

## EDUCATION

---

<b>University of Amsterdam</b> - <i>PhD Student</i>	May 2025 – Present
<b>Technical University of Munich</b> - <i>MSc in Informatics</i>	Oct 2022 – Mar 2025
<b>Koç University</b> - <i>BSc in Computer Engineering &amp; Mathematics</i>	Sep 2017 – Jun 2022

## RESEARCH EXPERIENCE

---

<b>Doctoral Researcher</b> - <i>University of Amsterdam</i> 👤 Prof. Ana Lucic 🔗 Mechanistic interpretability, AI4Science	May 2025 – Present
<b>MSc Thesis Student</b> - <i>University of Munich, Technical University of Munich</i> 👤 Prof. David Rügamer (LMU), Prof. Bastian Rieck (TUM) 🔗 Sampling neural network weights using geometric flow matching and GNNs	Jun 2024 – Jan 2025
<b>TUM DI-LAB</b> - <i>Technical University of Munich</i> 👤 Hannes Stärk, MSc. (MIT), Céline Marquet, MSc. (TUM) 🔗 Estimating free-energy differences between molecules using flow matching	Oct 2023 – Sep 2024
<b>Student Research Assistant</b> - <i>Technical University of Munich</i> 👤 Mahdi Dhaini, MSc. (Prof. Gjergji Kasneci) 🔗 Explainability, fairness, privacy in NLP	Jun 2023 – Mar 2025
<b>Guided Research</b> - <i>Technical University of Munich</i> 👤 Simon Geisler, MSc. (Prof. Stephan Günnemann) 🔗 Adversarial robustness of GNNs	Apr 2023 – Oct 2023
<b>Student Research Assistant</b> - <i>Helmholtz Munich, Institute of Computational Biology</i> 👤 Ignacio Ibarra, PhD. (Prof. Fabian Theis) 🔗 Inference of protein binding specificities	Nov 2022 – Mar 2023
<b>Research Intern</b> - <i>Koç University Cryptography, Security, and Privacy Research Group</i> 👤 Assoc. Prof. Alptekin Küpçü, Asst. Prof. Ercüment Çiçek 🔗 Privacy and confidentiality in federated machine learning	Jul 2020 – Oct 2022

## SELECTED PUBLICATIONS

- 
- Ege Erdogan**, Ana Lucic. "Group Equivariance Meets Mechanistic Interpretability: Equivariant Sparse Autoencoders", 2025; *NeurIPS Mechanistic Interpretability and Unifying Representations in Neural Models (UniReps) Workshops*.
  - Mahdi Dhaini, **Ege Erdogan**, Nils Feldhus, Gjergji Kasneci. "Gender Bias in Explainability: Investigating Performance Disparity in Post-hoc Methods", 2025; [doi.org/10.1145/3715275.3732192](https://doi.org/10.1145/3715275.3732192) ACM FAccT 2025.
  - Ege Erdogan**, Radoslav Ralev, Mika Rebensburg, Celine Marquet, Leon Klein, Hannes Stark "FreeFlow: Latent Flow Matching for Free Energy Difference Estimation", 2025; *Workshop on Machine Learning Multiscale Processes (at ICLR '25)*.
  - Ege Erdogan**, Unat Tekşen, Mehmet Salih Çeliktenyıldız, Alptekin Küpçü, A. Ercüment Çiçek. "SplitOut: Out-of-the-Box Training-Hijacking Detection in Split Learning via Outlier Detection", 2023; [arxiv.org/abs/2302.08618](https://arxiv.org/abs/2302.08618). *International Conference on Cryptology And Network Security (CANS '24)*.
  - Ege Erdogan**, Simon Geisler, Stephan Günnemann. "Poisoning × Evasion: Symbiotic Adversarial Robustness for Graph Neural Networks", 2023; [arxiv/2312.05502](https://arxiv.org/abs/2312.05502). *New Frontiers in Graph Learning Workshop (NeurIPS GLFrontiers 2023)*.
  - Ege Erdogan**, Alptekin Küpçü, A. Ercüment Çiçek. "SplitGuard: Detecting and Mitigating Training-Hijacking Attacks in Split Learning", 2022; [doi.org/10.1145/3559613.3563198](https://doi.org/10.1145/3559613.3563198). *The 21st Workshop on Privacy in the Electronic Society (at ACM CCS '22)*.
  - Ege Erdogan**, Alptekin Küpçü, A. Ercüment Çiçek. "UnSplit: Data-Oblivious Model Inversion, Model Stealing, and Label Inference Attacks Against Split Learning", 2022; [doi.org/10.1145/3559613.3563201](https://doi.org/10.1145/3559613.3563201). *The 21st Workshop on Privacy in the Electronic Society (at ACM CCS '22)*.

## TEACHING

---

- Undergraduate Tutor (x2)** - Koç University Feb. 2021 – June 2021 and Sep. 2021 - Jan. 2022  
Weekly tutoring sessions and help with course administration for the *Computer Networks* course.
- Undergraduate Teaching Assistant** - Koç University Sep. 2020 – Jan. 2021  
Weekly problem sessions and office hours for the *Introduction to Programming with Python*.

## NON-RESEARCH WORK EXPERIENCE

---

- Software Development Intern** - Proteams July 2020 – Aug. 2020  
Implemented features for different parts of a mobile application, including its back-end server, database, and admin panel.
- Summer Intern** - IBM July 2019  
Worked within the IBM Cloud & Cognitive team. Built a chatbot to answer questions by IBM new hires. Developed a web interface for a neural network music generation system.
- Software Development Intern** - Bitlo Cryptocurrency Exchange Feb. 2018 – April 2018  
Started learning and gained experience using the Spring framework to build web applications. Gained practice with the Java language.

## SKILLS & INTERESTS

---

- Languages:** Turkish (native), English (fluent), German (beginner)
- Computer Languages**
- Primary:** Python (PyTorch for ML)
  - Others:** Java/JavaScript (web development), Go (distributed systems)
- Research interests:** deep generative models, geometric DL, safe/trustworthy DL, DL applications in science.

## FULL PUBLICATIONS

---

- Ege Erdogan**, Ana Lucic. "Group Equivariance Meets Mechanistic Interpretability: Equivariant Sparse Autoencoders", 2025; *NeurIPS Mechanistic Interpretability and Unifying Representations in Neural Models (UniReps) Workshops*.
- Mahdi Dhaini, Juraj Vladika, **Ege Erdogan**, Zineb Attaoui, Gjergji Kasneci. "Can LLM-Generated Textual Explanations Enhance Model Classification Performance? An Empirical Study", 2025; [doi.org/10.1007/978-3-032-04549-2\\_16](https://doi.org/10.1007/978-3-032-04549-2_16) ICANN 2025.
- Mahdi Dhaini, Stephen Meisenbacher, **Ege Erdogan**, Florian Matthes, Gjergji Kasneci. "When Explainability Meets Privacy: An Investigation at the Intersection of Post-hoc Explainability and Differential Privacy in the Context of Natural Language Processing", 2025; [arxiv2508.10482](https://arxiv.org/abs/2508.10482) Preprint.
- Mahdi Dhaini, **Ege Erdogan**, Nils Feldhus, Gjergji Kasneci. "Gender Bias in Explainability: Investigating Performance Disparity in Post-hoc Methods", 2025; [doi.org/10.1145/3715275.3732192](https://doi.org/10.1145/3715275.3732192) ACM FAccT 2025.
- Ege Erdogan**, Radoslav Ralev, Mika Rebersburg, Celine Marquet, Leon Klein, Hannes Stark "FreeFlow: Latent Flow Matching for Free Energy Difference Estimation", 2025; *Workshop on Machine Learning Multiscale Processes (at ICLR '25)*.
- Mahdi Dhaini, **Ege Erdogan**, Smarth Bakshi, Gjergji Kasneci. "Explainability Meets Text Summarization: A Survey", 2024; [aclanthology/2024.inlg-main.49/](https://aclanthology.org/2024.inlg-main.49/) *The 17th International Natural Language Generation Conference (INLG '24)*.
- Ege Erdogan**, Unat Tekşen, Mehmet Salih Çeliktenyıldız, Alptekin Küpçü, A. Ercüment Çiçek. "SplitOut: Out-of-the-Box Training-Hijacking Detection in Split Learning via Outlier Detection", 2023; [arxiv.org/abs/2302.08618](https://arxiv.org/abs/2302.08618). *International Conference on Cryptology And Network Security (CANS '24)*.
- Ege Erdogan**, Simon Geisler, Stephan Günnemann. "Poisoning × Evasion: Symbiotic Adversarial Robustness for Graph Neural Networks", 2023; [arxiv/2312.05502](https://arxiv.org/abs/2312.05502). *New Frontiers in Graph Learning Workshop (NeurIPS GLFrontiers 2023)*.
- Mahdi Dhaini, Wessel Poelman, **Ege Erdogan**. "Detecting ChatGPT: A Survey of the State of Detecting ChatGPT-Generated Text", 2023; [aclanthology/2023.ranlp-stud.1](https://aclanthology.org/2023.ranlp-stud.1) *Student Research Workshop (at RANLP '23)*.
- Ege Erdogan**, Alptekin Küpçü, A. Ercüment Çiçek. "SplitGuard: Detecting and Mitigating Training-Hijacking Attacks in Split Learning", 2022; [doi.org/10.1145/3559613.3563198](https://doi.org/10.1145/3559613.3563198). *The 21st Workshop on Privacy in the Electronic Society (at ACM CCS '22)*.

11. **Ege Erdogan**, Alptekin K p , A. Erc ment  i ek. “UnSplit: Data-Oblivious Model Inversion, Model Stealing, and Label Inference Attacks Against Split Learning”, 2022; [doi.org/10.1145/3559613.3563201](https://doi.org/10.1145/3559613.3563201). *The 21st Workshop on Privacy in the Electronic Society (at ACM CCS '22)*.
12. **Ege Erdogan**, Can Arda Aydın,  znur  zkasap, Waris Gill. “Demo – Zelig: Customizable Blockchain Simulator”, 2021; [arXiv:2107.07972](https://arxiv.org/abs/2107.07972). *The 40th International Symposium on Reliable Distributed Systems (SRDS '21)*.