

Yonatan Slutzky

slutzky1@mail.tau.ac.il | yonislutzky.com | linkedin.com/in/yonis-slutzky | github.com/YoniSlutzky98

Research Interests

Deep Learning, Learning Theory, Sequence Models, Learning for Control

Education

Ph.D. (Direct Track), School of Computer Science & AI, Tel Aviv University 2023 – Present

- Advisor: Professor Nadav Cohen
- Topic: Theoretical and algorithmic foundations of deep learning
- Current GPA: 98/100

B.Sc. in Computer Science, Statistics and Operations Research, Tel Aviv University 2020 – 2023

- GPA: 95/100

Research Experience

Visiting Researcher, Long Program on Theoretical Advances in Reinforcement Learning and Control, Institute for Mathematical and Statistical Innovation (UChicago) March 2026 – June 2026

Honors and Awards

- US-Israel Binational Science Foundation (BSF) Prof. Rahamimoff Travel Grant for Young Scientists (2026)
- Israel AI Safety Research Prize (2025)
- Princeton Machine Learning Theory Summer School (2024, 1/80 Ph.D. students selected from 500+ worldwide)
- Tel Aviv University Center for AI and Data Science (TAD) Direct Ph.D. excellence program (2024)
- Summa cum laude graduate of Tel Aviv University B.Sc. Statistics and Operations Research (2023)
- Magna cum laude graduate of Tel Aviv University B.Sc. Computer Science (2023)
- Tel Aviv University School of Mathematical Sciences Award for outstanding undergraduate achievements (2023)
- Tel Aviv University Faculty of Exact Sciences Dean's honors list (2021, 2023)
- Tel Aviv University School of Computer Science & AI "First Steps in Research" program (2023)

Publications

Preprints

- Yonatan Slutzky, Yotam Alexander, Tomer Slor, Yoav Nagel and Nadav Cohen. 'Why Does Agentic Safety Fail to Generalize Across Tasks?'. *arXiv preprint arXiv:2605.06992*, 2026

Conference Proceedings

- Yotam Alexander, Yonatan Slutzky, Yuval Ran-Milo and Nadav Cohen. 'Do Neural Networks Need Gradient Descent to Generalize? A Theoretical Study'. *Conference on Neural Information Processing Systems (NeurIPS) 2025*
- Yonatan Slutzky[†], Yotam Alexander[†], Noam Razin and Nadav Cohen. 'The Implicit Bias of Structured State Space Models Can Be Poisoned with Clean Labels'. *Conference on Neural Information Processing Systems (NeurIPS) 2025, Spotlight Track (top 3%)*
- Nir Endy[†], Idan Daniel Grosbard[†], Yuval Ran-Milo[†], Yonatan Slutzky[†], Itay Tshuva, Raja Giryes. 'Mamba Knockout for Unraveling Factual Information Flow'. *Association for Computational Linguistics (ACL), 2025, SAC Highlight Track (top 2%)*

[†]Equal Contribution

Teaching

- Teaching assistant, Intro to Machine Learning, School of Computer Science & AI, Tel Aviv University (2025–present)
- Teaching assistant, Workshop on Deep Learning, School of Computer Science & AI, Tel Aviv University (2024–present)
- Teaching assistant, Foundations of Deep Learning, School of Computer Science & AI, Tel Aviv University (2024–present)
- Teaching assistant, Statistical Learning, School of Mathematical Sciences, Tel Aviv University (2023)

Service

Reviewer

- *International Conference on Machine Learning (ICML)*
- *Journal of Machine Learning Research (JMLR)*

Talks

- *Institute for Mathematical and Statistical Innovation (IMSI) Workshop on New Directions in Reinforcement Learning and Control, Chicago, IL, USA, 2026*
- *Israel Computer Vision Day, Tel Aviv, Israel, 2026*
- *Reichmann University AI-NLP Seminar, Herzliya, Israel, 2025*
- *ELLIS Deep Models and Optimization Group Seminar, Tübingen, Germany, 2025*
- *LMU Mathematical Data Science and Artificial Intelligence Group Seminar, Munich, Germany, 2025*