

# Katharina Deckenbach

RESEARCH ENGINEER · MACHINE LEARNING

## About me

Highly motivated and curious Machine Learning Master's graduate with a hard-working attitude. Excited about the potential of AI and driven by the goal of mitigating its risks to ensure future AI systems can be as beneficial as possible.

## Contact

✉ kathi.deckenbach@gmx.de  
☎ +49 151 57272060  
📖 72070 Tübingen  
🔗 kaethy  
🌐 katharina-deckenbach  
🏠 nBGXSEAAAAJ

## Languages

German	Native
English	C1
Spanish	C1
Swedish	B1
French	A1

## Programming

Python	
R	
Matlab	
Java	
Julia	

## Other Skills

Uncertainty Estimation	Git
PyTorch	Weights&Biases
Hugging Face	Presentation
Communication	

## EDUCATION

08/23 - 06/25 **M.Sc. Machine Learning**  
KTH ROYAL INSTITUTE OF TECHNOLOGY STOCKHOLM, SWEDEN  
Thesis: "Improving the Laplace Posterior Approximation via Sharpness-Aware Minimization".

09/21 - 03/22 **Erasmus Semester**  
UNIVERSIDAD POMPEU FABRA BARCELONA, SPAIN

10/19 - 03/23 **B.Sc. Cognitive Science**  
UNIVERSITY OF TÜBINGEN TÜBINGEN, GERMANY  
Thesis: "Are People Confident When Artificial Neural Networks Are Confident? ".

## EXPERIENCE

11/25 - present **Research Intern**  
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS TÜBINGEN, GERMANY  
• Research Internship with Sahar Abdelnabi's COMPASS group

07/25 - 10/25 **Research Engineer**  
KTH ROYAL INSTITUTE OF TECHNOLOGY STOCKHOLM, SWEDEN  
• Continued my thesis work on improving posterior approximation methods  
• Trained and evaluated models (e.g., ResNet, ViT, BERT) in PyTorch, using Laplace Approximation and SWAG

2025 **Participant**  
ML4GOOD AND ARBOx STRASBOURG, FRANCE AND OXFORD, UK  
• Two 10-day in-person bootcamps to up-skill in AI safety  
• Topics covered training ResNet and GPT-2 from scratch, RLHF, reasoning models, mechanistic interpretability, adversarial attacks, frontier model evaluations, reinforcement learning, steering vectors

04/22 - 07/22 **Teaching Assistant**  
UNIVERSITY OF TÜBINGEN TÜBINGEN, GERMANY  
• The course taught by Prof. Martin Butz covered various topics within ML and Cognitive Sciences such as Reinforcement Learning and Evolutionary Algorithms  
• Held a weekly tutorial session, created learning material for students and helped with exam correction

## SCHOLARSHIP

12/20 - 03/22 **Deutschlandstipendium**  
Public-private German Scholarship

## ENGAGEMENT

08/23 - 06/25 **Co-Organizer Stockholm AI Safety**  
• Organized biweekly meetings and a discussion group

08/23 - 06/25 **Treasurer Effective Altruism KTH**

09/18 - 08/19 **Voluntary Social Year**

## PUBLICATIONS

2025 E. Banzuzi, D. J. Khodaverdian, and K. Deckenbach, "A Reproducibility Study of Decoupling Feature Extraction and Classification Layers for Calibrated Neural Networks", *Transactions on Machine Learning Research June-2025*, 2025.