

Mateusz Cedro

mateuszcedro.com | mateusz.cedro@uantwerpen.be | [Google Scholar](https://scholar.google.com/citations?user=...) | [GitHub](https://github.com/mateuszcedro) | [LinkedIn](https://www.linkedin.com/in/mateuszcedro)

I am a second-year PhD student in Machine Learning, in particular **machine learning interpretability** and **explainable AI**. My current research focuses on **AI Safety and Alignment** of Large Language Models and semantic concepts in neural networks. My previous research concerned explanations of various ML models: LLMs [1], Graph Neural Networks [2], Computer Vision [3], and AI models in medicine [4,5]. I aim to **increase understanding** and **alignment of AI**.

Education

- 2024 – now **PhD student in Machine Learning**, University of Antwerp, Belgium, advisor: Prof. Dr. David Martens [↗](#)
- 2021 – 2023 **MSc, Data Science**, University of Warsaw, Poland
- 2018 – 2021 **BSc, Economics**, University of Warsaw, Poland

Publications and Preprints

- [1] **M. Cedro**, T. Ichmoukhamedov, S. Goethals, Y. He, J. Hinns, D. Martens. Cash or Comfort? How LLMs Value Your Inconvenience. Accepted to *Communications of the ACM*, 2026. [↗](#)
- [2] **M. Cedro**, D. Martens. GraphXAIN: Narratives to Explain Graph Neural Networks. *World Conference on Explainable Artificial Intelligence*, 2025. [↗](#)
- [3] **M. Cedro**, M. Chlebus. Beyond the Black Box: Do More Complex Deep Learning Models Provide Superior XAI Explanations? Under review, 2026. [↗](#)
- [4] M. Dobrzycka, A. Sulewska, J. Konopinska, P. Karabowicz, A. Charkiewicz, K. Golaszewska, **M. Cedro**, P. Biecek, J. Niklinski, A. Charkiewicz. learning-based identification of small RNA signatures in aqueous humor as a step toward precision diagnosis of glaucoma. *Annals of Medicine, Elevate*, 2025. [↗](#)
- [5] M. Dobrzycka, A. Sulewska, **M. Cedro**, P. Karabowicz, R. Charkiewicz, A. Charkiewicz, K. Golaszewska, P. Biecek, J. Niklinski, A. Charkiewicz, J. Konopinska. *Small RNA profiling in ocular liquids, enhanced with clinical and imaging data, to develop a multimodal machine learning model for glaucoma detection*. Under review, 2026.

Conference Presentations

- Antwerp Center on Responsible AI, 2026**. Cash or Comfort? How LLMs Value Your Inconvenience. (talk)
- Flanders AI Research Program, Research Day, 2025**. Risks of LLMs' Valuation of Human Discomfort. (poster)
- World Conference on Explainable Artificial Intelligence, 2025**. GraphXAIN: Narratives to Explain Graph Neural Networks. (talk)
- Antwerp Center on Responsible AI, 2025**. Narratives to Explain AI Models. (talk)

Work Experience

- Jan - Sept 2024 **AI Research Scientist**, MI2.AI, Poland
 - Developing and explaining multimodal machine learning models (tabular-image data) for medical diagnosis of glaucoma disease
 - Conducting ML explainability research under a supervision of Prof. Dr. Przemyslaw Biecek

- Jun - Nov 2023 **Machine Learning Engineer**, Synerise, Poland
- Developing BaseModel.AI, a foundation deep learning model for behavioural data modelling
 - Developing and tailoring deep learning models for clients' databases using PyTorch for classification, recommendation, and regression on datasets with millions of users
 - Cooperating within AI Research and ML Engineering teams
- Jul - Dec 2022 **Machine Learning Engineer**, Data Juice Lab, Poland
- Developing deep learning computer vision models for medical disease classification and segmentation in PyTorch
- Feb - Jul 2022 **Data Scientist**, Health Labs Care, Poland
- Machine learning models development and data analytics
 - Process optimization

Teaching

- 2026 **Data Engineering** (master course) - lab sessions & project supervision, 50 students/year
- 2024, 2025 **Python for Machine Learning** (master course) - lab sessions & project supervision, 140 students/year

Summer Schools

- Jun 2026 **Machine Learning Summer School on Safety & Reliability**, organized by Prof. Dr. Wojciech Samek, Prof. Dr. Bartosz Zieinski, and Prof. Dr. Jacek Tabor at the Jagiellonian University in Krakow, Poland. The summer school focused on problems and research advances in the field of AI Safety and Alignment. [↗](#)
- Jun 2024 **Summer School on Artificial Intelligence for Biomedical Applications**, organized by Prof. Dr. Franco Scarselli at the University of Siena, Italy. The summer school focused on Graph Neural Network modelling of medical data and Explainable AI for medicine. [↗](#)

References

Prof. Dr. David Martens [↗](#)

Full Professor of Data Science, University of Antwerp, Belgium
Chair of Antwerp Center on Responsible AI, Belgium
david.martens@uantwerpen.be

Prof. Dr. Przemyslaw Biecek [↗](#)

Full Professor of Computer Science, University of Warsaw, Poland
Director of the Centre for Credible AI, Warsaw University of Technology, Poland
p.biecek@uw.edu.pl