DENIZ BAYAZIT

PhD Candidate

deniz.bayazit@epfl.ch | bayazitdeniz.github.io | bayazitdeniz | bayazit-deniz | Swiss • Turkish

RESEARCH INTERESTS

My research focuses on natural language processing and artificial intelligence, with a particular interest in the underlying mechanisms and behavior of large language models (LLMs). I am also interested in understanding the dynamics of LLM training and identifying factors that influence models' generalization and decision-making processes.

EDUCATION

PhD in Computer Science — Sep 2021 — Present
EPFL — École Polytechnique Fédérale de Lausanne — Lausanne, Switzerland
- **Depth area:** AI/NLP
- **Coursework:** Reinforcement Learning, Machine Learning, Topics in NLP

MSc & BSc in Computer Science — Sep 2016 — May 2021
Brown University — Providence, RI, USA
- **Graduate:** AI/ML Pathway
- **Undergraduate:** AI/ML and Data Science Pathways with Honors
  Thesis: *Generalizing Natural Language Instruction Following to Aerial Robots and Arbitrary Environments*
- **Coursework:** AI, DL, RL, Reintegrating AI, Language Processing in Humans and Machines, Special Topics in Computational Linguistics & Computer Vision, Lexical Semantics, Introduction to Linguistics, Logic for Systems

RESEARCH EXPERIENCE

Doctoral Research Assistant — Sep 2021 — Present
EPFL NLP — Lausanne, Switzerland
Advisor: Antoine Bosselut [PI Website] [EPFL NLP Lab]

Undergraduate & Graduate Research Assistant — Jun 2018 — May 2021
H2R & LUNAR Laboratory — Providence, RI, USA
Advisor: Stefanie Tellex [PI Website] [H2R Lab], Ellie Pavlick [PI Website] [LUNAR Lab]

SELECTED PUBLICATIONS & PREPRINTS

- *Crosscoding Through Time: Tracking Emergence & Consolidation Of Linguistic Representations Throughout LLM Pretraining.* **arXiv 2025.**
  D Bayazit, A Mueller, A Bosselut. [PDF] [Code]
- *Discovering Knowledge-Critical Subnetworks in Pretrained Language Models.* **EMNLP 2024.**
  D Bayazit, N Foroutan, Z Chen, G Weiss, A Bosselut. [PDF] [Code] [Video] [Poster]
- *Could ChatGPT get an Engineering Degree? Evaluating Higher Education Vulnerability to AI Assistants.* **PNAS 2024.**
  B Borges*, N Foroutan*, **D Bayazit***, A Sotnikova* et al. [PDF] [Code]
- *MEDITRON-70B: Scaling Medical Pretraining for Large Language Models.* **arXiv 2023.**
  Z Chen et al.
  DAIR.AI Top ML Papers of the Week. [PDF] [Code]
- *PeaCoK: Persona Commonsense Knowledge for Consistent and Engaging Narratives.* **ACL 2023.**
  S Gao, B Borges, S Oh, **D Bayazit**, S Kanno, H Wakaki, Y Mitsufuji, A Bosselut.
  Outstanding Paper Award. [PDF] [Code] [Video]
- *Spatial Language Understanding for Object Search in Partially Observed Cityscale Environments.* **RO-MAN 2021.**
  K Zheng, **D Bayazit**, R Mathew, E Pavlick, S Tellex. [PDF] [Code] [Website] [Video]
- *Grounding Language to Landmarks in Arbitrary Outdoor Environments.* **ICRA 2020.**
  M Berg*, **D Bayazit***, R Mathew, A Rotter-Aboyoun, E Pavlick, S Tellex. [PDF] [Data] [Video]

TECHNICAL SKILLS

Programming: Python *[PyTorch, TensorFlow, scikit-learn, pandas, transformers, …]* (fluent); C, Scala, Java (familiar)
Operating Systems: Linux, macOS
Version Control & Cloud: Git, Bitbucket, Docker
Miscellaneous: Vim, Amazon Mechanical Turk, SQL, Excel, LaTeX

SELECTED RESEARCH PROJECTS

Discovering Knowledge-Critical Subnetworks in Pretrained Language Models — **PI: Antoine Bosselut**

We explore whether pretrained language models contain *knowledge-critical* subnetworks—sparse subgraphs whose removal suppresses specific memorized knowledge. We propose a multi-objective differentiable masking technique for weights and neurons to discover these subnetworks, enabling precise removal of targeted knowledge while preserving most of the model's original functionality.

Could ChatGPT get an Engineering Degree? Evaluating Higher Education Vulnerability — **PI: Antoine Bosselut**

We examine the challenges of AI assistants in higher education through the lens of vulnerability. We assess this by compiling a dataset of assessment questions from 50 EPFL STEM courses and evaluating how well GPT-3.5 and GPT-4 can answer them. Our findings show the need to rethink assessment design in response to advances in GenAI.

Grounding Language to Landmarks in Arbitrary Outdoor Environments — **PI: Stefanie Tellex & Ellie Pavlick**

This work tackles the challenge of generalizing robot instructions to new environments. We propose a framework that parses landmark references, assesses semantic similarities in a map, and converts natural language commands into drone motion plans. This enables robots to follow commands in unfamiliar environments, allowing untrained users to control them in large, outdoor areas using unconstrained natural language.

HONORS & AWARDS

- Teaching Assistant Award (Spring 2023)
- *Honorable Mention*, Computing Research Association — Outstanding Undergraduate Researcher Award (2020)
- Interdisciplinary Team Undergraduate Teaching and Research Award (Summer 2019, Fall 2019)

PROFESSIONAL SERVICE

Reviewer: ACL ARR Jul 2025 • ACL ARR May (EMNLP 2025) • ACL ARR Oct (NAACL 2024) • ACL ARR Jun (EMNLP 2024) • CSRR ACL 2022 Workshop • ICRA 2021 • ML-RSA NeurIPS 2020 Workshop • ICRA 2020

Mentorship: IC Buddy Program @EPFL 2022 • Out in CS @Brown 2021 • Women in CS @Brown 2020

TEACHING EXPERIENCE

Graduate TA — Modern NLP — Spring 2023, 2024
EPFL — Lausanne, Switzerland
Instructor: Antoine Bosselut. **Course Code:** CS 552. [Website]

Graduate TA — Introduction to NLP — Fall 2022, 2023, 2024
EPFL — Lausanne, Switzerland
Instructor: Jean-Cédric Chappelier, Martin Rajman, Antoine Bosselut. **Course Code:** CS 431. [Website]

Graduate TA — Language Processing in Humans and Machines — Spring 2021
Brown University — Providence, RI, USA
Instructor: Ellie Pavlick, Roman Feiman. **Course Code:** CSCI 2952I/CLPS 1850.

Undergraduate TA — Artificial Intelligence — Fall 2019
Brown University — Providence, RI, USA
Instructor: George Konidaris. **Course Code:** CSCI 1410.

LANGUAGES

- **Fluent:** English, French, Turkish
- **Elementary:** Spanish, Japanese