Derck Prinzhorn

derckprinzhorn@gmail.com | linkedin.com/derckprinzhorn

Education

University of Amsterdam 2023 - 2026
Master of Science in Artificial Intelligence (Grade: 8.0/10) *Amsterdam*

- **Relevant Coursework:** Machine Learning, Deep Learning, Reinforcement Learning, Computer Vision, Natural Language Processing, Information Retrieval, Interpretability & Explainability.

University of Amsterdam 2020 - 2023
Bachelor of Science in Artificial Intelligence (Grade: 8.2/10) *Amsterdam*

- **Relevant Coursework:** Programming, Linear Algebra, Calculus, Bayesian Statistics, Machine Learning, Reinforcement Learning, Computer Vision, Natural Language Processing, Information Retrieval.

Het Amsterdams Lyceum 2014 - 2020
VWO Gymnasium *Amsterdam*

Research experience

Research Intern Jul 2024 – present
The Netherlands Cancer Institute *Amsterdam*

- Working on AI for radiotherapy, supervised by Stefanos Achlatis.

Research Intern Jul 2024 – Oct 2024
Supervised Program for Alignment Research (SPAR) *Remote*

- Worked on AI Control, focusing on safety techniques to detect and mitigate suspicious outputs using trusted and untrusted models.
- Worked with red and blue teaming strategies to identify and mitigate backdoors.
- Integrated the auditing step from the AI control paper into the codebase with a limited auditing budget.
- Created an ensemble monitor with custom members targeting specific backdoor likelihoods and implemented a distribution benchmark to track backdoors across ensemble setups.
- Gained experience in caching management, cost-effective prompting, and reproducing academic papers.
- Supervised by Aryan Bhatt, alignment researcher at Redwood Research.

Research Intern Mar 2024 – Jun 2024
Deltares *Utrecht*

- Researched conformal prediction methods for discharge forecasting, supervised by Jing Deng and Hans Korving.
- This involved implementing appropriate methods, evaluating their performance and explaining them to meteorologists.

Research Intern Jan 2024 – May 2024
University of Amsterdam *Amsterdam*

- Researched uncertainty quantification methods, supervised by Putri van der Linden and Alexander Timans. Specifically, we introduced a novel perspective on conformal prediction for time series.
- Paper accepted to COPA, a workshop with a focus on conformal prediction and published in PMLR.

Industry experience

Solution Architect AI Apr 2023 – present
Politie Nederland *Utrecht*

- Designing centralized MLOps architecture, consisting of MLOps processes, tooling and workflows, including CI/CD pipelines, and AI governance frameworks.
- Introducing AI Safety initiatives through risk modelling and safety engineering.

Machine Learning Engineer Sep 2022 – Jan 2024
Dutch Nao Team *Amsterdam*

- Developed AI models for pose classification, object detection, sound detection and reinforcement learning, supervised by Arnoud Visser.

- Managed team activities, project backlogs and led scrum teams, resulting in 5x more members and a novel robot framework built from scratch in Rust.

Software Engineer Oct 2021 – Jan 2023
LeerLevels *Amsterdam*

- Developed grading algorithms, search engines, and recommendation systems.
- Supervised an app development project, resulting in an MVP mobile app.

Academic work

NeurIPS Poster Oct 2024
Reproducibility Study of FairAC

- Presenting as a poster at the Neural Information Processing Systems (NeurIPS) 2024 conference.

Workshop Paper June 2024
Conformal time series decomposition with component-wise exchangeability

- Accepted to the 13th Symposium on Conformal and Probabilistic Prediction with Applications (COPA 2024) and published in the Proceedings of Machine Learning Research (PMLR 2024).

Journal Paper June 2024
Reproducibility study of FairAC

- Published in the Transactions on Machine Learning Research (TMLR 2024) and accepted to the Machine Learning Reproduction Challenge (MLRC2023).

Bachelor Thesis June 2023
Benchmarking conformal prediction methods for time series regression

Honors and awards

AmsterdamAI Thesis Award Winner

- Awarded for outstanding bachelor thesis on conformal prediction for time series.

Teaching

Information Visualization Spring 2023
Teaching assistant for BSc course at UvA

Cognitive Modeling (Reinforcement Learning) Spring 2023
Teaching assistant for BSc course at UvA

Datastructures and Algorithms Winter 2022
Teaching assistant for BSc course at UvA

Machine Learning Project Winter 2022
Teaching assistant for BSc course at UvA

Introduction to Machine Learning Fall 2022
Teaching assistant for BSc course at UvA

Bayesian Statistics for Machine Learning Fall 2022
Teaching assistant for BSc course at UvA

Projects

AI Safety Hackathon, 2nd place | *LLMs*, *SAEs*, *TransformerLens* November 2023

- Developed a novel method to inspect, reverse engineering and steer Large Language Models.
- Our team achieved second place out of 8 teams.

Robotics Hackathon ERF2022, 2nd place | *Python*, *ROS2*, *Robotics* June 2022

- Created software for Lely Juno robot, achieving second place among robotics master students.

Machine Learning Project | *Python*, *Random Forests* January 2022

- Conducted ML project to identify drivers of real estate valuation growth for KR&A.

Volunteering and organizing

AI Safety Amsterdam (AISA) Sep 2023 – present
Core team *Amsterdam*

Google Developer Student Clubs UvA Dec 2023 – Jun 2024
Core team *Amsterdam*

Foundation Dutch Nao Team Jul 2023 – Mar 2024
Vice chair *Amsterdam*

- Refined board processes, managed recruitment, and developed partnerships

Programme Committee AI UvA Sep 2021 – Apr 2023
Member *Amsterdam*

- Contributed to AI program discussions, course evaluations, and resolving student-teacher issues

Stichting Hoormij Jun 2021 – May 2023
Board member *Houten*

- Advisor to the board of Hoormij.NVVS
- Focused on tinnitus and innovation strategies within the organization.

Tinnitus Jong Netwerk, Stichting Hoormij Jan 2021 – Apr 2022
Secretary *Houten*

- Established a committee for young people with tinnitus.

Stichting Studiezalen Feb 2020 – Oct 2021
Mentor *Amsterdam*

- Mentored high school students in coaching and homework tutoring.

School's cool Oct 2020 – Aug 2021
Mentor *Amsterdam*

- Mentored primary school students during their transition to high school, while managing language and arithmetic backlogs and home situation.

Skills

Languages: Dutch (Native), English (Professional)
Programming Languages: Advanced - Python; Basic - Rust, C++, HTML, CSS, JavaScript
Data Science and Machine Learning: Scientific Libraries - Numpy, Pandas, Scipy, Matplotlib, Astropy; ML Frameworks - Scikit-learn, PyTorch, TensorFlow, OpenCV, Jax, Statsforecast
Databases: SQL - PostgreSQL, MySQL, SQLite; NoSQL - JSON, Firebase (Cloud Firestore); Graph - Neo4j
Development and API Tools: API Development - Flask, Fastapi, Postman; Development Tools - Jupyter, GitHub, Git, Bash shell, Docker, Kubernetes
MLOps: Experiment Tracking - MLflow, Weights & Biases, Neptune; Orchestration - Metaflow, Kubeflow, Airflow