Oliver Sieberling

osieberl@mit.edu | Homepage | Google Scholar | GitHub

Education

Massachusetts Institute of Technology — Cambridge, USA
Ph.D. Computer Science — *Sep. 2025 – Present*
- Advisor: Yoon Kim

ETH Zurich — Zurich, Switzerland
B.Sc. Computer Science — *Sep. 2021 – Aug. 2025*
- Thesis: *Dynamic Monotone Functions on the (1+1)-Evolutionary Algorithm: Drift Minimization Does Not Maximize Hardness*. Advised by Johannes Lengler, Institute of Theoretical Computer Science.
- Bronze Medal, ICPC SWERC 2024.

Princeton University — Princeton, USA
Departmental Exchange Program — *Sep. 2023 – Jan. 2024*

Experience

MIT-IBM Watson AI Lab — Cambridge, USA
Research Intern — *June 2026 – Aug. 2026*
- Worked on improving linear attention, with Bharat Runwal and Rameswar Panda.

ETH Zurich — Zurich, Switzerland
Teaching Assistant — *Feb. 2024 – June 2024*
- Led weekly tutorials for *Algorithms & Probability*, taught by Angelika Steger and Rasmus Kyng.

Publications

- **Oliver Sieberling**, Bharat Runwal, David Jin, Ryan Chin, Rameswar Panda, Yoon Kim, *Triadic Linear Attention: Three-Dimensional Recurrent States for Long-Context Sequence Modeling*, *Preprint*, 2026.
- **Oliver Sieberling**, Bharat Runwal, Rameswar Panda, Yoon Kim, *Dynamic Short Convolutions Improve Transformers*, *NeurIPS*, 2026. Link.
- Maximilian Schlegel*, Rajai Nasser*, Seijin Kobayashi*, Yanick Schimpf*, **Oliver Sieberling***, Robert Obryk*, Kazuki Irie*, João Sacramento*, Johannes von Oswald*, *Retrieval-Centric Deep Learning in Growing Nonparametric Neural Networks*, *NeurIPS*, 2026.
- Zhaofeng Wu, **Oliver Sieberling**, Shawn Tan, Rameswar Panda, Yury Polyanskiy, Yoon Kim, *Variable-Width Transformers*, *Preprint*, 2026. Link.
- Johannes von Oswald*, Nino Scherrer*, Seijin Kobayashi, Luca Versari, Songlin Yang, Maximilian Schlegel, Kaitlin Maile, Yanick Schimpf, **Oliver Sieberling**, Alexander Meulemans, Guillaume Lajoie, Rif A. Saurous, Charlotte Frenkel, Razvan Pascanu, Blaise Agüera y Arcas, João Sacramento, *MesaNet: Sequence Modeling by Locally Optimal Test-Time Training*, *ICLR*, 2026. Link.
- Shengkun Tang, **Oliver Sieberling**, Eldar Kurtic, Zhiqiang Shen, Dan Alistarh, *DarwinLM: Evolutionary Structured Pruning of Large Language Models*, *COLM*, 2026. Link.
- Roberto Castro*, Andrei Panferov*, Soroush Tabesh, **Oliver Sieberling**, Jiale Chen, Mahdi Nikdan, Saleh Ashkboos, Dan Alistarh, *Quartet: Native FP4 Training Can Be Optimal for Large Language Models*, *NeurIPS*, 2025. Link.
- **Oliver Sieberling**, Denis Kuznedelev, Eldar Kurtic, Dan Alistarh, *EvoPress: Accurate Dynamic Model Compression via Evolutionary Search*, *ICML*, 2025. Link.
- Johannes Lengler$^{\alpha\beta}$, Leon Schiller$^{\alpha\beta}$, **Oliver Sieberling**$^{\alpha\beta}$, *Plus Strategies are Exponentially Slower for Planted Optima of Random Height*, *GECCO*, 2024. Link. Extended journal version published in *Algorithmica*, 2026.
- Marc Kaufmann$^{\alpha\beta}$, Maxime Larcher$^{\alpha\beta}$, Johannes Lengler$^{\alpha\beta}$, **Oliver Sieberling**$^{\alpha\beta}$, *Hardest Monotone Functions for Evolutionary Algorithms*, *EvoCOP*, 2024. Best Paper Award Nomination. Link. Extended journal version published in *SN Computer Science*, 2025.

*Equal contribution. $^{\alpha\beta}$Authors listed alphabetically (TCS).

Talks

- **Cohere Labs Community**: Dynamic Short Convolutions Improve Transformers. August 2026.
- **Pinterest**: Dynamic Short Convolutions Improve Transformers. August 2026.
- **EvoCOP 2024**: Hardest Monotone Functions for Evolutionary Algorithms. April 2024.

Academic Service

- **Reviewer**: NeurIPS 2026, ICML 2026, TMLR 2026.