Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2026.070103
ISSN 2790-0932 Vol. 7, Issue 1: 18-27

Research on the Application of Causal Reasoning Method in Content Compliance Experimental Evaluation

Huisheng Liu

Operations Research, Columbia University, New York, 10027, USA

Keywords: Causal reasoning, Generator validator framework, Multi-step reasoning, Cross modal causal reasoning, Causal chain prompts.

Abstract: This study improves the reliability of content conformity evaluation through causal reasoning. To address the limitations of large language models in reasoning reliability, a generator-validator collaborative framework is constructed. It integrates causal interpretation and reasoning tasks through a "generate-validate-correct" closed-loop mechanism, tackling challenges of high-dimensional sparsity and unstructured data. The traditional chain-of-thought method lacks process supervision, leading to logical discontinuity and error accumulation. Research has achieved breakthroughs through three-stage innovation: in the mathematical reasoning scenario, the generator constructs a structured reasoning path and maps it to mathematical expressions, the validator provides fine-grained feedback, and the model accuracy and reliability have been significantly improved through six datasets validation and ablation experiments; The dimensions of causal reasoning are unified for causal inference and explanation generation. The generator extracts text causal relationships, and the validator generates natural language explanations to optimize the inference chain. Multi dimensional evaluation indicators and new quality scores are used to verify effectiveness, and the ablation experiment clarifies the impact of feedback forms; Under the requirement of autonomous causal reasoning, a causal chain prompt framework is designed to transform the five step process of variable identification, relationship extraction, adjacency matrix initialization, independence evaluation, and hypothesis generation into executable prompt engineering. The intermediate process explicit output is achieved by integrating do calculus and d-separation principle, and breaking through data dependence by combining few sample learning. Experiments confirm the framework enhances reasoning accuracy, reduces long-term dependency errors, and improves interpretability. Future work will explore multimodal fusion and self-validation to advance reliable, interpretable cross-modal causal inference.

1. Introduction

This study focuses on the application of causal reasoning methods in evaluating content conformity experiments. The background is rooted in the essence of reasoning as a cognitive activity - forming conclusions through evidence, arguments, and logic. Although large models have shown potential applications in mathematics, common sense, symbols, logic, and multimodal reasoning, causal reasoning faces core challenges due to the high-dimensional sparsity [1], unstructured nature, and complex context of natural language data. LLM self supervised learning [2] easily confuses causal and statistical correlations, and lags behind human cognitive levels in performance; The traditional thinking chain method relies on result feedback and lacks process supervision, leading to logical discontinuity and error accumulation; Multi step reasoning suffers from long-range dependency defects, lack of self validation, and low domain adaptation efficiency; The transition from correlation to causality in causal reasoning is difficult, and data dependence and probabilistic output lead to explanatory loss. The existing literature challenges include the lack of feedback mechanisms, which makes it difficult to verify the inference path, multi-step inference bottlenecks that limit complex task processing, and causal inference ability evaluation indicators limited to text similarity. The motivation of this article is to improve the logical rigor, accuracy, and interpretability of the model through collaborative multi-step reasoning frameworks and reasoning paradigms. The specific goal is to build a generator validator collaborative framework to implement process based feedback optimization, integrate causal relationships and causal prompt chains to enhance interpretability, and break through the performance bottleneck of LLM in complex reasoning tasks. The contribution is mainly reflected in three aspects: firstly, proposing a generator validator collaborative multi-step reasoning framework, which addresses mathematical reasoning chain errors and achieves continuous optimization through stepwise reasoning paradigms and mathematical expression mapping; Secondly, deeply coupling causal reasoning with explanation generation tasks and introducing new causal explanation quality evaluation indicators to overcome the limitations of traditional indicators; Thirdly, propose an autonomous causal reasoning method based on causal chain prompts, integrate Pearl causality diagram theory [3] to design a five step reasoning process, and combine few sample learning to solve the problem of scarce training data.

2. Correlation theory

2.1 Reasoning task classification and prompt engineering method

As the core mechanism of cognitive processes, reasoning can be divided into six categories based on logical structure and conclusion certainty: deduction, induction, causation, analogy, causality, and probabilistic reasoning. Deductive reasoning [5] is based on the inevitability of formal logic, and the conclusion strictly follows the semantic implication relationship of the premise, which is typically applied in mathematical proofs; Inductive reasoning derives general conclusions from samples, and its credibility depends on the representativeness of the samples. It is commonly used in summarizing natural science laws; Causal reasoning selects hypotheses by constructing the best explanatory model, following the Occam's Razor principle[6]; Analogical reasoning plays a crucial role in interdisciplinary research by achieving knowledge transfer through cross domain structural similarity; Causal reasoning focuses on identifying causal relationships between events, involving the removal of confounding factors and the evaluation of causal effects, and is the foundation of scientific decision-making; Probabilistic reasoning is based on Bayesian framework to quantify uncertainty and is widely used in risk analysis and machine learning. These reasoning paradigms have significant differences in conclusion strength, applicable scenarios, and cognitive costs, and adaptation strategies need to be selected based on task requirements. In the field of prompt engineering, the breakthrough in the ability of big language models is due to innovative methods such as instruction learning, context learning, and thought chain. Instruction learning builds an input-output template ternary framework through natural language instructions, such as FLAN which improves zero sample generalization ability through multitasking instruction fine-tuning; Context learning achieves task adaptation without parameter updates through task descriptions and sample prompt templates, reflecting the meta learning potential of language models; The thought chain enhances the performance of complex tasks by introducing step-by-step deduction logic, and its topological structure has evolved from a single chain to tree, schema, and hybrid modal architectures, significantly improving interpretability in arithmetic, common sense, and symbolic reasoning. These prompt methods collectively push the performance boundaries of large language models in inference tasks, providing key technical support for the logical foundation of intelligent systems.

2.2 The Technological Evolution and Representative Architecture of Large Language Models

The technological evolution of big language models, represented by the GPT series, ChatGLM series, and LLaMa series, demonstrates a dual breakthrough in parameter scale expansion and architecture innovation. Since the first generation model GPT-1 (117 million parameters) in 2018, the GPT series has adopted a two-stage training strategy (unsupervised pre training+supervised fine-tuning) and gradually developed to GPT-4, with parameter scales reaching billions, achieving joint understanding and generation of graphics and text. Through reinforcement learning and human feedback mechanisms, the accuracy of content is improved, approaching human level in tasks such as complex mathematical problem solving and cross language understanding. The ChatGLM series was proposed by Tsinghua University, based on the Transformer architecture and using a joint training strategy of dynamic masking and sequence rearrangement. Its pre training objectives include autoregressive blank filling (as shown in equation (2-1), Zm is the permutation set of all possible index sequences of length m, szi is the first i label sequences, and the model predicts masked tokens through autoregression) and multi task pre training (document level and sentence level objectives). The GLM-130B model is pre trained on 400 billion token corpora and outperforms GPT-3 in MMLU benchmark few sample learning, with LAMBADA accuracy reaching 80.2%. The LLaMa series was developed by Meta. The first generation of LLaMa is based on an improved Transformer architecture, introducing pre normalization strategy and rotation position encoding. The 65B parameter version is comparable to GPT-3 in language understanding benchmark; LLaMa2 expands the context window to 4096 tokens, completes 2 trillion token training in combination with course learning, and integrates reinforcement learning classifiers to reduce harmful output rates to below 0.01%; LLaMa3 improves inference efficiency by 30% through group query attention technology and constructs a hybrid system of 8 expert sub networks. After multi-stage instruction fine-tuning, the accuracy of GSM8K mathematical inference benchmark reaches 85.3%, which is 14.2 percentage points higher than the previous generation. These models promote the evolution of natural language processing technology towards specialization and security through architectural innovation and open source ecosystem construction.

3. Research method

3.1 Theoretical Framework and Hierarchical Architecture of Causal Reasoning

The theoretical system of causal reasoning is based on two pillars: the latent outcome model and the structural causal model. The latent outcome model was proposed by Neyman and Rubin, which quantifies causal effects through counterfactual results. Its mathematical expression is

$$\tau_i = Y_i(1) - Y_i(0) \tag{1}$$

among them, Yi (1) and Yi (0) respectively represent the potential outcomes of individual i when receiving intervention (T=1) and when not receiving intervention (T=0), and the theoretical boundaries of observational studies are constructed based on the three hypotheses of "stability, non-interference, and consistency". The structural causal model, based on Pearl's research results, describes the data generation process (DGP) through three core modules: causal diagram, structural equation system, and counterfactual reasoning[7]. Its standard causal diagram includes intervention variables (T), confounding covariates (Z), and response variables (Y), forming a systematic mathematical framework for causal relationships between variables. Causal inference can be divided into three levels: correlation level focuses on variable correlation pattern mining, laying the foundation for tasks such as image recognition and natural language processing; The hierarchy of cause effect tracing evaluates the effectiveness by applying specific interference, such as observing changes in user click through rates after adjusting algorithms in intelligent recommendation systems; The causal hierarchy (counterfactual inference) explores the action strategies required to achieve specific results, such as intelligent customer service systems optimizing dialogue strategies through counterfactual analysis to enhance customer satisfaction. These theoretical frameworks and hierarchical structures together form the scientific foundation of causal reasoning, supporting complex system analysis and decision support applications.

3.2 Research on Mathematical Reasoning Optimization of Generator Validator Collaborative Closed Loop Framework

This article proposes a collaborative multi-step reasoning framework (CRMR) based on generator validator and a stepwise reasoning paradigm (SSR) [8], which optimizes the mathematical reasoning path through a "generation validation feedback" closed-loop mechanism. The CRMR framework consists of three main modules: the generator generates a preliminary thought process with logical steps, the validator gradually verifies and provides feedback on calculation errors, and the generator iteratively optimizes solutions based on feedback. The SSR paradigm decomposes unordered reasoning into ordered steps, such as mapping "4 boxes x 2 strips/box=8 scarves" to mathematical expressions. Generator training uses mean square error loss function

$$L_G = \frac{1}{N}\sum_{i=1}^{N}(a_i - gt_i)^2 \tag{2}$$

a_i is the predicted answer, gt_i is the true label); The validator constructs a dataset through sampling inference paths, using

$$L_V = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(f_{ij} - feedback_{ij}\right)^2 \tag{3}$$

Optimize feedback accuracy. The experiment was validated on datasets such as SVAMP and SingleEq, providing feedback on text integration issues, initial thinking processes, correction prompts, and instruction templates to achieve precise error localization and logical correction, significantly improving the accuracy and reliability of mathematical problem solving.

3.3 Experimental analysis of mathematical problem solving based on collaborative multi-step reasoning framework

In mathematical problem-solving tasks, the collaborative multi-step reasoning framework significantly improves the inference performance of the model through the interactive feedback mechanism between the generator and validator. Taking the GSM8K dataset as an example, the response comparison of ChatGLM2 model under three methods is demonstrated: when the generator independently answers, the inference step is incorrect due to the lack of verification module; After introducing the thought chain, the model can gradually deduce the correct answer; After combining with the validator, the generator can correct erroneous reasoning based on feedback and ultimately obtain accurate results. Table 1 further validates the universality of this method on multiple models - in open source models such as ChatGLM2, LLaMa-2-7B/13b, GPT 2-

Large, and DeepSeeker V3, the collaborative inference method achieved performance improvements of 5.84% -16.20% compared to the baseline, with GPT 2-Large achieving an accuracy improvement of 16.20% on the GSM8K dataset.(As shown in Table 1)

Table 1 Open source big model mathematical reasoning performance evaluation focuses on three major indicators

Model	Parameter Size	Method	AddSub	MultiArith	SingleEq	SingleOp	SVAMP	GSM8K
GPT2-Large	774M	Standard Prompt	40.00	83.17	44.23	52.48	52.20	66.85
GPT2-Large	774M	Self-Consistency	43.29	85.26	47.89	52.84	56.50	69.93
GPT2-Large	774M	CoT	47.34	90.72	52.94	63.23	64.40	73.63
GPT2-Large	774M	Generator+Verifier	56.20	89.00	56.82	77.56	67.70	78.97
ChatGLM2	6B	Standard Prompt	79.00	91.78	86.42	91.27	85.20	75.00
ChatGLM2	6B	Self-Consistency	80.00	93.42	87.34	91.83	85.50	75.36
ChatGLM2	6B	CoT	82.52	95.75	88.69	92.31	86.30	77.64
ChatGLM2	6B	Generator+Verifier	84.84	99.66	90.33	95.72	89.80	82.52
LLaMa-2	7B	Standard Prompt	80.69	90.16	88.28	92.34	87.50	78.61
LLaMa-2	7B	Self-Consistency	82.73	92.21	91.41	93.69	89.90	80.32
LLaMa-2	7B	CoT	83.60	95.23	92.35	94.73	91.20	82.57
LLaMa-2	7B	Generator+Verifier	86.04	98.34	94.39	97.27	93.90	85.05
LLaMa-2	13B	Standard Prompt	84.24	94.68	91.35	93.59	89.30	79.13
LLaMa-2	13B	Self-Consistency	86.12	95.43	92.47	95.03	91.20	81.49
LLaMa-2	13B	CoT	88.93	95.69	93.71	96.47	93.30	83.05
LLaMa-2	13B	Generator+Verifier	93.03	98.58	95.38	98.37	95.10	86.70
DeepSeek-V3	671B	Standard Prompt	86.08	94.17	85.32	88.61	85.80	84.23
DeepSeek-V3	671B	Self-Consistency	89.32	95.64	86.17	91.35	87.40	86.24
DeepSeek-V3	671B	CoT	92.18	96.08	89.02	92.75	88.90	87.97
DeepSeek-V3	671B	Generator+Verifier	95.76	98.25	91.36	94.13	92.10	90.58

The closed source model experiment showed that Gemini Pro and GPT-4 improved their accuracy by 11.55% and 11.71% respectively after applying the collaborative inference framework, indicating that even with strong baseline models, inference performance can still be optimized through feedback mechanisms. The multi-path sampling experiment of ChatGLM2 on six datasets shows that the collaborative inference method exhibits more stable performance gains with increasing sampling paths, especially on complex datasets such as MultiArith and AddSub. The impact of inference step length on performance is revealed by revealing that when the inference step size is 1-2, the model performs best on the SingleOp and AddSub datasets; After increasing the step size to 4, the accumulation of errors leads to a significant decrease in accuracy. Using the example of GPT2 Large with a stride of 3 on the AddSub dataset, it is demonstrated that an excessively long inference path may lead to semantic comprehension bias. The ablation experiment verified the effectiveness of the stepwise reasoning paradigm (SSR) - on five datasets, the SSR format had an accuracy rate 0.8% -6.5% higher than the original reasoning format, and its structured steps helped to accurately locate errors. The comparison between self reflection and collaborative reasoning shows that the performance improvement of the generator under the feedback of the validator reaches 6.3% -24.8%, which verifies the key role of the validator in compensating for the self reflection deficiency of the generator. The collaborative multi-step reasoning framework effectively improves the logical rigor and result accuracy of large language models in mathematical reasoning tasks through generator validator interaction, stepwise reasoning paradigm, and multi round feedback mechanism, providing interpretable optimization paths for complex reasoning tasks.

4. Results and discussion

4.1 Collaborative multi-step reasoning framework for causal reasoning in natural language

The digital health platform follows the core idea of "countless carriers", which means that the platform does not own the data ownership, but only has the right to use and operate it under authorization, helping data owners achieve value generation and realization. In terms of health management data storage, blockchain distributed storage supports synchronization and verification of content change validity among nodes, achieves access control and privacy protection through asymmetric encryption, and can build different architectures such as public chains, private chains, and consortium chains according to individual and group health service models, adapting to different privacy level requirements - public chains support arbitrary user queries and transaction confirmations, suitable for personal data service scenarios; Private chains restrict write permissions to specific institutions, while read permissions can be flexibly restricted, making them suitable for medical knowledge accumulation and cost control in medical or research institutions. To incentivize users to share health data, a consensus mechanism based on Proof of Health Data Asset Value (PoDAV) is designed: the value of health data assets is measured by reliability (reflecting data quality and demand matching) and liquidity (reflecting transaction frequency and scale). The PageRank algorithm [8] is used to rank users and select the top N as validators. After paying a deposit, a pseudo-random number is used to select new block initiators; The consensus process consists of two rounds of voting, both using Byzantine fault-tolerant methods - in the first round of preparation stage, validators cast preparation votes, and if more than 2/3 of the valid votes are cast, the process enters the second round of confirmation stage. Each round of voting validators can receive a reward of 1.5x tokens; Successfully adding a compliant new block will reward 1x tokens, while voting failure will result in a loss of 0.5x tokens. This mechanism promotes the circulation and value appreciation of health data assets through token incentives and punishments, and drives the transformation of health management from traditional models to data ownership, assetization, and personalized services.

4.2 Model experiment

Causal reasoning prioritizes inferring potential mechanisms rather than simple associations. The collaborative multi-step reasoning framework integrates causal reasoning and interpretation generation through three steps: the generator performs initial causal reasoning and outputs preliminary results; Validator generates auxiliary explanations to clarify causal relationships; The generator corrects the answer based on feedback from the validator.The generator is trained using cross entropy loss on the e-CARE dataset [9]. For samples containing premises and two hypotheses, the model predicts the probabilities P_1 and $P_2$of hypothesis 1 and hypothesis 2, with a loss function of:

$$L = -\frac{1}{N}\sum_{I=1}^{N}[Y_i \log P_1(X_i) + (1 - Y_i) \log P_2(X_i)] \tag{4}$$

among them,$Y_i \in \{0,1\}$ represents the correct hypothesis (0 is hypothesis 1, 1 is hypothesis 2).

As a sequence to sequence task, the validator optimizes the interpretation through cross entropy loss to generate:

$$L = -\sum_{t=1}^{T} \log p(Y_t | y_{<t}, X) \tag{5}$$

among them, Yt is the t-th morpheme of the explanation.

The experiment used the COPA (1000 daily causal inference multiple-choice questions) and e-CARE (over 21000 causal inference questions with conceptual explanations) datasets. The task format includes multiple-choice questions (selecting reasons/results) and binary classification (determining causal relationships).The BLEU metric quantifies the similarity between generated text and reference text through n-gram accuracy and short penalty (BP), and its calculation formula is

$$\mathrm{BLEU} = \exp\left(\sum_{n=1}^{N} \omega_n \log P_n\right) \cdot \mathrm{BP} \tag{6}$$

among them, P_n is the accuracy of n-gram matching, and $\omega_n$is the weight coefficient. ROUGE-N focuses on n-gram recall rate, defined as

$$\mathrm{ROUGE} - \mathrm{N} = \frac{\sum_{\mathrm{gram}_n \epsilon S} \mathrm{Count}_{\mathrm{match}}(\mathrm{gram}_n)}{\sum_{\mathrm{gram}_n \epsilon S} \mathrm{Count}(\mathrm{gram}_n)} \tag{7}$$

Directly measure the degree of overlap between the generated text and the reference text. Causal Explanation Quality (CEQ) innovatively introduces incremental evaluation of causal strength, through

$$\Delta_{cs} = cs(C, E|X) - cs(C, E) \tag{8}$$

The enhancement effect of quantitative explanation on causal relationship, where csC and EX are the causal strength of cause C and result E given explanation X, calculated based on corpus word level association and introduced with a penalty coefficient α=0.66 to correct the association strength. The three constructed a multi-level model performance evaluation system from three dimensions: text similarity, semantic overlap, and causal mechanism analysis, jointly supporting the quantitative verification of logical coherence and explanatory credibility in causal reasoning tasks.

4.3 Effect analysis

In causal inference tasks, testing of 11 mainstream models (such as LLaMa2, GPT series, DeepSeek-V3, etc.) showed that generative models (such as LLaMa2-13B, DeepSeek-V3) performed better than discriminative models (such as BERT, RoBERTa) on COPA and e-CARE datasets due to autoregressive mechanisms, with DeepSeek-V3 achieving the best results in all four evaluation dimensions (as shown in Table 2)

Verified the positive correlation between model size and causal reasoning ability. In the explanation generation experiment, GPT-2 outperformed RNN and BERT based based based on BLEU-1 (53.19), ROUGE-1 (26.85), and causal interpretation quality (0.102) indicators, demonstrating its advantages in short word matching and long text similarity between generated text and reference interpretation. The collaborative reasoning experiment improves performance through the "generator+validator" framework: compared with the baseline (generator reasoning alone) and the chain of thought (CoT) method, this framework achieves the highest accuracy in multiple choice/binary classification tasks of COPA and e-CARE, proving that the auxiliary interpretation provided by the validator can enhance the generator's analysis of complex causal relationships (as shown in Figure 1).

Table 2 Model Performance Comparison

Model	COPA	e-CARE	Choice	Binary Choice
Bart-base	57.80	63.07	71.47	64.25
Bert-base-cased	61.20	50.36	73.25	65.96
RoBERTa-base	68.20	64.20	70.78	59.34
XLNet-base-cased	64.50	72.59	75.92	69.63
ALBERT	54.30	65.28	72.16	66.71
GPT	53.20	58.41	63.36	62.26
GPT-2	69.00	71.53	68.73	72.91
ChatGLM2-6B	74.60	69.38	78.75	79.02
LLaMa2-7B	76.30	68.51	80.34	75.06
LLaMa2-13B	81.90	78.29	82.62	78.64
DeepSeek-V3*	90.20	89.52	87.35	85.93

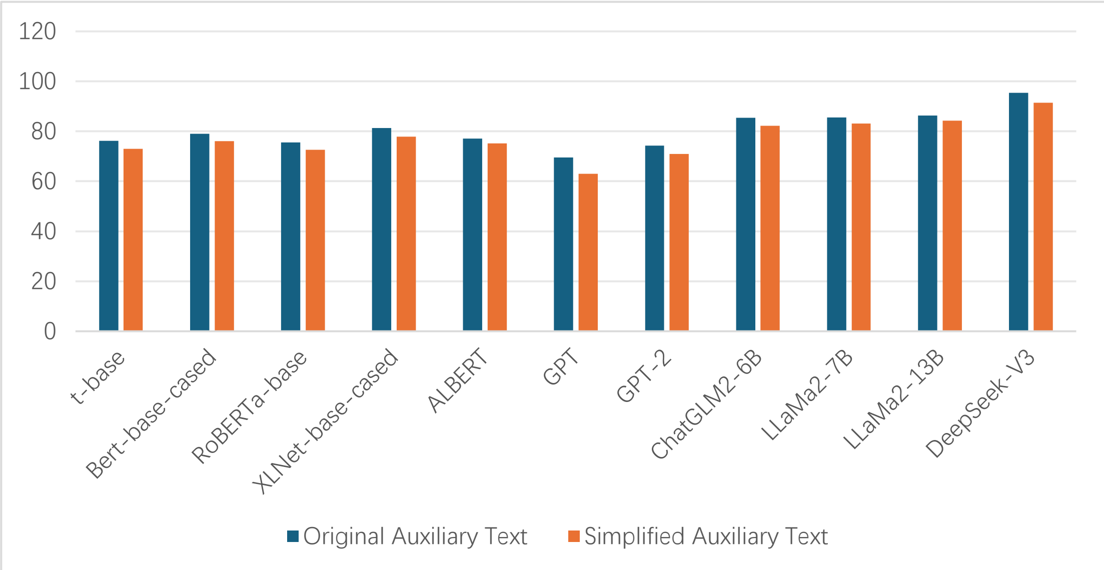

Figure 1 Model Performance Comparison under Different Auxiliary Text Conditions

The ablation experiment further validated the importance of high-quality explanations: simplifying the explanatory text generated by the validator significantly reduced the accuracy of model collaborative reasoning (such as DeepSeeker V3 dropping from 95.42% to 91.37%); Example comparison of original/simplified explanations); However, random feedback (non causal interpretation) leads to a significant decrease in the inference performance of the generator (such as DeepSeeker V3 dropping from 95.24% to 88.49%); Showing examples of normal/random feedback. These results collectively support the superiority of the "generator validator" collaborative framework in integrating causal reasoning and interpretation generation tasks, and promote the development of causal reasoning applications in practical scenarios such as intervention analysis and hybrid control.

5. Conclusion

This study focuses on the two core challenges of complex reasoning in large language models - insufficient reasoning reliability and causal reasoning ability deficiencies. An innovative generator validator collaborative framework is constructed to form a "generation validation correction" closed-loop mechanism, achieving deep integration of causal interpretation and reasoning tasks.

Specifically, a three-stage breakthrough is achieved: in the mathematical reasoning scenario, the generator constructs a structured multi-step reasoning path and maps it to an explicit mathematical expression, and the validator corrects the path through fine-grained feedback. Six datasets and ablation experiments have shown that this design significantly improves the accuracy and reliability of the model; In the dimension of causal inference, the task of unifying causal inference and explanation generation is carried out. The generator extracts causal relationships from the text, and the validator generates natural language explanations to optimize the inference chain. Multi dimensional evaluation indicators and a new type of causal explanation quality score are used to verify the effectiveness, and the impact of different feedback forms on performance is clarified through ablation experiments; To meet the demand for autonomous causal reasoning, a causal chain prompt framework is designed. Through a five step process of variable identification, relationship extraction, adjacency matrix initialization, independence evaluation, and hypothesis generation, structured causal reasoning is transformed into executable prompt engineering. The do algorithm and d-separation principle are integrated to achieve explicit output in the intermediate reasoning process, and the few sample learning mechanism is used to break through the dependence on large-scale annotated data. Experimental and robustness analysis confirm that the framework breaks through the bottleneck of causal reasoning in large models. Future research will explore multimodal fusion (such as geometric visual language joint modeling), endogenous self validation ability [10] (based on attention mechanism to locate logical loopholes), and PDAG causal inference enhancement (through subtask extension and fine-tuning integration), promoting the development of large-scale model inference ability towards more complex and reliable cross modal causal inference.

References

[1] *Wang W, Zhu Z. Homogeneity and Sparsity Analysis for High-Dimensional Panel Data Models. 2024.*

[2] *Mendoza R, Cruz I, Liu R, et al. Adaptive Self-Supervised Learning Strategies for Dynamic On-Device LLM Personalization. 2024.*

[3] *Wang J, Ma Y L. Research on Personal Credit Risk Assessment Methods Based on Causal Inference. 2024.*

[4] *Yan J, Smith S F, Li J. WinkTPG: An Execution Framework for Multi-Agent Path Finding Using Temporal Reasoning. 2025.*

[5] *Cai C, Zhao X, Liu H, et al. The Role of Deductive and Inductive Reasoning in Large Language Models. 2024.*

[6] *Mari C, Mari E. Occam's razor, machine learning and stochastic modeling of complex systems: the case of the Italian energy market. Quality & Quantity, 2024, 58(2). Bhattacharya A, Stumpf S, De Croon R, et al. Explanatory Debiasing: Involving Domain Experts in the Data Generation Process to Mitigate Representation Bias in AI Systems. 2024.*

[7] *FangNaiyu, QiuLemiao, ZhangShuyou, et al. PG-VTON: A Novel Image-Based Virtual Try-On Method via Progressive Inference Paradigm. IEEE Transactions on Multimedia, 2024.*

[8] *Püttschneider, Jens, Faulwasser T. On Dissipativity of Cross-Entropy Loss in Training ResNets. 2024.*

[9] *Püttschneider, Jens, Faulwasser T. On Dissipativity of Cross-Entropy Loss in Training ResNets. 2024.*

[10] *Lu, Z. (2025). Design and Practice of AI Intelligent Mentor System for DevOps Education. European Journal of Education Science, 1(3), 25-31.*

[11] *Yu, X. (2025). Application Analysis of User Behavior Segmentation in Enhancing Customer Lifetime Value. Journal of Humanities, Arts and Social Science, 9(10).*

[12] *Zheng, H. (2025). Research on Lifecycle Configuration and Reclamation Strategies for Edge Nodes Based on Microservice Architectures. Advances in Computer and Communication, 6(5).*

[13] *Dingyuan Liu. The Relationship between Household Consumption Pattern Changes under Disasters and the Recovery of Business Ecosystems. Academic Journal of Business & Management (2025), Vol. 7, Issue 12: 151-156.*

[14] *Li, J. (2025). The Impact of Distributed Data Query Optimization on Large-Scale Data Processing.*

[15] *Zhen Zhong. Big Data Engineering and Intelligent Analysis Framework for Compliance Investigation. Academic Journal of Computing & Information Science (2025), Vol. 8, Issue 11: 107-115*