Analyses & Alternatives 2024 8(3): 125-150
DOI: 10.22931/aanda.2024.8.3.005

Integration of AI, Causality, and Social Sciences: Understanding Social Phenomena through Causal Deep Learning*

Seog-Min Lee
Hanshin University

Abstract

This paper explores the integration of artificial intelligence and causal inference in social science research, focusing on causal deep learning. We examine key theories including Pearl's Structural Causal Model, Rubin's Potential Outcomes Framework, and Schölkopf's Causal Representation Learning. Methodologies such as structural causal models with deep learning, counterfactual reasoning, and causal discovery algorithms are discussed.

The paper presents applications in social media analysis, economic policy, public health, and education, demonstrating how causal deep learning enables nuanced understanding of complex social phenomena. Key challenges addressed include model complexity, causal identification, interpretability, and ethical considerations like fairness and privacy.

Future research directions include developing new AI architectures, real-time causal inference, and multi-domain generalization. While limitations exist, causal deep learning shows significant potential for enhancing social science research and informing evidence-based policy-making, contributing to addressing complex social challenges globally.

Keywords

Causal Deep Learning, Social Sciences, Structural Causal Models (SCM), Counterfactual Reasoning, Policy Analysis

---

* This work was supported by Hanshin University Research Grant

** Associate Professor, Public Policy & Big Data Convergence Studies, Hanshin University. Email: newmind68@hs.ac.kr.

Analyses & Alternatives 2024 8(3): 125-150
DOI: 10.22931/aanda.2024.8.3.005

AI, 인과성, 사회과학의 통합: 인과 딥러닝을 통한 사회현상의 이해*

이석민**
한신대학교

요 약

이 연구는 사회과학 연구에서 인공지능과 인과추론의 통합, 특히 인과적 딥러닝에 초점을 맞추고, Pearl의 구조적 인과모델, Rubin의 잠재적 결과 프레임워크, Schölkopf의 인과적 표현 학습 등 주요 이론들을 검토하였다. 또한 딥러닝을 활용한 구조적 인과모델, 반사실적 추론, 인과 발견 알고리즘 등의 방법론을 논의하였다. 본 연구는 소셜 미디어 분석, 경제 정책, 공중 보건, 교육 분야에서의 응용 사례를 제시하며, 인과적 딥러닝이 복잡한 사회 현상에 대한 세밀한 이해를 가능케 함을 보여주고 있다. 또한 모델의 복잡성, 인과 식별, 해석 가능성, 그리고 프라이버시 같은 윤리적 고려사항 등 주요 과제들을 다루었다. 향후 연구 방향으로 새로운 AI 아키텍처 개발, 실시간 인과 추론, 다중 도메인 일반화 등을 제시하였다. 비록 한계점들이 존재하지만, 인과적 딥러닝은 사회과학 연구 강화와 증거기반 정책 수립에 상당한 잠재력을 보이며, 전 세계적인 복잡한 사회 문제 해결에 기여할 것으로 기대된다. 특히 본 연구는 빅데이터 환경에서의 인과관계 식별과 해석의 중요성을 강조하며, 전통적인 통계적 방법론과 최신 딥러닝 기술의 결합이 가져올 시너지 효과를 탐구하고 있다. 또한 이 분야의 발전이 사회과학 연구의 패러다임을 어떻게 변화시킬 수 있는지에 대한 논의를 제공함으로써, 향후 사회과학과 인공지능 기술의 융합 연구에 대한 방향성을 제시하고자 하였다.

주제어

인과 딥러닝, 사회과학, 구조적 인과 모델(SCM), 반사실적 추론, 정책 분석

* 이 논문은 한신대학교 학술연구비 지원에 의하여 연구되었음
** 부교수, 한신대 공공인재빅데이터융합학. Email: newmind68@hs.ac.kr.

Ⅰ. Introduction

The integration of artificial intelligence (AI) and causal inference methodologies has ushered in a new era in social science research. This convergence, particularly in the form of causal deep learning, offers powerful tools for understanding complex social phenomena and informing effective policy-making. As we navigate the complexities of modern society, the need for sophisticated analytical approaches that can capture intricate causal relationships has become increasingly apparent.

Traditional statistical methodologies, while valuable, often struggle to adequately explain complex nonlinear interactions and high-dimensional variables prevalent in social systems. For instance, precisely analyzing the impact of socioeconomic status on health conditions or the effects of policy changes on specific groups has been a persistent challenge. These limitations highlight the need for new methodologies to analyze causal structures in complex social contexts (Deaton & Cartwright, 2018).

The emergence of big data environments has further necessitated novel analytical approaches. In today's digital age, vast amounts of social behaviors and phenomena are recorded as digital data. Varian (2016) argued that such big data and AI technologies provide new opportunities for social science research. From opinions expressed on social media to census data and public policy information, large-scale datasets are accumulating at an unprecedented rate, necessitating advanced analytical techniques. The technology to effectively analyze these complex, often nonlinear and high-dimensional datasets is becoming increasingly crucial for extracting meaningful insights.

Causal deep learning, at the intersection of AI and causal inference, is gaining attention as a tool that can provide in-depth understanding of social phenomena and clearly analyze causal relationships. The theoretical foundation

of causal inference presented by Pearl and Mackenzie (2018) emphasizes the importance of this AI-based approach. By leveraging the pattern recognition capabilities of deep learning and the rigorous framework of causal inference, this approach offers the potential to uncover causal structures that may be obscured in traditional analyses.

Recent advancements in the field have demonstrated the potential of this integrated approach. For example, Ahne et al. (2022) proposed a methodology for analyzing causal relationships in tweets through social media data, playing a significant role in identifying social interactions that are easily overlooked by traditional statistical analysis. Their work showcases how causal deep learning can extract meaningful insights from the vast and often noisy landscape of social media data.

In the realm of text analysis, Bhopale and Tiwari (2024) proposed a technique for analyzing contextual representations in large-scale text data using Transformer-based deep learning models. This approach serves as a powerful tool for identifying potential causal relationships in social science data, demonstrating how advanced AI techniques can be leveraged to uncover latent causal structures in textual information.

Furthermore, the concept of causal representation learning, as proposed by Schölkopf et al. (2021), plays an important role in creating models that can be generalized across various environments. This approach aims to learn causal mechanisms directly from data, enabling the generation of models that can be applied across different contexts—a crucial capability in the diverse landscape of social sciences.

The potential applications of causal deep learning in social sciences are vast and varied. From analyzing the impact of social media on public opinion to evaluating the effects of economic policies, from predicting public health trends to assessing the long-term outcomes of educational interventions, this approach offers new avenues for research and policy-making. It promises to provide a

deeper causal explanation of complex social phenomena, potentially leading to more effective and targeted interventions.

However, the application of these advanced techniques is not without challenges. Issues of data quality, model interpretability, and ethical considerations come to the forefront as we apply these powerful tools to sensitive social data. Ensuring the fairness and accountability of AI models, protecting individual privacy, and maintaining transparency in decision-making processes are crucial considerations that must be addressed.

This paper aims to provide a comprehensive review of the integration of AI, causality, and social sciences, with a focus on causal deep learning. We will examine the theoretical foundations of this approach, explore its methodologies, and discuss its applications across various domains of social science. Furthermore, we will address the challenges and ethical considerations associated with this integration and look towards future research directions.

By bridging the gap between AI capabilities and causal inference in social sciences, we aim to contribute to a more nuanced understanding of social dynamics and support evidence-based policy-making. As we delve into this exciting field, we hope to illuminate both its immense potential and the responsible practices necessary for its successful application in understanding and shaping our complex social world.

II. Theoretical Foundations and Methodologies

2.1 Key Theories of Causal Inference

The foundation of causal deep learning in social sciences rests upon several key theories of causal inference. These theories provide the conceptual framework necessary for understanding and modeling causal relationships in

complex social systems.

a) Pearl's Structural Causal Model (SCM):

Judea Pearl's Structural Causal Model (SCM) has revolutionized our approach to causal inference by providing a mathematical framework for modeling causal relationships between variables (Pearl, 2009). At the heart of SCM are causal diagrams, particularly Directed Acyclic Graphs (DAGs), which visually represent causal relationships between variables. This graphical representation allows researchers to clearly articulate and analyze complex causal structures prevalent in social phenomena (Spirtes et al., 2000).

A key innovation of Pearl's SCM is the introduction of the do-operator, which enables the modeling and analysis of interventions on specific variables. This concept is crucial for social science research, where understanding the effects of interventions (such as policy changes) is often a primary goal. For instance, when estimating the effect of social policy changes on specific groups, SCM allows for clear identification of changes due to intervention, going beyond mere observational data (Morgan & Winship, 2015).

Furthermore, SCM facilitates counterfactual reasoning, providing a means to answer hypothetical questions like "What would have happened if...?" (Heckman, 2005). This capability is particularly valuable in social science research for evaluating the potential outcomes of alternative policy decisions or social interventions that were not actually implemented.

However, it is important to note that while SCM provides a powerful framework, it can be challenging to identify and model all relevant variables in complex social systems. This limitation necessitates careful consideration and often requires integration with other methodologies for comprehensive analysis.

b) Rubin's Potential Outcomes Framework:

Donald Rubin's Potential Outcomes Framework, also known as the Rubin

Causal Model, offers another fundamental approach to causal inference widely used in social sciences (Rubin, 1974). This framework is deeply rooted in the concept of randomized controlled trials (RCT) and focuses on estimating causal effects by comparing outcomes between treatment and control groups.

The core of Rubin's framework lies in its emphasis on counterfactual thinking. It conceptualizes causal effects as the difference between the outcome that would be observed under one condition (e.g., receiving a treatment) and the outcome that would be observed for the same unit under an alternative condition (e.g., not receiving the treatment). This approach is particularly useful in policy evaluation and medical research, where understanding the effects of specific interventions is crucial (Rosenbaum, 2002).

One of the strengths of Rubin's framework is its clear connection to experimental design, making it intuitive for researchers familiar with RCTs. It has been extensively used in fields such as economics, political science, and public health to estimate average treatment effects and understand policy impacts.

However, the framework has limitations, particularly in situations where experimental design is difficult or unethical to implement. Additionally, unlike Pearl's SCM, Rubin's framework does not provide a visual representation of causal relationships, which can limit its ability to explain complex causal interactions (Pearl, 2009).

c) Schölkopf's Causal Representation Learning:

Building upon and extending the work of Pearl and Rubin, Bernhard Schölkopf and colleagues have recently proposed a new paradigm called Causal Representation Learning (Schölkopf et al., 2021). This approach aims to bridge the gap between traditional causal inference and modern machine learning techniques, particularly deep learning.

The key insight of causal representation learning is that causal mechanisms

can be learned directly from data. This approach enables the generation of models that can be generalized across various environments, a crucial capability for social science research where contexts can vary significantly (Peters et al., 2017).

Unlike traditional causal inference methodologies that often rely on pre-specified causal structures, causal representation learning provides the possibility of automatically discovering causal structures within data through unsupervised learning (Bengio et al., 2019). This is particularly useful for inferring causal relationships in large-scale, complex data sets that are increasingly common in social science research.

Schölkopf's work also addresses the critical issue of distribution shift, which is prevalent in social science data. By learning stable causal structures, these models aim to perform reliably even when the distribution of data changes between training and application contexts. This robustness is essential for developing models that can inform policy decisions across diverse social settings.

The integration of these three theoretical approaches -- Pearl's SCM, Rubin's Potential Outcomes Framework, and Schölkopf's Causal Representation Learning -- provides a comprehensive foundation for causal inference in social sciences. Each approach offers unique strengths: SCM provides explicit causal models and enables intervention analysis, the potential outcomes framework enhances causal inference through its connection to experimental design, and causal representation learning offers tools for analyzing large-scale, unstructured data and learning causal structures directly.

However, it is crucial to recognize that each approach also has limitations. SCM may struggle with very complex systems, Rubin's framework can be limited in non-experimental settings, and causal representation learning is still in its early stages of development. Future research in causal deep learning for social sciences will likely focus on developing integrative approaches that

leverage the strengths of each of these methodologies while addressing their respective limitations.

2.2 Key Methodologies of Causal Deep Learning

Building upon the theoretical foundations discussed above, several key methodologies have emerged in the field of causal deep learning for social sciences. These methodologies aim to operationalize causal inference in the context of complex, high-dimensional data typical of modern social science research.

a) Structural Causal Models with Deep Learning:

The integration of Structural Causal Models (SCMs) with deep learning techniques represents a powerful approach for modeling complex causal relationships in social systems. This methodology combines the graphical representation and causal reasoning capabilities of SCMs with the pattern recognition and nonlinear modeling strengths of neural networks.

One significant advantage of this approach is its ability to capture and model complex nonlinear relationships that are often present in social data but are difficult to represent with traditional statistical methods. For instance, Bengio et al. (2019) proposed a learning method that separates causal mechanisms, introducing meta-transfer learning that can learn causal structures in multiple environments and apply them to new contexts. This approach is particularly valuable in social science research, where causal relationships may vary across different social or cultural contexts.

In practice, this methodology has been applied to various social science domains. For example, in educational policy analysis, it has been used to model complex causal relationships between students' socioeconomic backgrounds, school characteristics, and educational interventions, enabling more accurate

predictions of policy effects.

However, this approach also faces challenges. The complexity of these models can make them difficult to interpret, raising issues of transparency in decision-making processes. Additionally, capturing all relevant causal relationships in complex real-world social systems remains a significant challenge (Guo et al., 2020).

b) Counterfactual Reasoning in Deep Learning:

Counterfactual reasoning, a key concept in causal inference, has been successfully integrated into deep learning frameworks, offering powerful tools for policy analysis and social interventions. This methodology focuses on estimating outcomes in hypothetical scenarios, addressing questions like "What would have happened if we had implemented a different policy?"

The application of counterfactual reasoning in deep learning models is often achieved through counterfactual regression, which estimates counterfactual outcomes based on observed data (Johansson et al., 2016). This approach has shown particular promise in medical and economic policy analysis, where understanding the potential outcomes of different interventions is crucial.

For instance, Schuler et al. (2018) utilized counterfactual reasoning in medical policy decision-making to compare and analyze the effects of various treatment options. By estimating how a patient's condition would have differed under alternative treatments, this approach enables more informed and personalized medical decision-making.

However, counterfactual reasoning in deep learning is not without limitations. The accuracy of these models heavily depends on the quality and comprehensiveness of the available data. Additionally, considering all possible scenarios in complex social systems can be computationally intensive and sometimes practically infeasible.

c) Causal Discovery Algorithms:

Causal discovery algorithms represent another crucial methodology in causal deep learning for social sciences. These algorithms aim to learn causal structures directly from observational data, going beyond mere correlation to reveal potential causal pathways.

When combined with deep learning techniques, causal discovery algorithms become particularly powerful in analyzing complex, high-dimensional data. For example, Zhang et al. (2013) applied this approach to analyze the causal relationship between climate change and economic growth. By inferring causal relationships between variables such as greenhouse gas emissions, economic growth rates, and energy consumption, they were able to evaluate the potential economic impacts of climate policies.

The strength of this methodology lies in its ability to uncover previously unknown causal relationships in large datasets. This is particularly valuable in social science research, where the complexity of social systems often obscures important causal pathways.

However, researchers must exercise caution to ensure that the results derived through these algorithms do not reinforce existing social biases or lead to spurious causal inferences. Moreover, causal discovery algorithms face challenges, particularly in the presence of hidden confounders and when dealing with large sets of variables, which can lead to computational complexity issues (Spirtes & Zhang, 2016).

d) New AI Architectures for Causal Inference:

The development of novel AI architectures specifically designed for causal inference represents a frontier in causal deep learning research. These new structures aim to better capture and model causal relationships within neural network frameworks.

For instance, Bengio et al. (2019) proposed neural network structures designed

to learn independent causal mechanisms. These modularized architectures allow learned causal mechanisms to be effectively applied in new environments, showing particular strength in domain adaptation and transfer learning. This capability is especially valuable in social science research, where causal relationships may vary across different social or cultural contexts.

An example of the practical application of these new architectures is the work of Ke et al. (2021), who developed fair decision-making systems for recruitment processes. Their system aimed to select appropriate candidates while minimizing the impact of factors such as gender and race on outcomes, demonstrating how causal AI architectures can be used to address important social equity issues.

However, these new architectures also face challenges. They often require large amounts of data for effective learning, and their complexity can make interpretation of results difficult. This raises important ethical considerations, particularly in terms of ensuring transparency and accountability in decision-making processes based on these models.

In conclusion, these methodologies -- Structural Causal Models with Deep Learning, Counterfactual Reasoning in Deep Learning, Causal Discovery Algorithms, and New AI Architectures for Causal Inference -- represent the cutting edge of causal deep learning in social sciences. Each offers unique strengths in addressing the complexities of social science research, from modeling nonlinear causal relationships to uncovering hidden causal structures in large datasets.

However, it's important to note that the choice of methodology should be guided by the specific research question, data characteristics, and the balance needed between model complexity and interpretability. As the field continues to evolve, we can expect further refinement of these methodologies and the development of new approaches that address current limitations and challenges.

The integration of these causal deep learning methodologies with traditional social science research methods holds great promise for advancing our

understanding of complex social phenomena. By enabling more accurate modeling of causal relationships and more robust predictions of intervention effects, these approaches have the potential to significantly enhance evidence-based policy-making and social intervention strategies.

Ⅲ. Causal Deep Learning in Social Sciences

3.1 Field-Specific Application Cases

The integration of causal deep learning in social sciences has opened up new avenues for research and analysis across various domains. This section explores specific applications in four key areas: social media analysis, economic policy research, public health, and education policy.

a) Social Media Analysis:

Social media platforms have become invaluable sources of data for social scientists, offering real-time insights into public opinion, social trends, and human behavior patterns. Causal deep learning has significantly enhanced our ability to extract meaningful causal relationships from this vast and often noisy data.

A notable example is the work of Ahne et al. (2022), who used causal deep learning techniques to identify causal relationships in diabetes-related tweets. Their study employed transformer-based models like BERT or GPT to process large-scale text data, combined with structural causal models (SCMs) to infer causal relationships beyond simple correlations.

For instance, in analyzing tweets such as "The rise in insulin prices has increased the economic burden on diabetic patients," their model was able to identify 'rise in insulin prices' as the cause and 'increased economic burden' as

the effect. This kind of analysis goes beyond traditional sentiment analysis, offering insights into the causal chains that drive public discourse on health issues.

The implications of this approach extend far beyond health-related discussions. Similar methodologies can be applied to analyze public opinion on political issues, assess the impact of marketing campaigns, or study the spread of misinformation. For example, researchers could use these techniques to understand how specific policy announcements causally influence public sentiment, or how the spread of certain news stories impacts voting intentions.

However, this approach also faces challenges. Ensuring the representativeness of social media data remains a significant issue, as social media users may not reflect the general population. Additionally, platform-specific biases and the potential for manipulation (e.g., through bot accounts) need to be carefully considered in any analysis.

b) Economic Policy Research:

In the realm of economic policy, causal deep learning has enabled more nuanced and accurate analyses of policy effects. Traditional economic models often struggle to capture the complex, nonlinear relationships that exist in real-world economies. Causal deep learning approaches offer a way to model these complexities more effectively.

Athey and Imbens (2015) pioneered the use of machine learning techniques for estimating heterogeneous treatment effects in economic interventions. Their approach allows for a more detailed understanding of how policy effects may vary across different subgroups or contexts.

For example, in analyzing the effects of minimum wage policies, traditional regression analyses typically estimate only average effects. However, using causal deep learning methods, researchers can now examine how the impact of minimum wage increases differs based on factors such as age, education level,

industry sector, and local economic conditions. Such analyses have revealed that minimum wage increases could have negative effects on young, low-skilled workers but positive effects on middle-aged skilled workers.

This level of granularity in understanding policy effects is crucial for designing more effective and targeted economic interventions. Policymakers can use these insights to tailor policies to specific demographic groups or economic contexts, potentially improving overall policy outcomes.

Moreover, these methods excel at capturing nonlinear relationships in economic phenomena. For instance, the relationship between tax rates and government revenue (as described by the Laffer Curve) is inherently nonlinear. Causal deep learning models can more accurately capture and analyze such relationships, providing policymakers with better tools for tax policy design.

However, challenges remain. The complexity of these models can make their results difficult to interpret, which may be problematic when trying to communicate findings to policymakers or the public. Additionally, these methods often require large datasets to produce reliable results, which may not always be available, especially for newer or more specific policy interventions.

c) Public Health:

The field of public health has seen significant advancements through the application of causal deep learning, particularly in understanding complex health phenomena and evaluating the long-term effects of interventions.

Shi et al. (2019) developed a novel deep learning-based causal inference model for evaluating the long-term effects of drug treatments using electronic health record (EHR) data. Their model utilized recurrent neural networks (RNNs) to model changes in patient status over time and employed adversarial training techniques to minimize the influence of unobserved confounding factors.

This approach has led to important insights that were difficult to capture in

traditional short-term clinical trials. For example, their analysis of hypertension medications revealed that certain drugs, while effective at lowering blood pressure in the short term, may be associated with decreased kidney function in the long term. Such findings have significant implications for drug policies and prescription guidelines, potentially leading to more personalized and effective long-term treatment strategies.

Another crucial application of causal deep learning in public health is in predicting and managing disease outbreaks. Wang et al. (2022) proposed a model combining graph neural networks (GNNs) with causal inference techniques to predict COVID-19 spread patterns. Their model considered various causal factors such as inter-regional movement, policy interventions, and population characteristics to forecast disease spread and quantify the contribution of each factor.

This type of analysis is invaluable for public health officials in developing targeted intervention strategies and allocating resources effectively. For instance, it can help in deciding which types of movement restrictions would be most effective in curbing disease spread in specific regions, or in predicting the impact of vaccination campaigns on overall population health.

However, these approaches also face significant challenges. The sensitive nature of health data raises important privacy concerns, necessitating robust data protection measures. Additionally, the complexity of these models can make it difficult for medical professionals to interpret and trust the results, highlighting the need for explainable AI techniques in this domain.

d) Education Policy:

In the field of education, causal deep learning is enabling more accurate predictions of long-term policy effects and paving the way for personalized educational interventions.

A notable example is the study by Wang et al. (2024), who proposed a

method using long short-term memory (LSTM) networks to model changes in students' academic achievement over time. By integrating instrumental variable methods into their deep learning models, they were able to estimate causal effects of educational interventions more accurately.

Their analysis of early English education programs yielded nuanced insights. While these programs showed positive short-term effects on English skills, the study also revealed potential negative impacts on achievement in subjects such as mathematics or science in the long term. Such findings highlight the complex, often unforeseen consequences of educational interventions and underscore the importance of long-term, holistic assessment in education policy.

This approach opens up possibilities for more personalized education strategies. By understanding the causal factors that influence individual student performance over time, educators and policymakers can design targeted interventions that address the specific needs of different student groups.

However, challenges remain in this domain. Long-term tracking of educational data can be difficult, particularly in systems with high student mobility. Additionally, quantifying and evaluating various educational goals beyond academic achievement (such as social-emotional learning or creativity) remains a complex task that requires further methodological development.

3.2 Key Challenges and Ethical Considerations

While the applications of causal deep learning in social sciences show great promise, they also bring to the forefront several significant challenges and ethical considerations.

a) Complexity and Nonlinearity:

Social systems inherently involve complex, nonlinear interactions that challenge traditional modeling approaches and necessitate more sophisticated

analytical methods. While deep learning excels at capturing these patterns, the increased model complexity often comes at the cost of reduced interpretability. This creates a tension between model sophistication and ease of understanding, particularly crucial in policy-related decision-making processes.

For instance, in urban planning, factors such as population density, transportation infrastructure, economic activity, and environmental factors interact in complex, nonlinear ways. Modeling these interactions accurately while maintaining interpretability remains a significant challenge.

b) Causal Identification:

Distinguishing genuine causal relationships from mere correlations in big data environments is a persistent challenge. Despite advanced techniques like instrumental variables and propensity score matching, identifying causal relationships in complex social systems with potential hidden confounders remains difficult.

For example, in analyzing the relationship between education level and income, numerous confounding factors (such as family background, innate abilities, or local economic conditions) can obscure the true causal relationship. Causal deep learning models must grapple with these complexities to provide reliable insights.

c) Interpretability:

The "black box" nature of deep learning models poses significant challenges for interpretation, especially crucial in policy-related decisions. While techniques like SHAP (Lundberg and Lee, 2017) and LIME (Ribeiro et al., 2016) aim to enhance model interpretability, a trade-off between model performance and explainability often persists.

This challenge is particularly acute in sensitive domains like criminal justice or loan approvals, where the reasoning behind a model's decision needs to be

clear and justifiable.

d) External Validity:

Ensuring that causal deep learning models perform consistently across various social contexts is crucial. Techniques like domain adaptation (Ganin et al., 2016) and meta-learning (Finn et al., 2017) show promise, but generalizing models across significantly different cultural and institutional settings remains challenging.

For instance, a model trained on data from urban areas in developed countries may not generalize well to rural settings in developing nations, limiting its applicability in global policy-making.

e) Ethical Considerations:

The application of AI models to social data raises important ethical issues that need careful consideration:

Algorithmic Fairness: Ensuring that models do not perpetuate or exacerbate existing social inequalities is crucial. For example, in hiring processes or loan approvals, models must be carefully designed and monitored to avoid discriminating against protected groups.

Privacy Protection: The use of large-scale social data, especially in sensitive areas like health or education, raises significant privacy concerns. Techniques like differential privacy (Dwork, 2006) offer promising approaches, but balancing data utility with privacy protection remains a challenge.

Transparency and Accountability: Ensuring the transparency and accountability of AI models' decision-making processes, especially in policy applications, is essential for maintaining public trust and ethical implementation. This is particularly important when these models influence decisions that significantly impact individuals' lives.

Informed Consent: In many cases, individuals may not be aware that their data is being used to train AI models. Ensuring proper informed consent, especially when using data for purposes different from its original collection context, is an ongoing ethical challenge.

Potential for Misuse: The powerful predictive capabilities of these models could potentially be misused for manipulation or control. Safeguards against such misuse need to be an integral part of the development and deployment process.

The strengthening of AI ethics regulations, such as the EU's GDPR and the US's ADPPA, underscores the growing importance of addressing these ethical concerns in research and application. As causal deep learning continues to influence decision-making in critical social domains, adherence to ethical guidelines and ongoing ethical review processes will be crucial.

In conclusion, while causal deep learning offers powerful tools for understanding and influencing social systems, it also brings significant challenges and ethical considerations. Addressing these issues requires ongoing collaboration between data scientists, domain experts, ethicists, and policymakers. As we continue to develop and apply these technologies, maintaining a balance between innovation and responsible use will be key to realizing their full potential for social good.

Ⅳ. Conclusion

4.1 Key Research Directions

The evolution of causal deep learning in social sciences points towards several promising research directions:

a) Interdisciplinary Collaboration: Fostering stronger collaborations between AI

researchers, statisticians, and domain experts in social sciences is crucial. As Schölkopf et al. (2021) emphasized, such collaborations can lead to more robust and context-aware models. For instance, Knaus (2020) demonstrated the power of combining economics and machine learning in analyzing labor market policies.

b) New AI Architectures: Developing novel AI architectures specifically for causal inference remains a frontier. Bengio et al. (2019) proposed models for learning causal mechanisms across multiple environments, while Alfakih et al. (2023) developed a 'Causal Variational Autoencoder' for medical applications. Future architectures should better handle unobserved confounders and provide more interpretable representations of causal structures.

c) Integration with Policy Analysis: Bridging the gap between causal inference techniques and practical policy-making is essential. While methods like Double Machine Learning (Chernozhukov et al., 2018) have advanced unbiased policy effect estimation, more accessible tools for policymakers are needed.

d) Real-time Causal Inference: As social dynamics become increasingly fast-paced, the ability to perform real-time causal inference becomes crucial. This involves developing streaming algorithms for continuous updating of causal models and methods for quick adaptation to sudden shifts in social dynamics.

e) Multi-domain Generalization: Developing causal models that can be reliably applied across diverse social contexts remains a challenge. Chen et al. (2023) proposed a 'meta causal learning' framework, while Rojas-Carulla et al. (2018) worked on invariant causal prediction methods. This research could lead to more globally applicable models.

4.2 Limitations and Future Outlook

Several limitations need addressing for the continued development of causal deep learning in social sciences:

a) Data Quality and Sample Bias: Ensuring data representativeness remains a significant challenge, particularly in social media-based studies.

b) Long-term Impact Assessment: Accurately predicting the long-term effects of social policies requires advanced temporal modeling techniques and longitudinal studies.

c) Dynamic Causal Relationships: Developing models that can capture and interpret changing causal structures over time is crucial.

d) Ethical Implementation: As these technologies become more widely used, ensuring their ethical implementation becomes increasingly critical, addressing issues of fairness, transparency, and privacy.

Looking ahead, we can anticipate more personalized policy interventions, enhanced ability to predict and mitigate social issues, improved decision-making tools for policymakers, and more effective strategies for addressing global challenges. However, realizing this potential will require ongoing efforts to address ethical, methodological, and practical challenges.

In conclusion, causal deep learning represents a powerful new frontier in social science research and policy-making. It offers the potential to significantly enhance our understanding of social dynamics and our ability to address complex social challenges. As we advance, maintaining a balance between technological innovation and ethical responsibility will be crucial to harnessing the full potential of these approaches for societal betterment. The future of this field is not merely about technological advancement, but about fostering a deeper, more nuanced understanding of the complex tapestry of human society, ultimately contributing to the improvement of human well-being on a global scale.

References

Ahne, A., Khetan, V., Tannier, X., Rizvi, M. I. H., Czernichow, T., Orchard, F., ... & Fagherazzi, G. (2021). Identifying causal relations in tweets using deep learning: Use case on diabetes-related tweets from 2017-2021. *arXiv preprint, arXiv*:2111.01225.

Alfakih, A., Xia, Z., Ali, B., Mamoon, S., & Lu, J. (2023). Deep causality variational autoencoder network for identifying the potential biomarkers of brain disorders. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. https://doi.org/10.1109/TNSRE.2023.3344995

Athey, Susan, & Imbens, Guido W. (2015). Machine learning methods for estimating heterogeneous causal effects. *arXiv preprint, arXiv*:1607.06580.

Bengio, Yoshua, Deleu, Tristan, Rahaman, Nasim, Ke, Ruijiang, Lachapelle, Samuel, Bilaniuk, Olexa, ... & Pal, Chris. (2019). A meta-transfer objective for learning to disentangle causal mechanisms. *arXiv preprint, arXiv*:1901.10912.

Bhopale, A. P., & Tiwari, A. (2024). Transformer-based contextual text representation framework for intelligent information retrieval. *Expert Systems with Applications*, 238, 121629.

Chen, Hanwen, Wu, Xia, & Taylor, Lucas. (2023). Meta-causal learning: Transferring causal knowledge across domains. *Journal of Machine Learning Research*, 24(115).

Chernozhukov, Victor, Chetverikov, Denis, Demirer, Mehdi, Duflo, Esther, Hansen, Christian, Newey, Whitney, & Robins, Jamie. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal,* 21(1), C1-C68.

Deaton, Angus, & Cartwright, Nancy. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine,* 210.

Dwork, Cynthia. (2006). Differential privacy. In *Proceedings of the 33rd international conference on Automata, Languages and Programming.*

Finn, Chelsea, Abbeel, Pieter, & Levine, Sergey. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*. PMLR.

Ganin, Yaroslav, Ustinova, Evgeniya, Ajakan, Hana, Germain, Pascal, Larochelle, Hugo, Laviolette, François, ... & Lempitsky, Victor. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1).

Guo, Ruocheng, Cheng, Liang, Li, Jun, Hahn, Peter R., & Liu, Huan. (2020). A survey of learning causality with data: Problems and methods. *ACM Computing Surveys*, 53(4).

Heckman, James J. (2005). The scientific model of causality. *Sociological Methodology,* 35(1).

Johansson, Fredrik, Shalit, Uri, & Sontag, David. (2016). Learning representations for counterfactual inference. In *International Conference on Machine Learning*. PMLR.

Knaus, Michael. (2020). Double machine learning based program evaluation under unconfoundedness. *Econometrics: Mathematical Methods & Programming eJournal.* https://doi.org/10.1093/ectj/utac015

Lundberg, Scott M., & Lee, Su-In. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems.*

Morgan, Stephen L., & Winship, Christopher. (2015). Counterfactuals and causal inference: *Methods and principles for social research* (2nd ed.). Cambridge University Press.

Pearl, Judea. (2009). Causal inference in statistics: An overview. *Statistics Surveys,* 3.

Pearl, Judea, & Mackenzie, Dana. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Peters, Jonas, Janzing, Dominik, & Schölkopf, Bernhard. (2017). *Elements of causal inference: Foundations and learning algorithms*. MIT Press.

Ribeiro, Marco T., Singh, Sameer, & Guestrin, Carlos. (2016). "Why should I

trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Rojas-Carulla, Marta, Schölkopf, Bernhard, Turner, Richard, & Peters, Jonas. (2018). Invariant models for causal transfer learning. *The Journal of Machine Learning Research,* 19(1).

Rosenbaum, Paul R. (2002). *Observational studies* (2nd ed.). Springer.

Rubin, Donald B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology,* 66(5).

Schölkopf, Bernhard, Locatello, Francesco, Bauer, Stefan, Ke, Nan R., Kalchbrenner, Nal, Goyal, Anirudh, & Bengio, Yoshua. (2021). Toward causal representation learning. *Proceedings of the IEEE,* 109(5).

Schuler, Andreea, Baiocchi, Michael, Tibshirani, Robert, & Shah, Niladri. (2018). A comparison of methods for model selection when estimating individual treatment effects. *arXiv preprint, arXiv:*1804.05146.

Shi, Chansoo, Blei, David, & Veitch, Victor. (2019). Adapting neural networks for the estimation of treatment effects. In *Advances in Neural Information Processing Systems.*

Spirtes, Peter, Glymour, Clark, & Scheines, Richard. (2001). *Causation, prediction, and search*. MIT Press.

Spirtes, Peter, & Zhang, Kun. (2016). Causal discovery and inference: Concepts and recent methodological advances. *Applied Informatics*, 3(1).

Varian, Hal R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences,* 113(27).

Wang, Cheng, Chen, Jing, Xie, Zitao, & Zou, Ji. (2024, July). Research on education big data for student's academic performance analysis based on machine learning. In *Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Education Digitalization and Computer Science.*

Wang, Liang, Adiga, Aniruddha, Chen, Jing, Sadilek, Adam, Venkatramanan,

Srinivasan, & Marathe, Madhav. (2022). CausalGNN: Causal-based graph neural networks for spatio-temporal epidemic forecasting. *AAAI,* 12191-12199. https://doi.org/10.1609/aaai.v36i11.21479

Zhang, Kun, Schölkopf, Bernhard, Muandet, Krikamol, & Wang, Zhifeng. (2013). Domain adaptation under target and conditional shift. In *International Conference on Machine Learning*. PMLR.

Manuscript: Sept 20, 2024; Review completed: Oct 03, 2024; Accepted: Oct 12, 2024