Hen Davidov

Oxford, UK · +44 7895 046068 · chend136@gmail.com
LinkedIn · Google Scholar

EDUCATION

University of Oxford — Oxford, UK
PhD in Statistics - Statistics and Machine Learning CDT, — 2025 – Present
- Advisors: Prof. Yarin Gal, Prof. Patrick Rebeschini, and Prof. Yee Whye Teh
- **Rhodes Scholar** (Class of 2025)

Technion - Israel Institute of Technology — Haifa, Israel
M.Sc. in Computer Science (Grade Avg: 95/100, Thesis Grade: 96/100) — 2023 – 2026
- Advisors: Prof. Yaniv Romano and Prof. Ron Kimmel

Technion - Israel Institute of Technology — Haifa, Israel
B.Sc. in Computer Science (Grade Avg: 91.5/100) — 2020 – 2023

HONORS & AWARDS

- **Rhodes Scholarship:** Prestigious international postgraduate award. [Press Coverage]
- **Forbes Israel 30 under 30 for 2025** [Press Coverage in Hebrew]
- **Jesus College Graduate Scholarship** (Total funding: £3,800)
- **TechAI Awards (2023, 2024):** Awarded by the Israeli Planning and Budgeting Committee for multi-disciplinary data science research (Total funding: $7,400).

PUBLICATIONS & WORKS UNDER REVIEW

- **H. Davidov**, S. Feldman, G. Shamai, R. Kimmel, Y. Romano. "Conformalized Survival Analysis for General Right-Censored Data." *ICLR* 2025.
- **H. Davidov***, S. Feldman*, G. Freidkin, Y. Romano. "Calibrated Predictive Lower Bounds on Time-to-Unsafe-Sampling in LLMs." *AISTATS* 2026 & *ICLR* 2026 workshop on Trustworthy AI.
- **H. Davidov**, N. Cohen, O. Kalinsky, Y. Fairstein, G. Kushilevitz, R. Yazdi, P. Rebeschini. "Knowing When to Quit: A Principled Framework for Dynamic Abstention in LLM Reasoning." *ICML* 2026 & *ICLR* 2026 workshop on Trustworthy AI
- Bat-Sheva Einbinder*, **Hen Davidov***, Yee Whye Teh, Yarin Gal, Yaniv Romano. "Selective Safety Steering via Value-Filtered Decoding." *ICML* 2026 workshop on AI4Good & Under review - *NeurIPS* 2026
- Yihong Chen, Zihuiwen Ye, Yi Ren, Luckeciano Carvalho Melo, Anushka Nair, Lin Li, Yug D Oswal, Yonatan Gideoni, **Hen Davidov**, Hao Fei, Xander Davies, Yarin Gal. "Knowledge Repositories and Reasoning Engines: AI Should Generalize Across Worlds." Under review - *NeurIPS* 2026
- G. Shamai, D. Aran, Y. Binenbaum, O. Cohen, S. Cohen, A. Piven, **H. Davidov**, E. Sabo, A. Cretu, A. Vituri, M. Shahar, H. Joensuu, P.-L. Kellokumpu-Lehtinen, D. Bychkov, N. Linder, J. Lundin, J. A. Sparano, and R. Kimmel. "Clinical Utility of Discordances Between Histomorphology and Molecular Biomarkers in Precision Oncology." Under review - *Journal of Clinical Oncology (ASCO)*

RESEARCH EXPERIENCE

University of Oxford — PhD Student
OATML lab & OxCSML group — 2025 – present
- **Dynamic Controlled Decoding (under review - NeurIPS 2026):** Developed method for inference-time intervention in LLM text generation for safety steering. Achieved best safety results while maintaining

Amazon Research & Development — Applied Scientist Intern
Personalization — 2025
- **Dynamic Abstention (ICML 2026):** Formulated dynamic mid-generation abstention via RL reward maximization, creating a training-lean algorithm compatible with any auto-regressive LLM, in PyTorch and HuggeingFace Transformers. Achieved higher safety accuracy with significantly lower token consumption than prompt-based and other baselines.

Technion - Israel Institute of Technology Research Assistant & M.Sc. student
Taub Faculty of Computer Science 2022 – 2026

- **Survival Analysis (ICLR 2025):** Developed conformalized uncertainty estimation for survival analysis data, advancing towards reliable AI-powered breast cancer prognosis. Implemented using PyTorch, SciPy, and scikit-learn.
- **Time-to-Unsafe-Sampling in LLMs (AISTATS 2026):** Utilized conformalized uncertainty estimation for efficient LLM safety quantification. Implemented in vLLM, Pytorch, and scikit-learn.
- **Digital Pathology (ASCO 2026):** Designed pipelines for million-pixel pathology images, improving training efficiency by 10x, using PyTorch, h5 and ray. Developed Transformer-based models for biomarker clustering using digital pathology foundation models.