eXplainable Artificial Intelligence (XAI) Workshop

November 12th, 2024

Karlsruhe Institute of Technology
Blücherstraße 17
76185 Karlsruhe

Organizers

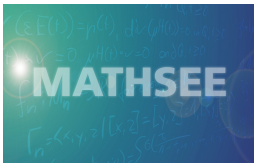
Prof. Dr. Steffen Rebennack

Lelisa Kebena Bijiga

Karlsruher Institut für Technologie (KIT), Institut für Operations Research
Stochastische Optimierung, Blücherstraße 17
Geb. 09.21, 76185 Karlsruhe

We would like to welcome you to the **eXplainable AI (XAI)** workshop at Blücherstraße campus. This meeting takes place on November 12$^{\text{th}}$, 2024 from 1:30pm to 6:00pm and is devoted to current research and advances in **XAI**.

We welcome researchers working on these topics from different career stages to discuss both the theoretical foundations and practical applications of **XAI**. This workshop takes place both onsite and online so that participants can access easily. We hope this workshop will provide valuable insights on making AI models more explainable and interpretable.



This workshop is co-hosted and sponsored by the KIT center "Mathematics in Sciences, Engineering, and Economics"(MathSEE).
Website: `https://www.mathsee.kit.edu`

Yours sincerely,

Steffen Rebennack
(Karlsruhe Institut of Technologie)

Workshop agenda

Tuesday, November 12, 2024, 13:30 - 18:00

Time	
	Session 1
	Session chair: **Prof. Dr. Steffen Rebennack**
	Online chair: **Dr. John Warwicker**
13:30 - 13:35	**Opening and Welcome**
13:35 - 14:30	**Dr. Vikram Sunkara** *Zuse Institute Berlin, Germany* A "Deep" Dive into how Neural Networks see Data
14:30 - 15:25	**Prof. Dr. Gitta Kutyniok** *Ludwig-Maximilians-Universität München, Germany* Mathematical Algorithm Design for Deep Learning under Societal and Judicial Constraints: The Algorithmic Transparency Requirement
15:25 - 15:55	——————————— Coffee break ———————————
	Session 2
	Session chair: **Lelisa Bijiga**
	Online chair: **Dr. Christian Füllner**
15:55 - 16:50	**Maximilian Fleissner** *Technical University of Munich, Germany* Explaining (Kernel) Clustering via Decision Trees
16:50 - 17:45	**Dr. Alessandro Renda** *University of Pisa, Italy* Increasing trust in AI through Federated Learning and model explainability
17:45 - 18:00	**Prof. Dr. Melanie Schienle** *Karlsruhe Institute of Technology, Germany* MathSEE Overview: Vision, Governance, Organization
18:00	——————— End of workshop ———————

The XAI Speakers

G. Kutyniok

Gitta Kutyniok received her Doctorate in Mathematics in 2000 from Paderborn University. She has held various academic positions across top universities including Princeton, Stanford,Yale and Georgia Institute of Technology. She received her habilitation in Mathematics at the University of Giessen in 2006. From 2008 - 2011, she has been a Full Professor for Applied Analysis at Osnabrück University, and Head of the Applied Analysis Group (AAG) and later she was given Einstein-Chair in Mathematics at the Technical University of Berlin from 2011 - 2020. From 2019 - 2023 she has been Adjunct Professor in Machine Learning at the University of Tromsø. Since 2020, she holds a Bavarian AI Chair for Mathematical Foundations of Artificial Intelligence at the Ludwig-Maximilians-University Munich. She has received various honors and awards including the von Kaven Prize by the DFG in 2007. She was invited as the Noether Lecturer at the ÖMG-DMV Congress in 2013, a plenary lecturer at the 8th European Congress of Mathematics (8ECM) in 2021. She was also honored by invited lectures at both the International Congress of Mathematicians 2022 (ICM 2022) and the International Congress on Industrial and Applied Mathematics (ICIAM) in 2023.
She became a SIAM Fellow in 2019, joined the Berlin-Brandenburg Academy of Sciences and Humanities in 2016, and was elected to the European Academy of Sciences in 2022. Gitta Kutyniok's research focuses on applied mathematics, artificial intelligence, and deep learning.

Maximilian Fleissner obtained his Master's degree in Mathematics at the Technical University of Munich (TUM), primarily focusing on statistics and machine learning. Maximilian is pursuing a PhD in the group of Prof. Debarghya Ghoshdastidar at TUM. His research interests include statistical learning theory, kernel methods, and explainable machine learning.

M. Fleissner

V. Sunkara

After studying Mathematics at University of Wollongong, Australia **Vikram Sunkara** received his PhD from the Australian National University for his work on Analysis and Numerics of the Chemical Master Equation in 2013. From 2012 - 2014 he has been a research associate in Numerical Mathematics at the Karlsruhe Institute of Technology (KIT), from 2014 - 2015 at the Department of Mathematics and Statistics at University of Adelaide (UOA)and from 2015 - 2021 at Mathematics of Complex Systems, Zuse Institute Berlin (ZIB) and Biocomputing Group (FU Berlin). Vikram is Head of Explainable A.I. for Biology at ZIB since 2021. His research focuses on modern mathematical biology, applied Mathematics, deep learning, and explainable AI methods for biology in areas like inflammation.

A. Renda

Alessandro Renda received the M.Sc. degree in Biomedical Engineering from the University of Pisa in 2017 and the Ph.D. degree in Smart Computing jointly awarded by the Universities of Florence, Pisa and Siena, in 2021, with a dissertation titled Älgorithms and Techniques for Data Stream Mining". He is currently assistant professor at the University of Pisa, working at the Department of Information Engineering (DII) as a member of the Artificial Intelligence-DII (AI-DII) LAB. His research interests include explainable artificial intelligence and federated learning, machine learning algorithms for data streams, applications of deep learning methodologies, and web/ social mining.

Mathematical Algorithm Design for Deep Learning under Societal and Judicial Constraints: The Algorithmic Transparency Requirement

Gitta Kutyniok, Ludwig-Maximilians-Universität München, Germany

Deep learning still has drawbacks in terms of trustworthiness, which describes a comprehensible, fair, safe, and reliable method. To mitigate the potential risk of AI, clear obligations associated to trustworthiness have been proposed via regulatory guidelines, e.g., in the European AI Act. Therefore, a central question is to what extent trustworthy deep learning can be realized. Establishing the described properties constituting trustworthiness requires that the factors influencing an algorithmic computation can be retraced, i.e., the algorithmic implementation is transparent. Motivated by the observation that the current evolution of deep learning models necessitates a change in computing technology, we derive a mathematical framework which enables us to analyze whether a transparent implementation in a computing model is feasible. We exemplarily apply our trustworthiness framework to analyze deep learning approaches for inverse problems in digital and analog computing models represented by Turing and Blum-Shub-Smale Machines, respectively. Based on previous results, we find that Blum-Shub-Smale Machines have the potential to establish trustworthy solvers for inverse problems under fairly general conditions, whereas Turing machines cannot guarantee trustworthiness to the same degree.

Explaining (Kernel) Clustering via Decision Trees

Maximilian Fleissner, Technical University of Munich, Germany

Despite the growing popularity of explainable and interpretable machine learning, there is still surprisingly limited work on inherently interpretable clustering methods. Recently, there has been a surge of interest in explaining the classic k-means algorithm using axis-aligned decision trees. However, interpretable variants of k-means have limited applicability in practice, where more flexible clustering methods are often needed to obtain useful partitions of the data. We investigate interpretable kernel clustering, and propose algorithms that construct decision trees to approximate the partitions induced by kernel k-means, a nonlinear extension of k-means. Our method attains worst-case bounds on the clustering cost induced by the tree. In addition, we introduce the notion of an explainability-to-noise ratio for mixture models. Assuming sub-Gaussianity of the mixture components, we derive upper and lower bounds on the error rate of a suitably constructed decision tree, capturing the intuition that well-clustered data can indeed be explained well with a decision tree.

A "Deep" Dive into how Neural Networks see Data

Vikram Sunkara, Explainable A.I. for Biology, Zuse Institute Berlin, Germany

Artificial Neural Networks (ANNs) have become a ubiquitous tool in society. They permeate through nearly all facets of our modern lives from leisure recommendations to more critical medical diagnosis. In this talk, we'll take a 'deep' dive into how neural networks perceive/process data and what mathematical patterns emerge inside ANNs. We will study a simple ANN, and describe its design through the lens of functional analysis; understand the learning mechanism with the lens of stochastics; decode its inner representation using the lens of geometry; and more importantly, with our new understanding and intuition, we will look at some real world applications in medicine and look into the architectures on what they have learnt.

Increasing trust in AI through Federated Learning and model explainability

Alessandro Renda, Department of Information Engineering, University of Pisa, Italy

Federated Learning (FL) lets multiple data owners collaborate in training a global model without any violation of data privacy, which is a crucial requirement for enhancing users' trust in Artificial Intelligence (AI) systems. Despite the significant momentum recently gained by the FL paradigm, most of the existing approaches in the field neglect another key pillar for the trustworthiness of AI systems, namely explainability. In this contribution, we discuss the concept of FL of eXplainable AI models (XAI), purposely designed to address the requirements of privacy and explainability simultaneously. We show how this goal can be achieved by tailoring either the learning procedure of inherently interpretable models or the application of post-hoc explanation methods to the federated setting.