Novelty Assessment Report

Paper: Exploratory Causal Inference in SAEnce
PDF URL: https://openreview.net/pdf?id=Ml8t8kQMUP
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-27

Abstract

Randomized Controlled Trials are one of the pillars of science; nevertheless, they rely on hand-crafted hypotheses and expensive analysis. Such constraints prevent causal effect estimation at scale, potentially anchoring on popular yet incomplete hypotheses. We propose to discover the unknown effects of a treatment directly from data. For this, we turn unstructured data from a trial into meaningful representations via pretrained foundation models and interpret them via a Sparse Auto Encoder. However, discovering significant causal effects at the neural level is not trivial due to multiple-testing issues and effects entanglement. To address these challenges, we introduce Neural Effect Search, a novel recursive procedure solving both issues by progressive stratification. After assessing the robustness of our algorithm on semi-synthetic experiments, we showcase, in the context of experimental ecology, the first successful unsupervised causal effect identification on a real-world scientific trial.

Core Task Landscape

This paper addresses: **Unsupervised Causal Effect Discovery from High-Dimensional Experimental Data**

A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Causal Structure Discovery from Observational Data**
- **Treatment Effect Estimation and Causal Inference**
- **Semi-Supervised and Unlabeled Causal Inference**
- **Causal Inference Under Distribution Shift**
- **Causal Feature Selection and Discovery**
- **Specialized Causal Inference Methods**

Complete Taxonomy Tree

- Unsupervised Causal Effect Discovery from High-Dimensional Experimental Data Survey Taxonomy
- Causal Structure Discovery from Observational Data
  - Constraint-Based and Score-Based Causal Discovery
  - High-Dimensional Multivariate Causal Discovery (4 papers)
    - [2] Information-based estimation of causality networks from high-dimensional multivariate time series (Akylas Fotiadis, 2023) View paper
    - [6] Causal discovery on vector-valued variables and consistency-guided aggregation (Ninad, 2025) View paper
    - [14] ALCM: Autonomous LLM-Augmented Causal Discovery Framework (Abbasian, 2024) View paper
    - [18] Weakly Supervised Causal Discovery Based on Fuzzy Knowledge and Complex Data Complementarity (Wenrui Li, 2024) View paper
  - Temporal and Dynamical Causal Discovery (4 papers)
    - [8] Causal Discovery in Nonlinear Dynamical Systems using Koopman Operators (Rupe, 2024) View paper
    - [10] Linear Scaling Causal Discovery from High-Dimensional Time Series by Dynamical Community Detection (Del Tatto, 2025) View paper
    - [44] Causal Inference on Process Graphs, Part I: The Structural Equation Process Representation (Reiter, 2023) View paper
    - [46] Robust inference of causality in high-dimensional dynamical processes from the Information Imbalance of distance ranks (Fortunato, 2023) View paper
  - Event Sequence and Sparse Causal Discovery (2 papers)
    - [12] One-Shot Multi-Label Causal Discovery in High-Dimensional Event Sequences (Schon, 2025) View paper
    - [23] Towards Practical Multi-label Causal Discovery in High-Dimensional Event Sequences via One-Shot Graph Aggregation (Lienhart, 2025) View paper
  - Local and Hybrid Causal Discovery (2 papers)
    - [41] Hybrid Local Causal Discovery (Zhaolong Ling, 2024) View paper
    - [47] Neural Networks as Universal Function Approximators for Causal Discovery with Reinforcement Learning (Zoccheddu, 2023) View paper
  - Deep Causal Representation Learning
  - Identifiable Causal Representation Learning (2 papers)
    - [11] Identifiable causal representation learning: Unsupervised, multi-view, and multi-environment (von KÃ¼gelgen, 2024) View paper
    - [38] Differentiable Causal Discovery For Latent Hierarchical Causal Models (Ng, 2024) View paper

- Variational and Generative Causal Models (2 papers)
  - [3] Causal discovery in physical systems from videos (Li, 2020) View paper
  - [7] Causal discovery for modular world models (A Lei, 2022) View paper
- Bivariate and Inverse Regression Causal Discovery (3 papers)
  - [27] Inference of cause and effect with unsupervised inverse regression (Eleni Sgouritsa, 2015) View paper
  - [36] Causal Order Discovery based on Monotonic SCMs (Izadi Ali, 2024) View paper
  - [45] Using unlabeled data to discover bivariate causality with deep restricted Boltzmann machines (Nataliya Sokolovska, 2018) View paper
- Domain-Specific Causal Discovery Applications (4 papers)
- [19] Causal Feature Selection for Materials Optimization: Identifying Causal Process Parameters via Machine Learning and High-Dimensional Hypothesis Testing (Liu, 2025) View paper
- [20] Unsupervised discovery of el nino using causal feature learning on microlevel climate data (Chalupka, 2016) View paper
- [24] Causal CloudScape (CCS): A Novel Approach to Causal Discovery of High-Dimensional Low-Level Cloud Performance Metrics (Zenner, 2024) View paper
- [49] Aristotle: stratified causal discovery for omics data (Mansouri Mehrdad, 2022) View paper

- Treatment Effect Estimation and Causal Inference
  - High-Dimensional Covariate Adjustment (4 papers)
  - [15] Probably approximately correct high-dimensional causal effect estimation given a valid adjustment set (Choo, 2024) View paper
  - [16] Outcome model free causal inference with ultra-high dimensional covariates. (Dingke Tang, 2020) View paper
  - [17] Dimension Reduction for Conditional Density Estimation with Applications to High-Dimensional Causal Inference (Mei Jian-hua, 2025) View paper
  - [40] High Dimensional Causal Inference with Variational Backdoor Adjustment (Israel, 2023) View paper
  - Multivariate and Conditional Treatment Effect Estimation (2 papers)
  - [1] A copula-based deep graphical causal model for multivariate conditional treatment effect estimation (Kim, 2025) View paper
  - [30] Causal Effect Identification in Cluster DAGs (Bareinboim, 2022) View paper
  - Unsupervised Causal Effect Discovery ★ (3 papers)
  - [0] Exploratory Causal Inference in SAEnce (Anon et al., 2026) View paper
  - [4] Knowledge discovery: from correlation to causation (Arab, 2024) View paper
  - [22] Making Interpretable Discoveries from Unstructured Data: A High-Dimensional Multiple Hypothesis Testing Approach (Carlson, 2025) View paper
- Semi-Supervised and Unlabeled Causal Inference
  - Semi-Supervised Regression and Causal Inference (3 papers)
  - [5] Semi-supervised regression analysis with model misspecification and high-dimensional data (Ye Tian, 2024) View paper
  - [9] Semiparametric semi-supervised learning for general targets under distribution shift and decaying overlap (Testa, 2025) View paper
  - [34] On Statistical Learning for Structural Data: Data Fusion and Semi-Supervised Learning (Jiang, 2025) View paper
  - Positive-Unlabeled and Control Group Construction (1 papers)
  - [33] Positive-Unlabeled Learning for Control Group Construction in Observational Causal Inference (Tsoumas, 2025) View paper
- Causal Inference Under Distribution Shift
  - Causal Domain Adaptation (2 papers)
  - [13] Causal inference-based adversarial domain adaptation for cross-domain industrial intrusion detection (Yongle Chen, 2024) View paper
  - [28] Sparse Causal Discovery with Generative Intervention for Unsupervised Graph Domain Adaptation (Luo Jun-Yu, 2025) View paper
  - Time Series Causal Domain Adaptation (1 papers)
  - [29] Time Series Domain Adaptation via Latent Invariant Causal Mechanism (Cai Rui-chu, 2025) View paper
- Causal Feature Selection and Discovery
  - Unsupervised Causal Feature Selection (2 papers)
  - [31] Beyond Correlation: Causal Multi-View Unsupervised Feature Selection Learning (Shen, 2025) View paper
  - [39] Causally-Aware Unsupervised Feature Selection Learning (Shen, 2024) View paper
  - Supervised Causal Feature Selection (1 papers)
  - [21] High-speed train fault detection with unsupervised causality-based feature extraction methods (Yubo Xu, 2021) View paper
- Specialized Causal Inference Methods
  - Uncertainty and False Discovery Control (2 papers)
  - [25] Uncertainty Assessment and False Discovery Rate Control in High-Dimensional Granger Causal Inference (Aditya Chaudhry, 2017) View paper
  - [48] Confidence in Causal Inference under Structure Uncertainty in Linear Causal Models with Equal Variances (David Strieder, 2023) View paper
  - Interpretable and Explainable Causal Inference (2 papers)
  - [26] Establishing Causal Relationship Between Whole Slide Image Predictions and Diagnostic Evidence Subregions in Deep Learning (Nan, 2024) View paper
  - [32] From Exploration to Explanation: ML-Driven Causal Discovery for Datacenter Reliability at Scale (Pavana Prakash, 2025) View paper
  - Tractable Causal Inference and Mechanism Learning (2 papers)
  - [37] Mechanism Learning: reverse causal inference in the presence of multiple unknown confounding through causally weighted Gaussian mixture models (Mao, 2024) View paper
  - [50] Causal Inference Using Tractable Circuits (Darwiche, 2022) View paper
  - Deep Learning for Causal Inference (2 papers)
  - [35] Deep learning for causal inference (Xiong, 2018) View paper
  - [42] Machine Learning Models for High-Dimensional Matched Data (Nooshin Shomal Zadeh, 2021) View paper
  - Causal Inference in Specialized Domains (1 papers)
  - [43] Causality in Unsupervised Learning: Methods and Applications in Cancer Genomics (Bayer, 2024) View paper

Narrative

Core task: unsupervised causal effect discovery from high-dimensional experimental data. The field addresses the challenge of identifying causal relationships and treatment effects when labeled outcome data or explicit intervention annotations are scarce or absent, particularly in settings where the dimensionality of covariates is large. The taxonomy reflects a broad landscape organized around six main branches. Causal Structure Discovery from Observational Data focuses on learning directed acyclic graphs and structural equation models from passive observations, often leveraging independence constraints or functional form assumptions. Treatment Effect Estimation and Causal Inference encompasses methods for quantifying intervention impacts, including both supervised approaches that rely on known treatment assignments and unsupervised variants that must infer effects without direct outcome labels. Semi-Supervised and Unlabeled Causal Inference explores hybrid settings where partial supervision or positive-unlabeled data guides discovery. Causal Inference Under Distribution Shift tackles robustness when training and test distributions differ, while Causal Feature Selection and Discovery aims to identify causally relevant variables in high-dimensional spaces. Specialized Causal Inference Methods includes domain-specific techniques and novel algorithmic frameworks, such as those integrating large language models or leveraging representation learning.

Several active lines of work highlight key trade-offs and open questions. One prominent theme is the tension between model flexibility and identifiability: representation learning approaches like Identifiable Causal Representation[11] and Latent Invariant Mechanism[29] seek to uncover latent causal factors from complex observations, yet must impose structural constraints to ensure uniqueness. Another contrast appears between methods that assume known interventions versus those that operate in fully unsupervised regimes. SAEnce Causal Inference[0] sits within the Unsupervised Causal Effect Discovery cluster, emphasizing the extraction of causal signals from experimental data without explicit outcome labels. It shares this unsupervised orientation with Correlation to Causation[4], which also aims to move beyond associational patterns, and contrasts with semi-supervised strategies like Semi-supervised Misspecification[5] that blend labeled and unlabeled information. The original work's focus on high-dimensional experimental settings positions it at the intersection of structure discovery and effect estimation, addressing scenarios where traditional supervised methods are infeasible yet experimental perturbations provide crucial leverage for causal identification.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Knowledge discovery: from correlation to causation

Authors: A Arab | **Year/Venue**: 2024 | **URL**: View paper

Abstract

â¦ Lastly, we introduce Hegel, a method for compound causal discovery. Hegel employs a â¦ discovery within unstructured, high-dimensional data sets, particularly images. Subgroup â¦

Relationship Analysis

Both papers belong to the Unsupervised Causal Effect Discovery category, focusing on discovering unknown causal effects from experimental data without predefined hypotheses. The original paper (Exploratory Causal Inference in SAEnce) specifically addresses discovering treatment effects from high-dimensional experimental data using sparse autoencoders and foundation models, proposing Neural Effect Search to handle multiple testing and entanglement issues. The candidate paper (Knowledge Discovery: from Correlation to Causation) is a broader doctoral thesis covering the entire knowledge discovery pipeline from hypothesis generation to validation, including subgroup discovery in images, biomarker discovery, temporal causal discovery, and compound causal discovery, representing a more comprehensive treatment of moving from correlational to causal analysis across multiple domains rather than focusing specifically on unsupervised effect discovery from high-dimensional experimental measurements.

---

2. Making Interpretable Discoveries from Unstructured Data: A High-Dimensional Multiple Hypothesis Testing Approach

Authors: Carlson, Jacob | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Social scientists are increasingly turning to unstructured datasets to unlock new empirical insights, e.g., estimating causal effects on text outcomes, measuring beliefs from open-ended survey responses. In such settings, unsupervised analysis is often of interest, in that the researcher does not want to pre-specify the objects of measurement or otherwise artificially delimit the space of measurable concepts; they are interested in discovery. This paper proposes a general and flexible framework ...

Relationship Analysis

Both papers belong to the Unsupervised Causal Effect Discovery category, focusing on discovering unknown causal effects from experimental data without predefined hypotheses. They overlap in their use of high-dimensional representations (the original uses sparse autoencoders on foundation model features, the candidate uses interpretable dictionaries from unstructured data) and multiple hypothesis testing frameworks to identify significant effects. The key difference is that the original paper specifically targets randomized controlled trials with neural representations and proposes Neural Effect Search to handle entanglement issues, while the candidate paper provides a more general framework for social science applications using k-FWER control under arbitrary dependence without the recursive stratification approach.

Contributions Analysis

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Formal differentiation of rationalist and empiricist approaches to causal inference

Description: The authors establish a formal framework distinguishing between rationalist approaches (hypothesis-driven causal inference with predefined outcomes) and empiricist approaches (data-driven discovery of treatment effects). They characterize these paradigms within statistical causality, showing how they complement each other in scientific discovery.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Mechanisms and mechanistic reasoning in medicine

URL: View paper

Brief Assessment

Mechanistic Reasoning Medicine[58] discusses rationalism and empiricism in medical methodology but does not address statistical causality frameworks or treatment effect estimation in randomized trials, which are central to the original paper's contribution.

---

2. Between rationalism and empiricism

URL: View paper

Brief Assessment

Rationalism and Empiricism[59] appears to be a philosophical work discussing epistemological frameworks broadly, not a technical paper on statistical causal inference methodologies distinguishing hypothesis-driven versus data-driven treatment effect discovery.

3. Causometry

URL: View paper

Brief Assessment

Causometry[64] briefly mentions rationalist and empiricist views in passing but does not provide a formal framework distinguishing these paradigms within statistical causality or characterize how they complement each other in scientific discovery.

4. Radical empiricism and machine learning research

URL: View paper

Brief Assessment

Radical Empiricism Machine[56] discusses 'data fitting' vs 'data interpreting' approaches in data science broadly, not a formal framework for rationalist versus empiricist paradigms in statistical causal inference specifically. The candidate focuses on philosophical contrasts in data science methodology rather than formalizing these approaches within the causal inference framework as the original paper does.

5. Aristotle's Induction and the Inference of First Principles

URL: View paper

Brief Assessment

Aristotle Induction Principles[63] addresses philosophical epistemology in Aristotle's theory of scientific knowledge and induction, not statistical causal inference frameworks. The domains are fundamentally different—ancient philosophy versus modern statistics.

6. Logical empiricism

URL: View paper

Brief Assessment

Logical Empiricism[57] discusses philosophical movements and views on causality, rationality, and explanation in general terms, but does not present a formal statistical framework distinguishing rationalist versus empiricist approaches to causal inference in the context of treatment effect estimation or randomized controlled trials.

7. Truth, knowledge, and entrepreneurship theory: arguments for a rationalist scientific epistemology

URL: View paper

Brief Assessment

Entrepreneurship Theory Rationalism[55] discusses rationalist versus empiricist epistemology in entrepreneurship theory, not statistical causality frameworks. The candidate focuses on philosophical approaches to entrepreneurship research rather than formal statistical methods for causal inference.

8. Causal learning in rats and humans: A minimal rational model

URL: View paper

Brief Assessment

Minimal Rational Causal[62] discusses rationalist versus empiricist epistemology in the context of causal learning in rats and humans, but focuses on cognitive modeling rather than statistical causality frameworks for scientific discovery.

9. Realism, empiricism and causal inquiry in International Relations: What is at stake?

URL: View paper

Brief Assessment

Realism Empiricism Relations[60] examines philosophical debates between scientific realism and empiricism in International Relations theory, not statistical causality frameworks. The paper focuses on metaphysical and epistemological positions regarding unobservable entities rather than formal statistical methodologies for causal inference.

10. Method and Analogy in Hellenistic Medicine

URL: View paper

Brief Assessment

Hellenistic Medicine Analogy[61] examines rationalist versus empiricist methodologies in ancient Greek medicine (3rd century BC), focusing on observation and analogy in medical practice. The ORIGINAL paper addresses modern statistical causality frameworks distinguishing hypothesis-driven versus data-driven causal inference in contemporary machine learning contexts—a fundamentally different domain and temporal context.

Contribution 2: Novel empiricist methodology using foundation models and sparse autoencoders

Description: The authors introduce a methodology that combines pretrained foundation models with sparse autoencoders to discover treatment effects in exploratory experiments. They identify and formalize the paradox of exploratory causal inference, showing how standard multiple testing fails when neural representations are entangled.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Prototype-Based Multiple Instance Learning for Gigapixel Whole Slide Image Classification

URL: View paper

Brief Assessment

Prototype Multiple Instance[74] focuses on interpretable multiple instance learning for histopathology image classification using sparse autoencoders to discover human-interpretable concepts. The original paper addresses exploratory causal inference in randomized controlled trials using foundation models and sparse autoencoders to discover treatment effects. These are fundamentally different applications with distinct methodological goals.

2. Sparse autoencoders for scientifically rigorous interpretation of vision models

URL: View paper

Brief Assessment

Sparse Autoencoders Vision[65] focuses on interpreting and manipulating visual features in vision models for model understanding and editing tasks, not on discovering treatment effects in randomized controlled trials or addressing exploratory causal inference challenges.

3. Sparse Auto-Encoder Interprets Linguistic Features in Large Language Models

URL: View paper

Brief Assessment

SAE Linguistic Features[67] focuses on interpreting linguistic mechanisms in LLMs using SAEs for linguistic feature extraction (morphology, syntax, semantics, pragmatics), not on discovering treatment effects in exploratory causal inference experiments. The candidate does not address causal effect estimation, randomized controlled trials, or the paradox of exploratory causal inference that the original paper tackles.

---

4. Saes can improve unlearning: Dynamic sparse autoencoder guardrails for precision unlearning in llms

URL: View paper

Brief Assessment

SAE Dynamic Guardrails[73] focuses on machine unlearning in LLMs using sparse autoencoders for targeted activation-based interventions, not on discovering treatment effects in exploratory causal inference experiments. The candidate addresses a fundamentally different problem domain (model safety/unlearning) than the original's causal discovery methodology.

---

5. Improving Steering Vectors by Targeting Sparse Autoencoder Features

URL: View paper

Brief Assessment

Targeting SAE Features[66] focuses on steering language model behavior using SAE features for model control, not on discovering treatment effects in randomized controlled trials or exploratory causal inference.

---

6. Can role vectors affect llm behaviour

URL: View paper

Brief Assessment

Role Vectors LLM[72] focuses on steering LLM behavior through role vectors derived from activation differences, not on discovering treatment effects in randomized controlled trials or addressing exploratory causal inference challenges.

---

7. Applying sparse autoencoders to unlearn knowledge in language models

URL: View paper

Brief Assessment

SAE Unlearn Knowledge[68] applies SAEs to remove knowledge from language models in a safety context, not to discover treatment effects in randomized controlled trials or exploratory causal inference.

---

8. Sparse autoencoders reveal temporal difference learning in large language models

URL: View paper

Brief Assessment

SAE Temporal Difference[71] focuses on discovering temporal difference learning mechanisms in LLMs for reinforcement learning tasks, not on discovering treatment effects in randomized controlled trials or addressing exploratory causal inference challenges.

---

9. SAIF: A Sparse Autoencoder Framework for Interpreting and Steering Instruction Following of Language Models

URL: View paper

Brief Assessment

SAIF Instruction Following[70] focuses on interpreting instruction-following mechanisms in LLMs using SAEs for model steering, not on discovering treatment effects in randomized controlled trials or addressing exploratory causal inference challenges.

---

10. A Deep Learning Framework for Causal Inference in Clinical Trial Design: The CURE AI Large Clinicogenomic Foundation Model

URL: View paper

Brief Assessment

CURE AI Foundation[69] focuses on clinical trial design and causal inference for treatment effect prediction in medical contexts, not on discovering treatment effects through sparse autoencoders applied to neural representations for exploratory causal inference.

Contribution 3: Neural Effect Search algorithm for iterative hypothesis testing

Description: The authors develop Neural Effect Search, a recursive stratification procedure that addresses multiple-testing issues and effect entanglement in neural representations. The algorithm iteratively identifies significant causal effects while controlling for dependencies between neurons through progressive stratification.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Applications of general multistage gatekeeping and graphical multiple testing strategies in a clinical trial setting

URL: View paper

Brief Assessment

Multistage Gatekeeping Strategies[52] focuses on clinical trial designs for testing multiple endpoints/doses with graphical representations, not on neural representations or effect entanglement in machine learning contexts.

---

2. A Multiparty Quantum Private Equality Comparison Scheme Relying on |GHZ3â© States

URL: View paper

Brief Assessment

Quantum Private Equality[51] addresses quantum cryptographic comparison protocols using entangled states, not iterative hypothesis testing for neural representations or multiple testing in causal inference.

---

3. Detection of genuine tripartite entanglement by multiple sequential observers

URL: View paper

Brief Assessment

Tripartite Entanglement Detection[54] addresses quantum entanglement detection in sequential measurement scenarios, which is fundamentally different from the ORIGINAL paper's iterative hypothesis testing for multiple testing with neural representations in causal inference.

4. Modeling Tactics as Operators: Effect-Grounded Representations for Lean Theorem Proving

URL: View paper

Brief Assessment

Tactics as Operators[53] addresses tactic representation learning in theorem proving, not iterative hypothesis testing for multiple testing with entangled representations in causal inference contexts.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Exploratory Causal Inference in SAEnce View paper
- [1] A copula-based deep graphical causal model for multivariate conditional treatment effect estimation View paper
- [2] Information-based estimation of causality networks from high-dimensional multivariate time series View paper
- [3] Causal discovery in physical systems from videos View paper
- [4] Knowledge discovery: from correlation to causation View paper
- [5] Semi-supervised regression analysis with model misspecification and high-dimensional data View paper
- [6] Causal discovery on vector-valued variables and consistency-guided aggregation View paper
- [7] Causal discovery for modular world models View paper
- [8] Causal Discovery in Nonlinear Dynamical Systems using Koopman Operators View paper
- [9] Semiparametric semi-supervised learning for general targets under distribution shift and decaying overlap View paper
- [10] Linear Scaling Causal Discovery from High-Dimensional Time Series by Dynamical Community Detection View paper
- [11] Identifiable causal representation learning: Unsupervised, multi-view, and multi-environment View paper
- [12] One-Shot Multi-Label Causal Discovery in High-Dimensional Event Sequences View paper
- [13] Causal inference-based adversarial domain adaptation for cross-domain industrial intrusion detection View paper
- [14] ALCM: Autonomous LLM-Augmented Causal Discovery Framework View paper
- [15] Probably approximately correct high-dimensional causal effect estimation given a valid adjustment set View paper
- [16] Outcome model free causal inference with ultra-high dimensional covariates. View paper
- [17] Dimension Reduction for Conditional Density Estimation with Applications to High-Dimensional Causal Inference View paper
- [18] Weakly Supervised Causal Discovery Based on Fuzzy Knowledge and Complex Data Complementarity View paper
- [19] Causal Feature Selection for Materials Optimization: Identifying Causal Process Parameters via Machine Learning and High-Dimensional Hypothesis Testing View paper
- [20] Unsupervised discovery of el nino using causal feature learning on microlevel climate data View paper
- [21] High-speed train fault detection with unsupervised causality-based feature extraction methods View paper
- [22] Making Interpretable Discoveries from Unstructured Data: A High-Dimensional Multiple Hypothesis Testing Approach View paper
- [23] Towards Practical Multi-label Causal Discovery in High-Dimensional Event Sequences via One-Shot Graph Aggregation View paper
- [24] Causal CloudScape (CCS): A Novel Approach to Causal Discovery of High-Dimensional Low-Level Cloud Performance Metrics View paper
- [25] Uncertainty Assessment and False Discovery Rate Control in High-Dimensional Granger Causal Inference View paper
- [26] Establishing Causal Relationship Between Whole Slide Image Predictions and Diagnostic Evidence Subregions in Deep Learning View paper
- [27] Inference of cause and effect with unsupervised inverse regression View paper
- [28] Sparse Causal Discovery with Generative Intervention for Unsupervised Graph Domain Adaptation View paper
- [29] Time Series Domain Adaptation via Latent Invariant Causal Mechanism View paper
- [30] Causal Effect Identification in Cluster DAGs View paper
- [31] Beyond Correlation: Causal Multi-View Unsupervised Feature Selection Learning View paper
- [32] From Exploration to Explanation: ML-Driven Causal Discovery for Datacenter Reliability at Scale View paper
- [33] Positive-Unlabeled Learning for Control Group Construction in Observational Causal Inference View paper
- [34] On Statistical Learning for Structural Data: Data Fusion and Semi-Supervised Learning View paper
- [35] Deep learning for causal inference View paper
- [36] Causal Order Discovery based on Monotonic SCMs View paper
- [37] Mechanism Learning: reverse causal inference in the presence of multiple unknown confounding through causally weighted Gaussian mixture models View paper
- [38] Differentiable Causal Discovery For Latent Hierarchical Causal Models View paper
- [39] Causally-Aware Unsupervised Feature Selection Learning View paper
- [40] High Dimensional Causal Inference with Variational Backdoor Adjustment View paper
- [41] Hybrid Local Causal Discovery View paper
- [42] Machine Learning Models for High-Dimensional Matched Data View paper
- [43] Causality in Unsupervised Learning: Methods and Applications in Cancer Genomics View paper
- [44] Causal Inference on Process Graphs, Part I: The Structural Equation Process Representation View paper
- [45] Using unlabeled data to discover bivariate causality with deep restricted Boltzmann machines View paper
- [46] Robust inference of causality in high-dimensional dynamical processes from the Information Imbalance of distance ranks View paper
- [47] Neural Networks as Universal Function Approximators for Causal Discovery with Reinforcement Learning View paper
- [48] Confidence in Causal Inference under Structure Uncertainty in Linear Causal Models with Equal Variances View paper
- [49] Aristotle: stratified causal discovery for omics data View paper
- [50] Causal Inference Using Tractable Circuits View paper
- [51] A Multiparty Quantum Private Equality Comparison Scheme Relying on |GHZ3â© States View paper
- [52] Applications of general multistage gatekeeping and graphical multiple testing strategies in a clinical trial setting View paper

- [53] Modeling Tactics as Operators: Effect-Grounded Representations for Lean Theorem Proving View paper
- [54] Detection of genuine tripartite entanglement by multiple sequential observers View paper
- [55] Truth, knowledge, and entrepreneurship theory: arguments for a rationalist scientific epistemology View paper
- [56] Radical empiricism and machine learning research View paper
- [57] Logical empiricism View paper
- [58] Mechanisms and mechanistic reasoning in medicine View paper
- [59] Between rationalism and empiricism View paper
- [60] Realism, empiricism and causal inquiry in International Relations: What is at stake? View paper
- [61] Method and Analogy in Hellenistic Medicine View paper
- [62] Causal learning in rats and humans: A minimal rational model View paper
- [63] Aristotle's Induction and the Inference of First Principles View paper
- [64] Causometry View paper
- [65] Sparse autoencoders for scientifically rigorous interpretation of vision models View paper
- [66] Improving Steering Vectors by Targeting Sparse Autoencoder Features View paper
- [67] Sparse Auto-Encoder Interprets Linguistic Features in Large Language Models View paper
- [68] Applying sparse autoencoders to unlearn knowledge in language models View paper
- [69] A Deep Learning Framework for Causal Inference in Clinical Trial Design: The CURE AI Large Clinicogenomic Foundation Model View paper
- [70] SAIF: A Sparse Autoencoder Framework for Interpreting and Steering Instruction Following of Language Models View paper
- [71] Sparse autoencoders reveal temporal difference learning in large language models View paper
- [72] Can role vectors affect llm behaviour View paper
- [73] Saes can improve unlearning: Dynamic sparse autoencoder guardrails for precision unlearning in llms View paper
- [74] Prototype-Based Multiple Instance Learning for Gigapixel Whole Slide Image Classification View paper