

Andreas Plesner

aplesner@ethz.ch | aplesner.github.io | linkedin.com/in/andreas-plesner-540a54141/

Education

ETH Zurich

Expected September 2026

PhD Student

- Research areas: differentiable logic gate networks (energy-efficient, ultra-low-latency models), adversarial ML, and LLM post-training.
- Advisor: Prof. Dr. Roger Wattenhofer
- Supervised over 50 student projects
- Have accepted papers at NeurIPS, ACL, ICLR, TMLR, and ICS.
- Have reviewed papers for NeurIPS, ICML, ICLR, ICCV, TMLR, CVPR, and EMNLP.

Technical University of Denmark

September 2022

MSc. in Mathematical Modelling and Computing (GPA: 10.6 / 12.0)

- Selected for the Honours Programme based on outstanding academic success
- Relevant Coursework: Deep Learning, High-performance Computing, Large-scale Modelling, Graph Theory, Functional Analysis
- **Thesis title:** Using data-driven state-of-the-art machine learning and conformal prediction for track irregularities from observed dynamics of in-service railway vehicles

BSc. in Engineering in Mathematics and Technology (GPA: 11.6 / 12.0)

- Relevant Coursework: Mathematical software programming, Stochastic Processes, Algorithms and Data Structures II, Mathematical modelling
- **Thesis title:** Estimating golf swing parameters from radar data using deep neural networks

Experience

Handshake AI

San Francisco, California, USA

Research Contractor (remote)

May 2026 – September 2026

- Post-training agentic models on long rollout tasks.
- Improving the infrastructure of rollout engines to speed up post-training runs.

Research Intern

January 2026 – May 2026

- Post-training LLMs in domains with verifiable and semi-verifiable (rubrics) rewards.
- Determined the impact of imperfect verification on downstream performance after post-training.

Google

Zurich, Switzerland

Student Researcher

April 2024 – December 2024

- Extended the flicker detection work from the prior internship, focusing on unsupervised models and large-scale evaluation.
- Created efficient and highly accurate models that could be deployed on-device.
- Performed extensive error analysis and transfer studies to analyze the failure modes of the final model.

Software Engineering Intern

May 2023 – September 2023

- Researched anomaly detection for on-device flicker detection in Android.
- Designed an end-to-end data pipeline that processes the raw samples into formats suitable for ML models.
- Created a supervised ML model for flicker detection to determine the viability of the problem and data format.

Technical University of Denmark

Kongens Lyngby, Denmark

Teaching Assistant

September 2018 – May 2022

- TA for 7 courses including Graph Theory, Machine Learning and Data Mining, and Stochastic Simulation.

Oticon Medical

Smørum, Denmark

Student Assistant

October 2020 – February 2022

- Created GUIs in PyQt in Python with backend processing, and developed simulation software in MATLAB

Selected Publications

- *Light Differentiable Logic Gate Networks*. Presented at **ICLR** 2026. Optimized the parameterization of differentiable logic gate networks that speeds up training by a factor of 8.5 and reduces memory usage during training by a factor of 4.
- *Mind the Gap: Removing the Discretization Gap in Differentiable Logic Gate Networks*. Presented at **NeurIPS** 2025. Introduced Gumbel-softmax noise during training of differentiable logic networks. This helps models converge in 4.5x fewer training steps and helps the model converge to more accurate solutions.
- *ACORD: An Expert-Annotated Retrieval Dataset for Legal Contract Drafting*. Presented at **ACL** 2025. Introduced ACORD, the first expert-annotated dataset for benchmarking legal contract clause retrieval. Benchmarked state-of-the-art models and demonstrated that they still have significant limitations before they can be used in real-world applications.
- *Breaking reCAPTCHA v2*. Presented at COMPSAC 2024. Showed that a simple image classification model (YOLOv8) can be used to solve CAPTCHAs from Google's reCAPTCHA v2 better than humans.
- *Accurate Computation of the Logarithm of Modified Bessel Functions on GPUs*. Presented at **ICS** 2024. Presented new algorithms that compute the logarithm of modified Bessel functions with high accuracy on GPUs, enabling their use in applications such as machine learning and scientific computing.

Research interests

- **How do we create efficient and effective models?** My research has centered around 1) creating extremely efficient models, such as logic gate networks, and making them functional, and 2) helping models learn better solutions, such as learning better LoRA adaptors or determining the impact of verifiers in post-training with semi-verifiable rewards.
- **What are the limits of model generalization?** I investigate how training data, model architecture, and training method shape a model's capacity to perform accurately when faced with out-of-distribution (OOD) or adversarial scenarios.
- **Code and hardware optimization:** I aim to increase research capacity by writing code that uses the performance of modern hardware well. This goal is motivated by a personal passion for writing efficient code (making a program 100x faster is very satisfying!).

Skills

Languages: English (C2), Danish (native)

Programming Languages: Python and C/C++

Frameworks and Libraries: PyTorch, matplotlib, NumPy, SciPy, CUDA, OpenMP, OpenMPI

All Publications

Saku Peltonen, August Bøgh Rønberg, **Andreas Plesner**, and Roger Wattenhofer. GraphARC: A Comprehensive Benchmark for Graph-Based Abstract Reasoning. In *The 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD*, 2026.

Severin Bratus, David F. Jenny, **Andreas Plesner**, and Roger Wattenhofer. A Survey on the Abstraction and Reasoning Corpus. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL <https://openreview.net/forum?id=qzFxBcK9Cg>. **Survey Certification**.

Simon Jonas Bühner, **Andreas Plesner**, Till Aczel, and Roger Wattenhofer. BitLogic: A Framework for Gradient-Based LUT-Native Neural Networks. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL <https://openreview.net/forum?id=ZbsSZAfDod>.

Andreas Plesner, Francisco Guzmán, and Anish Athalye. An imperfect verifier is good enough: Learning with noisy rewards. *arXiv preprint arXiv:2604.07666*, 2026.

Lukas Rüttgers, Till Aczel, **Andreas Plesner**, and Roger Wattenhofer. Light differentiable logic gate networks. In *The Fourteenth International Conference on Learning Representations, ICLR*, 2026.

- Shakir Yousefi, **Andreas Plesner**, Till Aczel, and Roger Wattenhofer. Mind the Gap: Removing the Discretization Gap in Differentiable Logic Gate Networks. In *39th Conference on Neural Information Processing Systems (NeurIPS 2025)*, San Diego, California, USA, December 2025a.
- Samuel Räber, Till Aczel, **Andreas Plesner**, and Roger Wattenhofer. Keep It Real: Challenges in Attacking Compression-Based Adversarial Purification. In *Reliable ML from Unreliable Data Workshop (ReliableML@NeurIPS 2025)*, San Diego, California, USA, December 2025a.
- Till Aczel, Lorenzo Vettor, **Andreas Plesner**, and Roger Wattenhofer. The Unwinnable Arms Race of AI Image Detection. In *Workshop on Structured Probabilistic Inference & Generative Modeling (SPIGM@NeurIPS2025)*, San Diego, December 2025.
- Marco Zimmerli, **Andreas Plesner**, Till Aczel, and Roger Wattenhofer. The Impact of Scaling Training Data on Adversarial Robustness. In *Reliable ML from Unreliable Data Workshop (ReliableML@NeurIPS 2025)*, San Diego, California, USA, December 2025.
- Simon Bühner, **Andreas Plesner**, Till Aczel, and Roger Wattenhofer. Recurrent Deep Differentiable Logic Gate Networks. In *The 2nd International Workshop on Edge and Mobile Foundation Models (EdgeFM@MobiCom 2025)*, Hong Kong, November 2025.
- Steven H. Wang, Maksim Zubkov, Kexin Fan, Sarah Harrell, Yuyang Sun, Wei Chen, **Andreas Plesner**, and Roger Wattenhofer. ACORD: An Expert-Annotated Retrieval Dataset for Legal Contract Drafting. In *The 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, Vienna, Austria, July 2025.
- Shakir Yousefi, **Andreas Plesner**, Till Aczel, and Roger Wattenhofer. Mind the Gap: Removing the Discretization Gap in Differentiable Logic Gate Networks. In *Workshop on Methods and Opportunities at Small Scale (MOSS@ICML 2025)*, Vancouver, Canada, July 2025b.
- Samuel Räber, **Andreas Plesner**, Till Aczel, and Roger Wattenhofer. Human Aligned Compression for Robust Models. In *The 5th Workshop of Adversarial Machine Learning on Computer Vision: Foundation Models + X (AdvML@CVPR 2025)*, Nashville, Tennessee, USA, June 2025b.
- Andreas Plesner**, Turlan Kuzhagaliyev, and Roger Wattenhofer. FLIP Reasoning Challenge. In *First Workshop on Open Science for Foundation Models (SCI-FM@ICLR2025)*, Singapore, April 2025.
- Joël Mathys, **Andreas Plesner**, Jorel Elmiger, and Roger Wattenhofer. Synthetic Data for Blood Vessel Network Extraction. In *Will Synthetic Data Finally Solve the Data Access Problem? (SynthData@ICLR2025)*, Singapore, April 2025.
- Béni Egressy, **Andreas Plesner**, and Roger Wattenhofer. Lending in Financial Networks. In *25th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA)*, Kyoto, Japan, November 2024.
- Andreas Plesner**, Hans Henrik Brandenborg Sørensen, and Søren Hauberg. Accurate computation of the logarithm of modified bessel functions on GPUs. In *Proceedings of the 38th ACM International Conference on Supercomputing*, pages 213–224, 2024a.
- Andreas Plesner**, Tobias Vontobel, and Roger Wattenhofer. Breaking reCAPTCHA v2. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1047–1056. IEEE, 2024b.
- Andreas Plesner**, Allan P Engsig-Karup, and Hans True. Detecting railway track irregularities using conformal prediction. In *International Conference on Artificial Neural Networks*, pages 295–309. Springer Nature Switzerland Cham, 2024c.
- Anders Reenberg Andersen, Bo Friis Nielsen, and **Andreas Plesner**. An approximation of the inpatient distribution in hospitals with patient relocation using markov chains. *Healthcare Analytics*, 3:100145, 2023.
- Hans True, Lasse Engbo Christiansen, **Andreas Plesner**, Andreas Lønstrup Ammitzbøll, and Bjørn Jerram Dahl. On the problem of the dynamical reactions of a rolling wheelset to real track irregularities. *Railway Engineering Science*, 31(1):1–19, 2023.

Tobias Idor Boklund, Laurits Fromberg, Christian Henriksen, Benjamin Jüttner, Peder Jørgensgaard Olesen, **Andreas Plesner**, and Kenneth Scheel. Correction of fluorescent probe degradation in biodistribution studies. *Mathematics in Industry Reports*, 2022.

Anders Reenberg Andersen and **Andreas Plesner**. Optimization of the organizational structure in hospitals to account for patients with multiple diseases. *Artificial Intelligence in Medicine*, 130:102327, 2022.

True Hans, Lasse Engbo Christiansen, **Andreas Plesner**, Andreas Lønstrup Ammitzbøll, and Bjørn Jerram Dahl. Why is it so difficult to determine the lateral position of the rails by a measurement of the motion of an axle on a moving vehicle? In *Second International Conference on Rail Transportation*, 2021.