[PDF] [Mechanistic Interpretability for AI Safety--A Review](#)

L Bereska, E Gavves - arXiv preprint arXiv:2404.14082, 2024

Understanding AI systems' inner workings is critical for ensuring value alignment and safety. This review explores mechanistic interpretability: reverse-engineering the computational mechanisms and representations learned by neural networks into …

[Combinational regularity analysis (CORA): An introduction for psychologists.](#)

A Thiem, L Mkrtchyan, Z Sebechlebská - Psychological Methods, 2024

Increasingly, psychologists make use of modern configurational comparative methods (CCMs), such as qualitative comparative analysis (QCA) and coincidence analysis (CNA), to infer regularity-theoretic causal structures from psychological data …

[PDF] [A Guide to Feature Importance Methods for Scientific Inference](#)

FK Ewald, L Bothmann, MN Wright, B Bischl… - arXiv preprint arXiv …, 2024

While machine learning (ML) models are increasingly used due to their high predictive power, their use in understanding the data-generating process (DGP) is limited. Understanding the DGP requires insights into feature-target associations …

[PDF] [Quantum non-classicality in the simplest causal network](#)

P Lauand, D Poderini, R Rabelo, R Chaves - arXiv preprint arXiv:2404.12790, 2024

Bell's theorem prompts us with a fundamental inquiry: what is the simplest scenario leading to the incompatibility between quantum correlations and the classical theory of causality? Here we demonstrate that quantum non-classicality is possible in a …

[PDF] [Audit of God: Hometown Connections and Building Damage in the Sichuan Earthquake](#)

Y Cao - 2024

This study documents how corruption can result in large-scale welfare consequences by exacerbating the damage from catastrophic events. Using an original dataset of 1,050 buildings from the 2008 Sichuan earthquake in China, I …

[PDF] [On the Use of Network Control Techniques in Pursuit of Influence Spread in Complex Networks](#)

A Sadaf - 2024

A critical element in the process of control and influence spread is the selection of the seed nodes from which influence and control spreads. To be able to meet project aim and to develop more effective seed selection methods, first we need to understand …

[PDF] [Specific Solutions to General Problems in Data Science and Ecology](#)

E Saberski - 2024

Nature is hard to predict. Rules and relationships you discover about a system today may be totally different tomorrow. These relationships do not change randomly over time; rather, they change as the state of the system evolves. In a deterministic view of …

[PDF] [Statistical Methods for Offline Deep Reinforcement Learning](#)

D Wang - 2024

Reinforcement learning (RL) has been a rapidly evolving field of research over the past years, enhancing developments in areas such as artificial intelligence, healthcare, and education, to name a few. Regardless of the success of RL, its …

[PDF] [Robust Representation Learning for Out-of-Distribution Extrapolation in Relational Data](#)

Y Zhou - 2024

Recent advancements in representation learning have significantly enhanced the analysis of relational data across various domains, including social networks, bioinformatics, and recommendation systems. In general, these methods assume …

[PDF] [Causal Neuro-Symbolic AI: A synergy between Causality and Neuro-Symbolic methods](#)

U Jaimini, C Henson, A Sheth

Causal Neuro-Symbolic AI combines the benefits of causality with Neuro-Symbolic Artificial Intelligence(NeSyAI). More specifically, it (1) enriches NeSyAI systems with explicit representations of causality,(2) integrates causal knowledge with domain …

[PDF] [On the statistical analysis of experiments with manipulation checks](#)

MB Mathur

Social scientists are often interested in the effect of an internal state, such as ego depletion or social stress, that cannot be directly assigned in an experiment. Instead, they might assign participants to a manipulation intended to produce this internal …

[PDF] [LLM-Enhanced Causal Discovery in Temporal Domain from Interventional Data](#)

P Li, X Wang, Z Zhang, Y Meng, F Shen, Y Li, J Wang… - arXiv preprint arXiv …, 2024

In the field of Artificial Intelligence for Information Technology Operations, causal discovery is pivotal for operation and maintenance of graph construction, facilitating downstream industrial tasks such as root cause analysis. Temporal causal discovery …

[PDF] [Inference of Causal Networks using a Topological Threshold](#)

F Barroso, D Gomes, GJ Baxter - arXiv preprint arXiv:2404.14460, 2024

We propose a constraint-based algorithm, which automatically determines causal relevance thresholds, to infer causal networks from data. We call these topological thresholds. We present two methods for determining the threshold: the first seeks a …

[PDF] [Identifying sparse treatment effects](#)

Y Jeong, E Fox, R Johari - arXiv preprint arXiv:2404.14644, 2024

Based on technological advances in sensing modalities, randomized trials with primary outcomes represented as high-dimensional vectors have become increasingly prevalent. For example, these outcomes could be week-long time-series …

[PDF] [Mining Invariance from Nonlinear Multi-Environment Data: Binary Classification](#)

A Goddard, K Du, Y Xiang - arXiv preprint arXiv:2404.15245, 2024

Making predictions in an unseen environment given data from multiple training environments is a challenging task. We approach this problem from an invariance perspective, focusing on binary classification to shed light on general nonlinear data …

[Kausale Inferenz](#)

E Brox, M Lechner, C Steckenleiter - 2024

In diesem Kapitel diskutieren wir die Grundlagen der kausalen Inferenz. Dabei wird besonderer Wert auf die Beschreibung der unterschiedlichen Annahmen gelegt, unter denen es im Prinzip möglich wäre, kausale Effekte zu schätzen. Welche dieser …

[PDF] [Adaptive rationality in communication](#)

C Pilgrim - 2023

Modern telecommunications have transformed the way that people communicate. The situation is dynamic, with a rapidly changing technological and cultural landscape. Furthermore, interactions between this landscape and human behaviour …

[PDF] [HAL Id: hal-04391010](#)

F Poggiolesi

To explain phenomena in the world is a central human activity and one of the main goals of rational inquiry. There are several types of explanation: one can explain by drawing an analogy, as one can explain by dwelling on the causes (see eg see …

[Reliable Estimation of Causal Effects Using Predictive Models](#)

MH Ali, YL Biannic, PH Wuillemin - International Journal on Artificial Intelligence Tools, 2024

In recent years, machine learning algorithms have been widely adopted across many fields due to their efficiency and versatility. However, the complexity of predictive models has led to a lack of interpretability in automatic decision-making …

[How transportability analysis can be useful for cumulative theory testing in management research](#)

GKF Lee - Journal of Management Scientific Reports, 2024

Moving management science forward by publishing research aimed at theory testing and refinement is the mission of the Journal of Management Scientific Reports. One way to test and refine a theory is through constructive replication studies. If $X \rightarrow Y$ (the …

[Modeling Treatment Effect with Cross-Domain Data](#)

B Han, YL Zhang, L Yu, B Chen, L Li, J Zhou - Pacific-Asia Conference on Knowledge …, 2024

Abstract Treatment effect estimation has received increasing attention recently. However, the issue of data sparsity often poses a significant challenge, limiting the feasibility of modeling. This paper aims to leverage cross-domain data to mitigate the …

[Distributed Discovery of Causal Networks in Pervasive Environments](#)

S Mariani, F Zambonelli - 2024 IEEE International Conference on Pervasive …, 2024

In pervasive computing environments, learning the causal network of relationships between environmental variables is crucial to support situation recognition and planning. However, this may be impossible when computing nodes have only partial …

[PDF] [SIDEs: Separating Idealization from Deceptive 'Explanations' in xAI](#)

E Sullivan - 2024

Explainable AI (xAI) methods are important for establishing trust in using black-box models. However, recent criticism has mounted against current xAI methods that they disagree, are necessarily false, and can be manipulated, which has started to …

A Primer on Dominance Analysis

F Bittmann - 2024

Regression models are highly popular in empirical research and come in many different forms to fit virtually any distribution, variable, or research question. Usually, these models also compute how much variation in the outcome variable can be …

Criticality-aware Deconfounded Classification of Long-tailed Multi-label 12-lead Electrocardiogram

T Deb, I Sahu, A Ukil, A Pal, S Khandelwal, U Garain - 2024 IEEE International …, 2024

We often observe long-tailed distribution in real-world classification problems and consequently, maintaining balanced predictive performance across all the classes is a research challenge. Further, we find, particularly in time series classification tasks …

Synergies Between Machine Learning and Reasoning-An Introduction by the Kay R. Amel group

I Baaj, Z Bouraoui, A Cornuéjols, T Denœux… - International Journal of …, 2024

This paper proposes a tentative and original survey of meeting points between Knowledge Representation and Reasoning (KRR) and Machine Learning (ML), two areas which have been developed quite separately in the last four decades. First …

From algorithms to action: improving patient care requires causality

WAC van Amsterdam, PA de Jong, JJC Verhoeff… - BMC Medical Informatics …, 2024

In cancer research there is much interest in building and validating outcome prediction models to support treatment decisions. However, because most outcome prediction models are developed and validated without regard to the causal aspects …

Concept Paper for a Digital Expert: Systematic Derivation of (Causal) Bayesian Networks Based on Ontologies for Knowledge-Based Production Steps

MML Pfaff-Kastner, K Wenzel, S Ihlenfeldt - Machine Learning and Knowledge …, 2024

MAKE | Free Full-Text | Concept Paper for a Digital Expert: Systematic Derivation of (Causal) Bayesian Networks Based on Ontologies for Knowledge-Based Production Steps Previous Article in Journal Enhancing Legal Sentiment Analysis: A Convolutional Neural …

Mediating effect of BMI on the association of economic status and coexistence of hypertension and diabetes in Bangladesh: A counterfactual framework-based …

F Ahmmed, MJ Hossain, MTF Khan, MMR Manik… - Health Science Reports, 2024

Abstract Background and Aims Non-communicable diseases such as hypertension and diabetes are matters of huge concern worldwide, with an increasing trend in prevalence over the previous decade. First of all, this study aimed to evaluate the …

Evolutionary Causal Discovery with Relative Impact Stratification for Interpretable Data Analysis

O Deng, S Nishimura, A Ogihara, Q Jin - arXiv preprint arXiv:2404.16361, 2024

This study proposes Evolutionary Causal Discovery (ECD) for causal discovery that tailors response variables, predictor variables, and corresponding operators to research datasets. Utilizing genetic programming for variable relationship parsing …

[CausalBO: A Python Package for Causal Bayesian Optimization](#)

J Roberts, MA Javidian - SoutheastCon 2024, 2024

This paper introduces CausalBO, a Python package developed to enhance the applicability and utility of the Causal Bayesian Optimization (CBO) algorithm. The original CBO algorithm, developed by Virginia Aglietti et al.[1], integrated causality …

[PDF] [In Silico Sociology: Forecasting COVID-19 Polarization with Large Language Models](#)

A Kozlowski, H Kwon, J Evans

By training deep neural networks on massive archives of digitized text, large language models (LLMs) learn the complex linguistic patterns that constitute historic and contemporary discourses. We argue that LLMs can serve as a valuable tool for …

[PDF] [Improving Students' Understanding through Metacognition About Instructor Feedback: A Causal Modeling Approach](#)

SY Wood, V Cross

Previous research has documented learning benefits from metacognitive engagement and suggested many instructional practices to prompt metacognition in students. For example, a scaffolded curriculum that includes effective feedback may …

[PDF] [Axiomatic Causal Interventions for Reverse Engineering Relevance Computation in Neural Retrieval Models](#)

C Chen, J Merullo, C Eickhoff - 2024

Neural models have demonstrated remarkable performance across diverse ranking tasks. However, the processes and internal mechanisms along which they determine relevance are still largely unknown. Existing approaches for analyzing neural ranker …

[The development of human causal learning and reasoning](#)

MK Goddu, A Gopnik - Nature Reviews Psychology, 2024

Causal understanding is a defining characteristic of human cognition. Like many animals, human children learn to control their bodily movements and act effectively in the environment. Like a smaller subset of animals, children intervene: they learn to …

[Advances in Knowledge Discovery and Data Mining: 28th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2024, Taipei, Taiwan, May 7 …](#)

DN Yang

Recent years have seen rapid progress in the field of representation learning, where self-supervised learning methods such as SimCLR [4] and the MoCo family [5, 10] have gained popularity and shown remarkable abilities to extract features …

[HTML] [Time-lagged panel models in psychotherapy process and mechanisms of change research: Methodological challenges and advances](#)

F Falkenström - Clinical Psychology Review, 2024

In recent years, there has been increasing interest in utilizing time-lagged panel models to study mechanisms of change in psychotherapy. These models offer valuable insights into the dynamic relationships between variables over time and …

[PDF] [Bayesian Learning of Causal Networks for Unsupervised Fault Diagnosis in Distributed Energy Systems](#)

F Castelletti, F Niro, M Denti, D Tessera, A Pozzi - IEEE Access, 2024

Distributed energy generation systems, key for producing electricity near usage points, are essential to meet the global electricity demand, leveraging diverse sources like renewables, traditional fuels, and industrial waste heat. Despite their …

[PDF] [RECONCILING MODEL-X AND DOUBLY ROBUST APPROACHES TO CONDITIONAL INDEPENDENCE TESTING BY ZIANG NIU, ABHINAV CHAKRABORTY](#)

O DUKES, E KATSEVICH

Model-X approaches to testing conditional independence between a predictor and an outcome variable given a vector of covariates usually assume exact knowledge of the conditional distribution of the predictor given the covariates. Nevertheless, model …

[PDF] [TARIFVERTRÄGE UND TARIFFLUCHT IN BRANDENBURG](#)

T Schulten, R Bispinck, M Lübker, Ş Erol

ZUSAMMENFASSUNG In Brandenburg arbeiten im Jahr 2022 noch 47 Prozent aller Beschäftigten in einem Betrieb mit Tarifvertrag. In den letzten Jahrzehnten ist die Tarifbindung kontinuierlich zurückgegangen. Dies hat zu einer deutlichen …