Theo X. Olausson

theoxo [at] csail [dot] mit [dot] edu
https://people.csail.mit.edu/theoxo/
https://scholar.google.com/citations?user=e7K3ZagAAAAJ
https://github.com/theoxo
32 Vassar Street, Cambridge, MA 02139

Education

09/2021 – Present **Massachusetts Institute of Technology** – Cambridge, MA
PhD in Electrical Engineering and Computer Science
Advisor: Armando Solar-Lezama, Distinguished Professor of Computing, COO of CSAIL
Minor Subject: Mathematical Statistics

09/2016 – 05/2021 **University of Edinburgh** – Edinburgh, UK
Master of Informatics
Advisor: Vijay Nagarajan, Reader (now Professor at U. of Utah)
Rank in Cohort: 1

Experience

09/2025 – 02/2026 **Apple**, Paris, France
Research Intern, Machine Learning Research.
Reinforcement learning and masked diffusion models (MDMs). Two (co-)first-author ICML'26 papers; engineering on large-scale multimodal MDM pre-training project (in review).

05/2024 – 04/2025 **Flatiron Institute (Simons Foundation)**, New York, NY
Pre-Doctoral Fellow → Guest Researcher, Center for Computational Mathematics.
Applications of optimal transport to long-context vision and language models.

05/2022 – 08/2022 **Microsoft Research**, Redmond, WA
Research Intern, Deep Learning.
Early work on self-correcting code language models. Led to first-author ICLR'24 paper.

Awards & Fellowships

2025–2026 **Expert/Top/Gold Reviewer**, TMLR'25, NeurIPS'25, ICML'26

2023 **Outstanding Paper Award**, EMNLP'23

2022 **Top Picks from the 2022 Computer Architecture Conferences**, IEEE Micro

2021 **Presidential Fellowship**, Massachusetts Institute of Technology

Master of Informatics Class Prize, University of Edinburgh

Outstanding Master's Dissertation, University of Edinburgh

2020 **ICSA Studentship**, University of Edinburgh

Outstanding Bachelor's Dissertation, University of Edinburgh

Leadership & Mentorship

2025–2026 **Lead Organizer**, VerifAI: The Workshop on AI Verification in the Wild, ICLR'25 and '26.

2023 **Co-President**, EECS Graduate Student Association, MIT.

2022–2023 **Mentor**, EECS Graduate Application Assistant Program (GAAP), MIT.

Preprints & In Submission

∗ denotes equal contribution.

2026 **Conformal Language Modeling via Posterior Sampling.** *In review.* (N. Emmenegger*, **TX. Olausson***, A. Solar-Lezama, C. Podimata.)

Searching Effectively with Latent Sketching. *In review.* (L. Hernandez Cano*, **TX. Olausson***, N. Fulton, A. Solar-Lezama.)

A Tale of Two Temperatures: Simple, Efficient, and Diverse Sampling from Diffusion Language Models. *In review; electronic preprint available.* (**TX. Olausson***, M. Jazbec*, X. Wang, A. Solar-Lezama, CA. Naesseth, S. Mandt, E. Nalisnick.) arXiv:2604.09921

The Design Space of Tri-Modal Masked Diffusion Models. *In review; electronic preprint available.* (L. Béthune, V. Turrisi, BK. Mlodozeniec, …, **TX. Olausson**, …, J. Ramapuram.) arXiv:2602.21472

Conference Publications

2026 **Amortizing Maximum Inner Product Search with Learned Support Functions.** In *The Forty-Third International Conference on Machine Learning (ICML 2026).* (**TX. Olausson**, J. Monteiro, M. Klein, M. Cuturi.) arXiv:2603.08001

Learning Unmasking Policies for Diffusion Language Models. In *The Forty-Third International Conference on Machine Learning (ICML 2026).* (M. Jazbec*, **TX. Olausson***, L. Béthune, P. Ablin, M. Kirchhof, J. Monteiro, V. Turrisi, J. Ramapuram, M. Cuturi.) arXiv:2512.09106 — Oral & Spotlight

2024 **The Counterfeit Conundrum: Can Code Language Models Grasp the Nuances of Their Incorrect Generations?** In *Findings of the 62nd Annual Meeting of the Association for Computational Linguistics (Findings: ACL 2024).* (A. Gu, WD. Li*, N. Jain*, **TX. Olausson***, C. Lee*, K. Sen, A. Solar-Lezama.) arXiv:2402.19475

Is Self-Repair a Silver Bullet for Code Generation? In *The Twelfth International Conference on Learning Representations (ICLR 2024).* (**TX. Olausson**, JP. Inala, C. Wang, J. Gao, A. Solar-Lezama.) arXiv:2306.09896

LILO: Learning Interpretable Libraries by Compressing and Documenting Code. In *The Twelfth International Conference on Learning Representations (ICLR 2024).* (G. Grand, L. Wong, M. Bowers, **TX. Olausson**, JB. Tenenbaum, J. Andreas.) arXiv:2310.19791

2023 **LINC: A Neuro-Symbolic Approach for Logical Reasoning by Combining Language Models with First-Order Logic Provers.** In *The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023).* (**TX. Olausson***, A. Gu*, B. Lipkin*, C. Zhang*, A. Solar-Lezama, JB. Tenenbaum, R. Levy.) arXiv:2310.15164 — Outstanding Paper

Top-Down Synthesis For Library Learning. In *The 50th ACM SIGPLAN Symposium on Principles of Programming Languages (POPL 2023).* (M. Bowers, **TX. Olausson**, L. Wong, G. Grand, JB. Tenenbaum, K. Ellis, A. Solar-Lezama.) arXiv:2211.16605

2022 **HeteroGen: Automatic Synthesis of Heterogeneous Cache Coherence Protocols.** In *The 28th IEEE International Symposium on High-Performance Computer Architecture (HPCA 2022).* (N. Oswald, V. Nagarajan, D. Sorin, V. Gavrielatos, **T. Olausson**, R. Carr.) — IEEE Micro Top Picks

Skills

Programming Languages: Python, Rust, OCaml, Haskell, C/C++

Tools and Frameworks: PyTorch, JAX, HuggingFace ecosystem, distributed training (DP/TP/PP/FSDP), Slurm, Git, Linux

Languages: English (fluent), Swedish (native), French (conversational), Italian (basic)