelines to diagnose metric bias and data-quality problems, supporting production launch decisions.

Microsoft, Commercial Software Engineering — *Zürich, Switzerland*

Data & Applied Scientist / ML Engineer — 2020 – 2024

- Built and evaluated production ML systems across **10+ engagements**: NLP, retrieval and ranking, LLM/RAG, transformer fine-tuning, and debugging.
- Fine-tuned and benchmarked **BERT and DistilBERT** for multimodal video classification using transcripts, OCR, and metadata; DistilBERT retained approximately **95% of BERT's predictive performance** with **60% lower inference latency**.
- Improved a production retrieval system for a global media organization, reducing latency by **40%** and increasing engagement by **25%**.
- Developed a GenAI-assisted healthcare document pipeline, delivering about **5× faster** processing and **50% higher extraction accuracy**, supported by systematic quality evaluation.

Constructor Academy — *Zürich, Switzerland*

Consultant & Instructor, GenAI and Statistics (part-time) — 2022 – 2024

- Taught ML, deep learning, NLP, generative AI, and statistical experimentation; mentored students on model development and evaluation.

ETH Zürich — *Zürich, Switzerland*

Postdoctoral Researcher, Data Science — 2018 – 2020

- Co-authored empirical research on media influence, combining **NLP text similarity and instrumental-variable methods** to distinguish effects on political framing from changes in topics or measured partisanship.
- Built reproducible NLP pipelines and tested measurement choices, specifications, and robustness.

Washington University in St. Louis *St. Louis, MO, USA*

Doctoral Researcher & Instructor 2012 – 2018

- Co-developed a **game-theoretic model of strategic behavior** and characterized its equilibria for research published in the *American Journal of Political Science*.
- Built Python/AWS pipelines to analyze more than **22 million Wikipedia revisions** and large social-media and news corpora; developed contextual text-similarity methods to investigate information manipulation.
- Appointed **Adjunct Professor of Statistics (2017)**; also served as a teaching and research assistant.

Central European University *Budapest, Hungary*

Lead Data Engineer, Climate & Energy Policy 2011 – 2012

- Built climate and energy-policy data pipelines, combining NASA geospatial data, GIS, emissions models, and economic scenarios.
- Co-authored *Best Practice Policies for Low Carbon & Energy Buildings* (GBPN, 2012; **119 Google Scholar citations**, Oct. 2026).

Technical Skills

- **Programming & ML:** Python, PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, NumPy, pandas, SQL, R, C++, C#.
- **Models & evaluation:** Transformer fine-tuning, NLP, RAG, information retrieval, controlled-user simulation, LLM behavioral evaluation, benchmark and rubric design, robustness and failure analysis, production ML.
- **Causal & experimental methods:** Structural causal models, DAGs, counterfactual inference, interventions, causal identification, A/B testing, treatment effects, heterogeneous effects, uncertainty quantification, experimental design.
- **Infrastructure:** AWS, Azure, Docker, Kubernetes, MLflow, Git, data pipelines, model deployment, reproducible experimentation.

Education

Washington University in St. Louis (St. Louis, USA) *2018*

PhD, Political Science
MA, Statistics
MA thesis: modeling restricted choice sets in multinomial logit.

Central European University (Budapest, Hungary) *2011*

MA, Political Science, with Merit
Primary specialization: causal inference and quantitative research design.

New Economic School (Moscow, Russia) *2010*

MA, Economics — Data Analysis

Lomonosov Moscow State University (Moscow, Russia) *2007*

BSc, Applied Mathematics & Computer Science

Selected Publications & Research

[J1] J. Barceló and **E. Labzina**. "Do Islamic State's Deadly Attacks Disengage, Deter, or Mobilize Supporters?" *British Journal of Political Science*, 2020.

[J2] O. V. Chyzh and **E. Labzina**. "Bankrolling Repression? Modeling Third-Party Influence on Protests and Repression." *American Journal of Political Science*, 2018.

[D1] **E. Labzina**. "Bankrolled Elites, Controlled State Television, and Payroll Trolls on Wikipedia: Examining Russia's Strategies in Manipulation of Opinions in the context of Ukrainian Revolution in 2014." PhD dissertation, Washington University, 2018.

[B1] **E. Labzina** and N. Schofield. "Application of the Variable Choice Logit Model to the British General Election of 2010." In *The Political Economy of Governance*, Springer, 2015.

[J3] **E. Labzina**. "Cost and Benefits of Natural Experiments in Political Science." *CEU Political Science Journal*, 2011.

Additional Training & Activities

- **Advanced ML training:** Deep RL (ETH Zürich); LLMs and RAG (DeepLearning.AI).
- **Invited technical talk:** *Revolutionizing ML Pipelines for Videos: LLMs and Zero-Shot Learning*, Constructor Learning NLP Event, Munich, 2023.
- **Technical writing:** Creating Your First Search Engine from Scratch , technical article and accompanying code, 2023.
- **Research community and teaching:** Co-organized **SICSS Zürich (2019)** at ETH; taught machine learning and computational methods.
- **Academic presentations:** Research presentations at PolMeth (MIT), APSA, SPSA, and Visions in Methodology; invited CEU alumni talk (2025), and a Washington University professionalization panel (2026).
- **Peer review:** 25+ manuscript reviews, including *APSR*, *AJPS*, *JPR*, and *JCR*.
- **Languages:** Russian (native), English (fluent), German (intermediate), Spanish (basic).