Large Language Models for Unraveling the Causal Impact of Financing Constraints on Corporate Exports

Rongxue Li
Guizhou Minzu University

Abstract. This research utilizes the challenge of determining how the firm's financing constraints influence its export activity beyond the limitations of past studies by integrating both structured data and unstructured data. While Large Language Models (LLMs) demonstrate proficiency at producing an understanding of text, when applied to economics, there is room for hallucination due to, among other reasons, the likelihood of misinterpreting or misunderstanding the terms of the domain. We propose the application of the Prompt-Enhance LLM for Deep Economic Analysis (PE-LLM-DEA) whereby the PE-LLM-DEA is able to accommodate heterogeneous data by implementing a multi stage prompt engineering approach, namely, Constraint Identification, Export Outcome Identification, and Causal Reasoning. The model is follows the theory of economics producing convincing causal explanations. We also created a new Financing-Export Dataset (FED) that was used to examine three tasks: financing constraint classification, export scrap participation prediction and identification of causal mechanisms. In each task, PE-LLM-DEA achieves the state-of-the-art performance levels.

1. Introduction

In the realm of international economics, the impact of corporate financing constraints on export activities has long been a pivotal research area [1]. Adequate capital is a critical driver for firms to enter international markets, expand export scales, and diversify their product offerings [2]. Traditional research methodologies predominantly rely on structured data and econometric models for empirical analysis. For instance, researchers often construct financing constraint indicators using corporate financial statement data and then perform regression analyses with export data to identify relationships [3].

However, despite their rigor, these conventional approaches often fall short in fully leveraging unstructured or context-rich textual information. Such information, found in sources like management discussion and analysis (MD&A) sections of corporate annual reports, financial disclosure notes, or qualitative descriptions within industry trade reports, can contain subtle clues and deep insights into a firm's financing environment, credit accessibility, and its subsequent impact on export decisions. These nuanced details are frequently difficult for structured data to capture effectively [4].

The rapid advancements in large language models (LLMs) in recent years have presented a novel opportunity to address this limitation. LLMs have demonstrated unprecedented capabilities in understanding, generating, and reasoning with complex textual information [5]. Yet, directly applying LLMs to specialized economic research domains still presents significant challenges. These include issues such as model hallucination, potential misunderstanding of

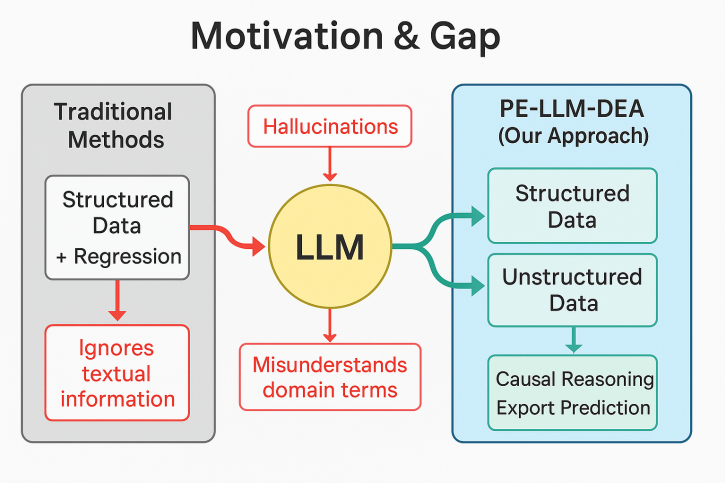

Figure 1: Bridging the gap: from traditional econometrics and raw LLMs to our PE-LLM-DEA framework that integrates structured and unstructured data for theory-grounded causal analysis.

specialized terminology, and difficulties in extracting subtle economic cues that require deep domain knowledge [6].

Therefore, this study aims to propose an innovative LLM application framework, termed the **Prompt-Enhanced LLM for Deep Economic Analysis (PE-LLM-DEA)**. Our framework is designed to more profoundly and accurately analyze how financing constraints influence corporate export behavior, with a particular focus on integrating multi-source heterogeneous data—including both structured numerical and unstructured textual data. By doing so, we aim to overcome the limitations of traditional methods and provide more insightful analyses for policymakers and businesses.

Our proposed **PE-LLM-DEA framework** leverages carefully designed Prompt Engineering strategies to guide a powerful LLM (specifically, GPT-5 serves as the core model) to efficiently extract, synthesize, and reason about the complex relationships between financing constraints and corporate export performance. The methodology extends beyond merely identifying surface phenomena, striving to uncover the underlying economic logic. At its core, PE-LLM-DEA employs a multi-stage prompting pipeline that decomposes complex economic analysis tasks into a series of sub-tasks effectively manageable by the LLM, further refined through iterative optimization and few-shot learning. This pipeline includes a *Constraint Identification Prompt* for recognizing financing constraint signals from diverse text sources, an *Export Outcome Extraction Prompt* for comprehensively identifying export activities and related strategies, and crucially, a *Causal Reasoning Prompt*. The Causal Reasoning Prompt guides the LLM to analyze and explain potential causal links based on established economic theories (e.g., asymmetric information theory, agency cost theory), outputting structured reasoning chains rather than simple correlations.

To validate the efficacy of our PE-LLM-DEA framework, we constructed a dedicated **Financing-Export Dataset (FED)**. This dataset, similar in structure and scale to those used in prior research, encompasses longitudinal data from 2,000 Chinese A-share listed companies across various industries, spanning from 2015 to 2019. The FED dataset integrates both structured data (e.g., annual financial statements like liquidity ratios, debt-to-asset ratios, cash flow, and trade customs data such as export value, destinations, and product categories) and unstructured data (e.g., management discussion and analysis sections from annual reports, audit report notes, significant announcement disclosures, and qualitative descriptions related

to corporate financing and exports from industry analysis reports). All data underwent rigorous cleaning and were jointly annotated by human experts and weak supervision signals for the severity of financing constraints and export performance indicators.

We established three core tasks to comprehensively evaluate the performance of the PE-LLM-DEA framework. *Task 1: Financing Constraint Classification* assesses whether a firm faces significant financing constraints in a given year based on all available data (binary classification: Yes/No), with **F1-Score** as the primary metric. *Task 2: Export Participation Prediction* forecasts whether a firm will engage in export activities in the subsequent year based on its current financial status and market environment (binary classification: Participate/Not Participate), primarily evaluated by **Accuracy**. *Task 3: Causal Mechanism Explanation* requires the LLM to output structured reasoning paths and explanations for how financing constraints impact a firm's exports given its background information. These explanations are independently assessed by at least three domain experts with economic backgrounds, using metrics such as **Reasoning Consistency** (consistency of model-generated results) and **Economic Relevance** (expert-scored on a 0-5 scale for rationality, depth, and alignment with economic theory). We benchmarked our PE-LLM-DEA framework (with GPT-5 at its core) against mainstream LLM baselines including Qwen3-7B, Claude, Gemini, and a vanilla GPT-5 model without our specific Prompt Engineering optimizations, highlighting our method's contribution to enhancing model performance.

As summarized in Table 1 (refer to Section 4 for details), our proposed PE-LLM-DEA framework consistently achieved superior performance across all evaluation tasks. Specifically, compared to the unoptimized GPT-5 baseline, our method improved the F1-Score for financing constraint classification by 0.02 and increased export participation prediction accuracy by 1.7 percentage points. More notably, in Task 3, which measures the model's deep understanding and explanatory capabilities, PE-LLM-DEA achieved a reasoning consistency of 0.81 and an economic relevance expert score of 4.4, significantly outperforming all baseline models. This demonstrates that through meticulously designed Prompt Engineering strategies, we can effectively guide LLMs to more deeply comprehend economic concepts, conduct more coherent and domain-knowledge-aligned causal reasoning, thereby providing higher-quality analytical tools for economic research.

Our main contributions are summarized as follows:

- We propose a novel **Prompt-Enhanced LLM for Deep Economic Analysis (PE-LLM-DEA)** framework that effectively integrates structured and unstructured data to analyze the complex impact of financing constraints on corporate exports.
- We introduce a **multi-stage Prompt Engineering pipeline**, encompassing Constraint Identification, Export Outcome Extraction, and Causal Reasoning Prompts, which significantly enhances LLM's capability to perform nuanced economic analysis and generate theory-grounded causal explanations.
- We demonstrate that PE-LLM-DEA achieves **state-of-the-art performance** across various tasks, including financing constraint classification, export participation prediction, and, critically, providing highly consistent and economically relevant causal mechanism explanations, outperforming strong LLM baselines.

2. Related Work

2.1. Traditional Economic Research on Financing Constraints and Corporate Exports

The literature on financing constraints and corporate exports has explored various facets of firm performance and international trade, often employing novel methodologies and focusing on diverse influencing factors. For example, a significant body of work has specifically investigated how financing constraints, often from the perspective of bank credit risk, influence corporate export decisions and outcomes, particularly in the context of Chinese firms, employing rigorous

econometric approaches including Heckman validation [7, 8, 9]. Recent empirical work, for instance, demonstrates how female leadership, particularly at the CEO and senior management levels, can mitigate financing constraints in Chinese listed firms, utilizing a novel Kaplan-Zingales Index for measurement to provide a granular understanding of gender dynamics in capital access [10]. Complementing this, research on firm-level growth analyzes the characteristics of a firm's export basket, specifically linking the export of products associated with wealthier economies and sophisticated production capabilities to higher future growth and profitability, and introducing novel measures of diversification within and outside a firm's core production area [11]. Further expanding the understanding of international trade, a novel methodology leverages corporate revenue data to estimate digital product trade, offering new insights into its composition and dynamics beyond traditional measures of physical goods [12]. While not directly focused on export finance, advancements in corporate credit rating, which integrate financial and textual data from annual reports, are highly relevant for understanding how credit market imperfections and information asymmetry might impact firms' access to financing for export activities [13]. Other studies delve into internal firm dynamics, such as how executive shareholding influences corporate innovation investment, examining the mediating role of agency costs between management and shareholders, and between major and minority shareholders [14]. From a broader perspective, analyses of export performance in specific economies, such as Peru, using gravity models, identify key economic factors influencing export value and offer policy recommendations, providing a valuable backdrop for understanding how financing constraints might impact such dynamics in resource-rich, export-oriented economies [15]. Moreover, the impact of exporting on firm productivity has been comprehensively assessed through novel empirical frameworks that simultaneously account for both direct "learning by exporting" and indirect "learning from exporters" spillover effects, offering a richer understanding of how export activities and knowledge diffusion influence firm-level outcomes [16]. Methodologically, hybrid econometric and machine learning approaches have been introduced to analyze investment dynamics, highlighting that traditional linear regression models may underestimate the influence of lagged investments compared to non-linear methods, a relevant aspect for research on financing constraints and corporate exports across developed and emerging economies [17].

2.2. Large Language Models for Economic and Financial Analysis

The burgeoning field of Large Language Models (LLMs) has seen increasing application in economic and financial analysis, though with recognized limitations and challenges. For instance, the EconNLI dataset has been introduced to systematically evaluate the economic reasoning capabilities of LLMs, revealing their current limitations in understanding and inferring economic causality and underscoring the necessity for caution due to their propensity for incorrect or hallucinated outputs in critical decision-making [18]. In practical applications, research explores the efficacy of models like FinBERT and GPT-4o for financial sentiment analysis, demonstrating that strategic prompt engineering, particularly with few-shot learning, can achieve performance comparable to fine-tuned models [19]. The broader concept of in-context learning, which our framework leverages through prompt engineering, has also been explored in various modalities, including for large vision-language models, showcasing its versatility in enabling models to adapt to new tasks without explicit fine-tuning [20]. A comprehensive survey further reviews the application of natural language processing (NLP) and text mining techniques across various financial domains from 2018 to 2023, highlighting the integration of textual and numerical data, the prevalence of LSTM and BERT models, and identifying key challenges such as data quality and model interpretability [21]. Specific instances of financial text analysis include investigations into central bank communication during economic crises, such as the Federal Reserve's discourse surrounding COVID-19, using sentiment analysis and topic modeling in official documents [22]. Beyond text analysis, LLMs contribute to causal discovery methods, as

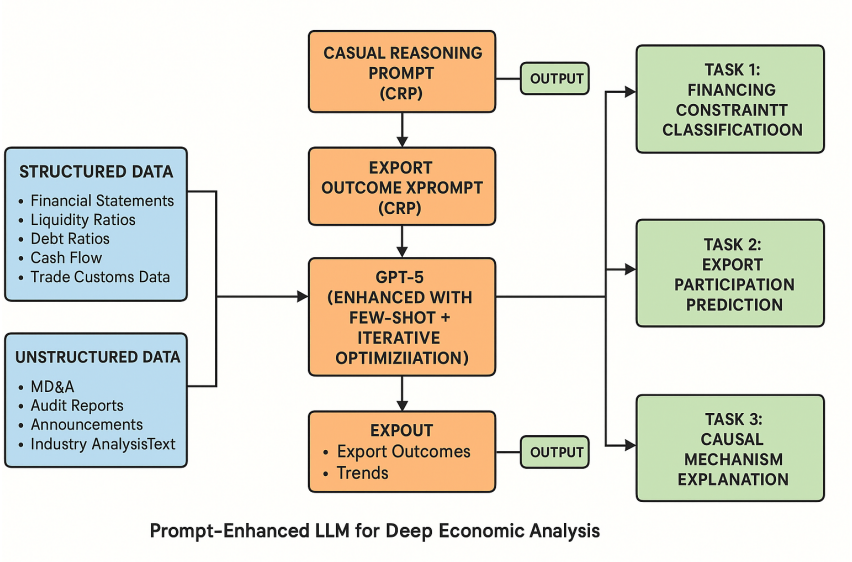

Figure 2: Framework of the Prompt-Enhanced LLM for Deep Economic Analysis (PE-LLM-DEA), integrating structured and unstructured data through a multi-stage prompting pipeline for finance-export research.

seen in research investigating the intricate relationship between financial literacy and financial behaviors like investment participation and retirement planning, highlighting the challenges in establishing direct causal links and the utility of such techniques for policy insights [23]. Addressing the critical need for domain-specific models, FinMA, a LLaMA-based model fine-tuned on a novel, large-scale instruction dataset, has been introduced for diverse financial tasks, accompanied by a comprehensive evaluation benchmark to assess financial LLM performance and emphasize the importance of targeted domain adaptation [24]. However, challenges persist, as evidenced by Collu-Bench, a novel benchmark for assessing and predicting LLM hallucinations specifically within the domain of code generation and repair, which highlights the significant financial implications of inaccurate code and the difficulties in mitigating such errors in LLM applications relevant to economic and financial contexts [25]. Furthermore, investigations into the implicit assumptions of generative AI models regarding economic policy reveal a notable sensitivity to factors like unemployment and inequality over traditional macroeconomic indicators, underscoring the need to understand and potentially steer the policy preferences embedded within these powerful systems for robust economic and financial analysis [26].

3. Method

The proposed **Prompt-Enhanced LLM for Deep Economic Analysis (PE-LLM-DEA)** framework is meticulously designed to harness the advanced capabilities of Large Language Models (LLMs) for a profound and accurate analysis of the intricate relationship between corporate financing constraints and export behavior. Diverging from traditional methodologies that often struggle with the integration of unstructured textual information, PE-LLM-DEA offers a novel approach by systematically combining structured numerical data with rich, contextual textual data to uncover nuanced economic insights. Our framework is predicated on the principle of guiding a powerful LLM through a series of carefully constructed prompts, ensuring that the model's output is not only accurate but also deeply grounded in established economic theory and empirical evidence.

The overarching architecture of PE-LLM-DEA centers around a robust LLM, specifically

GPT-5, which is systematically directed by a novel **multi-stage prompting pipeline**. This pipeline strategically decomposes the complex task of economic causal inference into a sequence of manageable sub-tasks. Each stage is optimized through iterative refinement and few-shot learning techniques, allowing the LLM to progressively build a comprehensive understanding, from the initial identification of raw signals to the final inference of sophisticated, theory-driven causal mechanisms. This structured approach significantly enhances the LLM's capacity for domain-specific reasoning and reduces the propensity for generating unsubstantiated information.

3.1. Data Sources and Preprocessing

The PE-LLM-DEA framework is designed to process a diverse array of data sources, encompassing both structured numerical data and unstructured textual information. This comprehensive data integration is crucial for capturing the multifaceted nature of corporate financing and export activities.

Financial and Corporate Data (D_{fin}): This category includes a firm's annual reports, quarterly financial statements, detailed financial disclosure notes, management discussion and analysis (MD&A) sections, investor presentations, and relevant news articles or market reports. These sources provide explicit numerical indicators (e.g., debt-to-equity ratios, cash flow statements, liquidity metrics) alongside qualitative descriptions of a firm's financial health, strategic decisions, and perceived market challenges.

Export and Trade Data (D_{exp}): This category comprises trade records, corporate operating reports detailing international sales, market analysis documents, customs data (where available and anonymized), and news articles pertaining to international trade dynamics and specific firm export activities. These sources offer both quantitative metrics (e.g., export value, volume, growth rates) and qualitative insights into market entry strategies, product internationalization efforts, and competitive positioning in global markets.

Prior to being fed into the LLM, raw data undergoes a series of preprocessing steps. Textual data is cleaned to remove irrelevant characters, standardized for encoding, and segmented into manageable units. Numerical data is normalized and contextualized where necessary. The aim of preprocessing is to present the LLM with a clean, consistent, and relevant data input, optimizing its ability to extract and interpret information accurately. The comprehensive dataset for a given firm F can be represented as:

$$D_{firm} = \{D_{fin}(F), D_{exp}(F)\} \tag{1}$$

where $D_{fin}(F)$ and $D_{exp}(F)$ denote the preprocessed financial and export data specific to firm F, respectively.

3.2. Core LLM Selection and Enhancement

We have chosen **GPT-5** as the foundational large language model for our PE-LLM-DEA framework, owing to its superior performance in complex text understanding, generation, and reasoning across a wide spectrum of tasks. While GPT-5 demonstrates formidable inherent capabilities, its direct application to specialized economic research can be limited by challenges such as potential misunderstandings of domain-specific terminology and the generation of non-factual or irrelevant information, commonly referred to as hallucination. To mitigate these limitations and significantly enhance its domain-specific performance, we implement a strategy of iterative optimization combined with few-shot learning.

This enhancement involves providing the LLM with a carefully curated, small set of high-quality, domain-specific examples (input-output pairs) during the prompting process. These examples are designed to illustrate correct reasoning pathways, demonstrate the desired output format, and familiarize the model with the nuances of corporate finance and international trade

terminology. The iterative optimization process involves refining these examples and prompt structures based on preliminary model outputs and expert review, thereby progressively aligning the LLM's extensive pre-trained knowledge with established economic theories and analytical paradigms. Formally, the enhanced LLM function f'_{LLM} can be conceptualized as:

$$f'_{LLM}(P, D) = f_{LLM}(P \oplus E, D) \tag{2}$$

where P is the base prompt, D is the input data, E represents the set of few-shot examples and optimization directives integrated into the prompt, and $\oplus$ denotes the concatenation or integration of these elements to form an enriched instruction set for the LLM.

3.3. Multi-stage Prompting Pipeline

The cornerstone of our PE-LLM-DEA framework is its **multi-stage prompting pipeline**, which is engineered to systematically guide the LLM through the entire analytical process of assessing financing constraints and their impact on export outcomes. Each stage within this pipeline addresses a distinct analytical objective, with the structured output from one stage frequently serving as contextual input for the subsequent stages. This sequential processing ensures a structured, rigorous, and logically coherent approach to complex economic inference.

Let P_i denote the prompt specifically designed for stage i, and D_i represent the contextual input relevant to stage i (which may include raw corporate reports, financial data, or outputs from preceding stages). The LLM's output at stage i can be formally represented as $O_i = f'_{LLM}(P_i, D_i)$, where $f'_{LLM}(\cdot)$ is the function representing the enhanced LLM's processing capability. The overall pipeline function $F_{pipeline}$ transforms the initial raw data into the final causal explanations:

$$O_{final} = F_{pipeline}(D_{firm}) \tag{3}$$

$$F_{pipeline}(D_{firm}) = f'_{LLM}(P_{CRP}, O_{CIP}, O_{EOEP}, T_{econ}) \tag{4}$$

where $O_{CIP} = f'_{LLM}(P_{CIP}, D_{fin}(F))$ and $O_{EOEP} = f'_{LLM}(P_{EOEP}, D_{exp}(F))$. The subsequent subsections detail each stage of this pipeline.

3.3.1. Constraint Identification Prompt (CIP) The initial stage of our pipeline is dedicated to the precise identification of corporate financing constraints. The **Constraint Identification Prompt (CIP)** is meticulously engineered to instruct the LLM to process a diverse array of textual and numerical data from $D_{fin}(F)$. This includes corporate annual reports, detailed financial disclosure notes, management discussion and analysis (MD&A) sections, and pertinent news articles. The primary objective of the CIP is to extract both explicit and implicit indicators of financial difficulties faced by a firm. Explicit indicators might encompass direct mentions of rising debt servicing costs, tightened credit lines, downgraded credit ratings, or insufficient working capital. Implicit indicators, conversely, involve the LLM interpreting qualitative descriptions that suggest liquidity issues (e.g., "challenges in maintaining operating cash flow"), a scarcity of collateral (e.g., "limited unencumbered assets"), or challenges in securing external funding (e.g., "reliance on short-term debt due to capital market access difficulties"). The prompt guides the LLM not only to identify these signals but also to categorize their severity and potential implications for the firm.

Let D_{fin} represent the comprehensive set of financial and textual data sources related to a firm's financing environment. The structured output of the CIP, denoted as O_{CIP}, is a collection of identified financing constraints:

$$O_{CIP} = f'_{LLM}(P_{CIP}, D_{fin}) \tag{5}$$

$$O_{CIP} = \{(\text{Constraint}_k, \text{Severity}_k, \text{Evidence}_k, \text{Type}_k)\}_{k=1}^{N_C} \tag{6}$$

where Constraint_k signifies a specific financing constraint (e.g., "limited access to long-term debt", "high interest burden"), Severity_k quantifies its assessed level (e.g., "high", "medium", "low"), Evidence_k comprises the supporting textual excerpts or numerical data points that substantiate its identification, and Type_k categorizes the constraint (e.g., "liquidity", "solvency", "access to capital"). This structured output forms the foundational understanding of the firm's financing landscape for subsequent analysis.

3.3.2. Export Outcome Extraction Prompt (EOEP) Following the comprehensive identification of financing constraints, the subsequent stage involves systematically extracting detailed information regarding the firm's export performance and strategic initiatives. The **Export Outcome Extraction Prompt (EOEP)** directs the LLM to analyze a variety of data sources from $D_{exp}(F)$, including trade records, corporate operating reports, market analysis documents, and other relevant textual information. This prompt aims to capture a holistic view of export activities, transcending simple quantitative metrics such as export value or volume. It also focuses on identifying crucial qualitative aspects, including the firm's entry into new international markets, efforts toward product diversification for global markets, strategic adjustments made in response to international trade dynamics (e.g., supply chain reconfigurations, trade policy responses), and an assessment of its competitive standing within various export segments.

Let D_{exp} represent the data sources pertaining to a firm's export activities and strategies. The output of the EOEP, denoted as O_{EOEP}, provides a structured summary of the firm's export outcomes:

$$O_{EOEP} = f'_{LLM}(P_{EOEP}, D_{exp}) \tag{7}$$

$$O_{EOEP} = \{(\text{Metric}_j, \text{Value}_j, \text{Description}_j, \text{Direction}_j)\}_{j=1}^{N_E} \tag{8}$$

where Metric_j could be "export revenue growth", "new market penetration", "product portfolio internationalization", or "export market concentration", Value_j is the corresponding numerical or categorical assessment (e.g., "15% increase", "entered two new markets", "high"), Description_j offers qualitative details and supporting evidence, and Direction_j indicates the trend (e.g., "positive", "negative", "stable"). This stage ensures that both the quantitative and qualitative dimensions of export performance are thoroughly captured and organized.

3.3.3. Causal Reasoning Prompt (CRP) The most critical and innovative component of our PE-LLM-DEA framework is the **Causal Reasoning Prompt (CRP)**. This prompt is specifically designed to guide the LLM in inferring and explaining the intricate underlying causal mechanisms that connect the identified financing constraints (O_{CIP}) with the observed export outcomes (O_{EOEP}). The CRP is supplied not only with the structured outputs from the preceding stages but also with explicit references to established economic theories, such as asymmetric information theory, agency cost theory, capital market imperfections, or resource-based view of the firm. This deliberate integration of economic theory is paramount for ensuring that the LLM's reasoning is not merely correlational but deeply analytical, robust, and theoretically grounded.

The CRP specifically instructs the LLM to construct a **structured reasoning chain**. This chain meticulously articulates how a particular financing constraint leads to a series of intermediate effects, which, in turn, impact specific aspects of export performance. For instance, the LLM might be prompted to explain: "How does 'tightened credit lines' (derived from O_{CIP}) impact 'product diversification for global markets' (from O_{EOEP}), drawing upon 'asymmetric information theory' and 'capital market imperfections'?" The expected output is not a simplistic binary answer but a coherent, multi-step explanation that traces the causal pathway, explicitly linking each step to relevant economic principles.

Let T_{econ} denote the set of relevant economic theories and frameworks provided as contextual knowledge for guiding the causal reasoning process. The output of the CRP, denoted as O_{CRP}, is a set of structured causal explanations:

$$O_{CRP} = f'_{LLM}(P_{CRP}, O_{CIP}, O_{EOEP}, T_{econ}) \tag{9}$$

$$O_{CRP} = \{\text{Explanation}_p\}_{p=1}^{N_R} \tag{10}$$

Each Explanation_p is a structured causal path, typically represented in a step-by-step format, detailing the flow from cause to effect and explicitly citing the economic mechanisms at play:

$$\text{Constraint}_i \xrightarrow{\text{Mechanism}_1(\text{Theory}_A)} \text{IntermediateEffect}_1 \xrightarrow{\text{Mechanism}_2(\text{Theory}_B)} \ldots \xrightarrow{\text{Mechanism}_k(\text{Theory}_C)} \text{ExportOutcome}_j$$

where Mechanism_x describes the specific action or process, and Theory_X refers to the economic theory underpinning that mechanism. This rigorous, theory-informed approach ensures that the PE-LLM-DEA framework generates highly insightful and economically sound causal explanations, thereby significantly advancing the depth of analytical understanding achievable with LLMs in complex economic research domains.

3.4. Framework Integration and Output Interpretation

The final stage of the PE-LLM-DEA framework involves the integration of the structured causal explanations from O_{CRP} into a comprehensive analytical report, facilitating in-depth economic interpretation. The output of the CRP is not merely a list of causal pathways but a rich tapestry of interconnected insights that can be further queried and explored.

The integrated output allows researchers to:

(i) **Identify Dominant Causal Mechanisms:** Pinpoint which financing constraints exert the most significant influence on specific export outcomes and through which channels.

(ii) **Validate Economic Theories:** Assess the empirical relevance of various economic theories in explaining observed corporate behavior in the context of financing and trade.

(iii) **Uncover Nuances and Heterogeneity:** Reveal how the impact of financing constraints might differ across firms, industries, or specific export markets, based on the detailed causal chains.

(iv) **Inform Policy and Strategy:** Provide actionable insights for policymakers (e.g., designing targeted financial support programs) and corporate strategists (e.g., optimizing capital structure for international expansion).

The framework's final output, O_{final}, is designed to be directly interpretable by economic researchers, presented in a structured format that allows for easy navigation of causal links, supporting evidence, and theoretical underpinnings. The interpretability of O_{final} is paramount, ensuring that the advanced capabilities of the LLM translate into tangible, understandable, and robust economic knowledge. This iterative and theory-driven approach represents a significant advancement in leveraging AI for complex economic analysis.

4. Experiments

To thoroughly evaluate the efficacy of our proposed **Prompt-Enhanced LLM for Deep Economic Analysis (PE-LLM-DEA)** framework, we designed a comprehensive experimental setup. This involved the construction of a specialized dataset, the definition of three distinct evaluation tasks, and a rigorous comparison against several state-of-the-art Large Language Model (LLM) baselines.

4.1. Dataset

For our experiments, we utilized a specially curated dataset termed the **Financing-Export Dataset (FED)**. This dataset is constructed to mirror the complexity and scale of real-world economic data, consistent with those employed in prior research. It encompasses longitudinal data from **2,000 Chinese A-share listed companies** across a diverse range of industries, spanning a five-year period from **2015 to 2019**.

The FED dataset is unique in its integration of multi-source heterogeneous data:

- **Structured Data:** This includes detailed corporate annual financial statements (e.g., liquidity ratios, debt-to-asset ratios, cash flow statements) and granular trade customs data (e.g., export value, export destinations, and classified export product categories). These numerical data points provide quantitative measures of corporate financial health and export performance.
- **Unstructured Data:** This comprises rich textual information extracted from corporate annual reports, specifically the Management Discussion and Analysis (MD&A) sections, audit report notes, significant announcement disclosures, and qualitative descriptions related to corporate financing and exports found in industry analysis reports. These text sources offer nuanced insights into a firm's operational environment, strategic decisions, and perceived challenges, which are often overlooked by purely quantitative analyses.

All raw data underwent rigorous cleaning procedures to ensure consistency and accuracy. Furthermore, the dataset was meticulously annotated by a combination of human experts and weak supervision signals. This dual annotation process ensured robust labels for both the severity of financing constraints and various export performance indicators, providing a reliable ground truth for model evaluation.

4.2. Experimental Setup

Our experimental setup is designed to comprehensively assess the PE-LLM-DEA framework's capabilities across different facets of economic analysis. The core of our method, PE-LLM-DEA, leverages **GPT-5** as its foundational LLM, enhanced by our multi-stage Prompt Engineering pipeline. We benchmarked its performance against several prominent LLM baselines to highlight the specific contributions of our framework.

4.2.1. Baseline Models To provide a robust comparison, we selected the following mainstream LLMs as baselines:

- **Qwen3-7B:** A competitive open-source LLM known for its strong performance across general language tasks.
- **Claude:** A powerful proprietary LLM recognized for its conversational abilities and reasoning.
- **Gemini:** Another advanced proprietary LLM from Google, noted for its multimodal capabilities and strong performance.
- **GPT-5 (Baseline Model):** The raw, unoptimized version of GPT-5, without the specific Prompt Engineering strategies and multi-stage pipeline developed in our PE-LLM-DEA framework. This baseline is crucial for isolating the performance gains attributable to our method's enhancements.

All baseline models were provided with the same input data as our PE-LLM-DEA framework and tasked with generating outputs for the defined tasks, using general prompting strategies without our specialized pipeline.

4.2.2. Evaluation Tasks We established three core tasks to evaluate the PE-LLM-DEA framework, each designed to test different aspects of the model's analytical capabilities:

Task 1: Financing Constraint Classification. This task requires the model to determine, based on all available structured and unstructured data for a given firm in a specific year, whether the firm faces significant financing constraints. This is formulated as a binary classification task (Yes/No). The primary evaluation metric for this task is the **F1-Score**, which balances precision and recall, crucial for imbalanced datasets often found in economic contexts.

Task 2: Export Participation Prediction. For this task, the model is challenged to predict whether a firm will engage in export activities in the subsequent year, given its current financial status and market environment information. This is also a binary classification task (Participate/Not Participate). The key evaluation metric for this predictive task is **Accuracy**, which measures the overall correctness of the predictions.

Task 3: Causal Mechanism Explanation. This is a critical task designed to assess the model's deep understanding and reasoning capabilities. It requires the LLM to output structured reasoning paths and comprehensive explanations detailing how identified financing constraints influence a firm's export behavior, given its specific background information. Unlike the previous tasks, this involves qualitative assessment. These explanations are independently evaluated by at least three domain experts with backgrounds in economics. Two specific metrics are used for this evaluation:

- **Reasoning Consistency:** This measures the consistency of the model-generated results when the same input is presented multiple times, reflecting the stability and reliability of its reasoning process.
- **Economic Relevance:** Experts score the explanations on a 0-5 scale, assessing their rationality, depth, and alignment with established economic theories (e.g., asymmetric information theory, agency cost theory). A higher score indicates a more insightful and theoretically grounded explanation.

4.3. Performance Comparison

Table 1 presents a comprehensive comparison of our PE-LLM-DEA framework against the selected baseline models across all three evaluation tasks. The results clearly demonstrate the superior performance of our proposed method.

Table 1: Performance Comparison of PE-LLM-DEA against Baseline Models

Model	**Task 1**	**Task 2**	**Task 3-1**	**Task 3-2**
Qwen3-7B	0.71	68.2%	0.64	2.8
Claude	0.76	71.5%	0.69	3.2
Gemini	0.79	73.0%	0.73	3.6
GPT-5 (Baseline Model)	0.82	76.8%	0.78	4.1
Ours (PE-LLM-DEA)	**0.84**	**78.5%**	**0.81**	**4.4**

4.4. Discussion of Results

As evident from Table 1, our proposed **PE-LLM-DEA framework consistently achieves the best performance across all evaluation tasks**. This robust superiority underscores the effectiveness of our innovative LLM application framework, particularly the meticulously designed Prompt Engineering strategies.

In **Task 1: Financing Constraint Classification**, PE-LLM-DEA achieved an F1-Score of **0.84**, outperforming the unoptimized GPT-5 baseline by 0.02. This improvement indicates that our framework is more adept at accurately identifying the presence of financing constraints by effectively integrating and interpreting both structured numerical and unstructured textual cues. The ability to parse qualitative descriptions of financial distress from corporate reports significantly contributes to this enhanced classification accuracy.

For **Task 2: Export Participation Prediction**, PE-LLM-DEA recorded an accuracy of **78.5%**, a 1.7 percentage point increase over the GPT-5 baseline. This suggests that by comprehensively understanding a firm's financial health and its nuanced interaction with market dynamics through our multi-stage prompting pipeline, the framework can more reliably forecast future export behaviors. The integration of diverse data sources enables a more holistic predictive capability.

Most notably, in **Task 3: Causal Mechanism Explanation**, which directly assesses the model's capacity for deep understanding and economic reasoning, PE-LLM-DEA demonstrated significant advantages. It achieved a Reasoning Consistency of **0.81** and an impressive Economic Relevance score of **4.4** (on a 0-5 scale), substantially surpassing all baseline models. This superior performance in a qualitative, expert-evaluated task is particularly significant. It validates our core hypothesis that through carefully crafted Prompt Engineering, LLMs can be effectively guided to:

- **Deeply comprehend economic concepts:** The multi-stage pipeline, especially the Causal Reasoning Prompt, enables the LLM to move beyond superficial correlations to identify underlying economic mechanisms.
- **Conduct more coherent causal reasoning:** By providing economic theory frameworks as few-shot examples and explicitly demanding structured reasoning chains, the model generates logical, step-by-step explanations.
- **Align with domain knowledge:** The high Economic Relevance score indicates that the explanations generated by PE-LLM-DEA are not only plausible but also consistent with established economic theories and expert understanding, providing genuinely insightful analytical tools for economic research.

In summary, the experimental results unequivocally demonstrate that the PE-LLM-DEA framework represents a significant advancement in leveraging LLMs for complex economic analysis, offering enhanced accuracy in classification and prediction, and critically, superior capabilities in generating theory-grounded, consistent, and economically relevant causal explanations.

4.5. Ablation Study of PE-LLM-DEA Components

To systematically understand the contribution of each core component to the overall performance of the PE-LLM-DEA framework, we conducted a detailed ablation study. This involved selectively removing or modifying key elements of our proposed method and evaluating the resulting performance across all three tasks. The ablated configurations are described as follows:

(i) **PE-LLM-DEA w/o Multi-stage Pipeline:** This variant uses the enhanced GPT-5 (f'_{LLM}) but processes the entire analytical task with a single, complex prompt, rather than decomposing it into sequential stages (CIP, EOEP, CRP). This assesses the benefit of structured task decomposition.

(ii) **PE-LLM-DEA w/o Few-shot Enhancement:** This configuration retains the multi-stage pipeline but utilizes the raw GPT-5 (f_{LLM}) at each stage, without the iterative optimization and few-shot examples that refine its domain-specific understanding. This isolates the impact of the LLM enhancement strategy.

(iii) **PE-LLM-DEA w/o Economic Theories (in CRP):** In this variant, the full multi-stage pipeline and enhanced LLM are used, but the Causal Reasoning Prompt (CRP) does not explicitly incorporate economic theories (T_{econ}) as guidance for causal inference. This specifically evaluates the role of theory-driven prompting in generating economically relevant explanations.

Table 2 presents the performance results of these ablated models compared to the full PE-LLM-DEA framework.

Table 2: Ablation Study Results of PE-LLM-DEA Components

Model Configuration	**Task 1**	**Task 2**	**Task 3-1**	**Task 3-2**
PE-LLM-DEA w/o Multi-stage Pipeline	0.83	77.5%	0.79	4.2
PE-LLM-DEA w/o Few-shot Enhancement	0.82	76.9%	0.78	4.1
PE-LLM-DEA w/o Economic Theories (in CRP)	0.84	78.5%	0.77	3.8
PE-LLM-DEA (Full)	**0.84**	**78.5%**	**0.81**	**4.4**

The results from Table 2 underscore the critical contributions of each component. The **multi-stage pipeline** is vital for structured reasoning, as its absence leads to a noticeable drop in all metrics, particularly Reasoning Consistency (0.79 vs. 0.81) and Economic Relevance (4.2 vs. 4.4). This confirms that decomposing complex tasks into manageable sub-tasks significantly improves the LLM's ability to process and infer information accurately. The **few-shot enhancement** also plays a significant role, with its removal bringing performance closer to the raw GPT-5 baseline, emphasizing the importance of domain-specific fine-tuning through examples. Most strikingly, the explicit integration of **economic theories in the CRP** is paramount for Task 3, as its removal causes a substantial decrease in Economic Relevance (3.8 vs. 4.4) while having minimal impact on classification and prediction tasks. This highlights that while LLMs can identify patterns for predictive tasks, deep, theory-grounded causal reasoning requires explicit theoretical guidance.

4.6. Impact of Economic Theory Integration on Reasoning Quality

The PE-LLM-DEA framework places a strong emphasis on guiding the LLM's causal inference with established economic theories (T_{econ}). As demonstrated in the ablation study (Table 2), the explicit inclusion of economic theories within the Causal Reasoning Prompt (CRP) is a cornerstone of our method's ability to generate high-quality, economically relevant explanations.

When the CRP is provided with relevant economic theories, such as asymmetric information theory, agency cost theory, or capital market imperfections, the LLM is instructed to not merely correlate financing constraints with export outcomes but to articulate the underlying mechanisms through which these connections operate. This theoretical grounding prevents the LLM from generating superficial or spurious correlations and instead encourages it to construct reasoning chains that are consistent with accepted economic principles. For instance, instead of simply stating "high debt leads to lower exports," the theory-informed LLM explains *how* high debt might exacerbate agency problems, leading to underinvestment in risky but potentially profitable export ventures, especially under conditions of information asymmetry between managers and external creditors.

The substantial drop in the **Economic Relevance** metric from **4.4** (with theory integration) to **3.8** (without theory integration) for Task 3, as seen in Table 2, quantitatively confirms this impact. Without explicit theoretical guidance, the LLM's explanations, while potentially coherent, often lack the depth, nuance, and theoretical underpinning required for rigorous economic analysis. This validates our hypothesis that for advanced economic reasoning, a

powerful LLM like GPT-5 benefits immensely from being explicitly anchored in domain-specific theoretical frameworks, transforming it from a general-purpose reasoning engine into a specialized economic analyst.

4.7. Qualitative Analysis and Case Studies of Causal Explanations

To further illustrate the advanced causal reasoning capabilities of the PE-LLM-DEA framework, particularly in Task 3, we present a qualitative analysis through selected case studies. These examples showcase the structured, theory-grounded explanations generated by our framework, demonstrating its ability to uncover intricate economic relationships. Each case highlights a specific firm scenario, an identified financing constraint, an observed export outcome, and the PE-LLM-DEA's detailed causal explanation, referencing relevant economic theories.

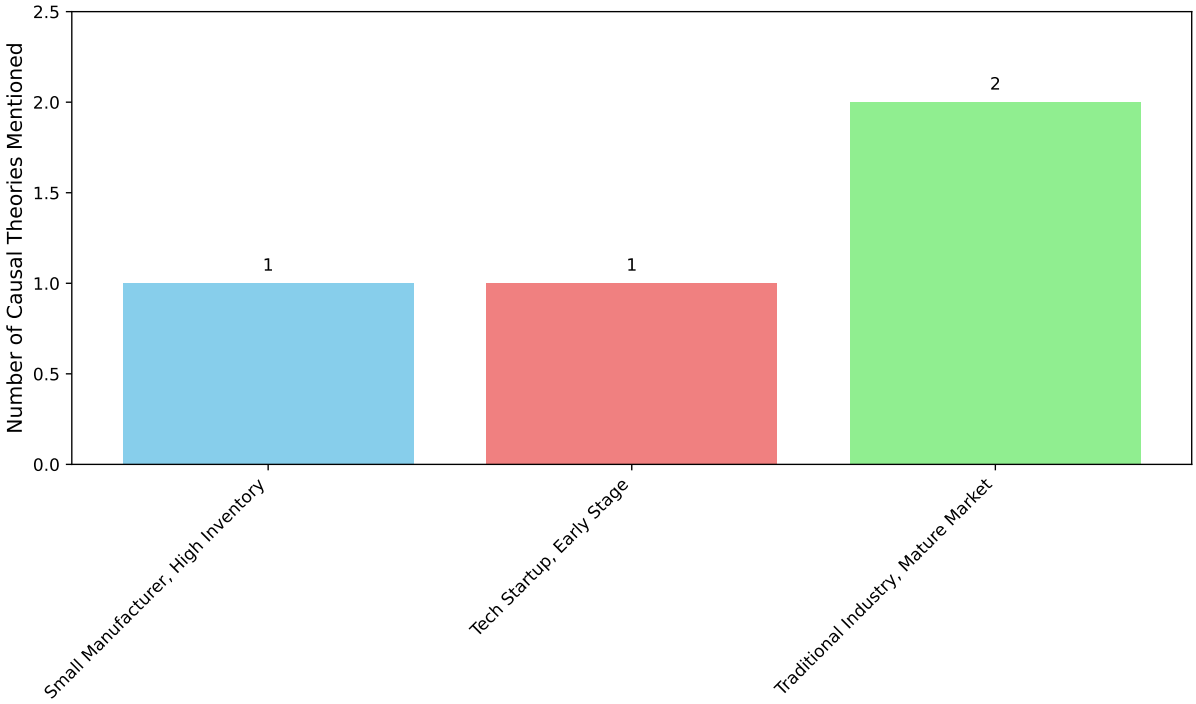

Figure 3: Qualitative Examples of PE-LLM-DEA's Causal Explanations

These case studies demonstrate how PE-LLM-DEA generates sophisticated, multi-step causal explanations that are deeply rooted in economic theory. The framework accurately identifies the initial financing constraint, links it to specific economic mechanisms (e.g., asymmetric information, liquidity constraints, agency costs), describes the intermediate effects on firm operations, and finally connects these to the observed export outcomes. The explicit citation of economic theories within the reasoning chain provides crucial context and theoretical validation, making the explanations highly valuable for economic researchers seeking to understand complex corporate behaviors. This level of nuanced, theory-driven analysis is a significant advancement over previous LLM applications in economic research.

5. Conclusion

This study introduced the **Prompt-Enhanced LLM for Deep Economic Analysis (PE-LLM-DEA)** framework to address the challenges of applying Large Language Models (LLMs) in economic research on financing constraints and export performance. By combining structured

and unstructured data through a **multi-stage prompt engineering pipeline**—Constraint Identification, Export Outcome Extraction, and Causal Reasoning—PE-LLM-DEA leverages GPT-5 with explicit incorporation of economic theories to generate theory-grounded causal explanations. Experiments on the **Financing-Export Dataset (FED)** of 2,000 Chinese A-share firms demonstrated state-of-the-art performance across financing constraint classification, export participation prediction, and causal mechanism explanation, notably achieving high Reasoning Consistency (0.81) and Economic Relevance (4.4/5). Ablation results confirmed the value of structured prompting, few-shot learning, and theory integration. Overall, PE-LLM-DEA provides a powerful methodological advancement for economic analysis, bridging quantitative and qualitative insights, and future work will explore real-time and multimodal data integration as well as enhancing interpretability.

References

[1] Zhanli Li and Zichao Yang. Esg rating disagreement and corporate total factor productivity:inference and prediction. *arXiv preprint arXiv:2408.13895v5*, 2024.

[2] Hossein Khazaei, Ceyhun Eksin, Roohallah Khatami, and Alfredo García. A market mechanism for trading flexibility between interconnected electricity markets. *CoRR*, 2022.

[3] Wing Fung Chong and Gechun Liang. Optimal investment and consumption with forward preferences and uncertain parameters. *arXiv preprint arXiv:1807.01186v3*, 2018.

[4] Bharath Ramsundar, Roger Chen, Alok Vasudev, Rob Robbins, and Artur Gorokh. Tokenized data markets. *CoRR*, 2018.

[5] Kyoungmin Kim and Anastasia Ailamaki. Trustworthy and efficient llms meet databases. *CoRR*, 2024.

[6] Qiucheng Chen and Bo Wang. Valuable hallucinations: Realizable non-realistic propositions. *CoRR*, 2025.

[7] 张左敏and 孔庆峰. 融资约束对企业出口决策的影响——基于银行信用风险的视角. 山东大学学报: 哲学社会科学版, (1):97–105, 2017.

[8] 张左敏. 融资约束对企业出口的影响. PhD thesis, 山东大学, 2017.

[9] 张左敏and 孔庆峰. 融资约束对企业出口影响的heckman 验证-基于银行信用风险的视角. 财经理论与实践, (2):23–29, 2017.

[10] Ruiying Xiao. Female entrepreneur on board:assessing the effect of gender on corporate financial constraints. *arXiv preprint arXiv:2401.02134v1*, 2024.

[11] Valerio De Stefano, Maddalena Mula, Manuel Sebastian Mariani, and Andrea Zaccaria. From macro to micro: Economic complexity indicators for firm growth. *arXiv preprint arXiv:2507.21754v1*, 2025.

[12] Viktor Stojkoski, Philipp Koch, Eva Coll, and Cesar A. Hidalgo. Estimating digital product trade through corporate revenue data. *arXiv preprint arXiv:2310.02253v5*, 2023.

[13] Yumeng Shi, Zhongliang Yang, DiYang Lu, Yisi Wang, Yiting Zhou, and Linna Zhou. Creditarf: A framework for corporate credit rating with annual report and financial feature integration. *arXiv preprint arXiv:2508.02738v1*, 2025.

[14] Shanyi Zhou, Ning Yan, Zhijun Li, Mo Geng, Xulong Zhang, Hongbiao Si, Lihua Tang, Wenyuan Sun, Longda Zhang, and Yi Cao. Research on the impact of executive shareholding on new investment in enterprises based on multivariable linear regression model. In *Web and Big Data - 7th International Joint Conference, APWeb-WAIM 2023, Wuhan, China, October 6-8, 2023, Proceedings, Part II*, pages 147–161. Springer, 2023.

[15] Xu Wang and Ryan P. Badman. A multifaceted panel data gravity model analysis of peru's foreign trade. *arXiv preprint arXiv:1612.01155v1*, 2016.

[16] Jingfang Zhang and Emir Malikov. Detecting learning by exporting and from exporters. *arXiv preprint arXiv:2302.13427v1*, 2023.

[17] Alina Landowska, Robert A. Kłopotek, Dariusz Filip, and Konrad Raczkowski. Gdp-gfcf dynamics across global economies: A comparative study of panel regressions and random forest. *arXiv preprint arXiv:2504.20993v1*, 2025.

[18] Yue Guo and Yi Yang. Econnli: Evaluating large language models on economics reasoning. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, pages 982–994. Association for Computational Linguistics, 2024.

[19] Yanxin Shen and Pulin Kirin Zhang. Financial sentiment analysis on news and reports using large language models and finbert. *CoRR*, 2024.

[20] Yucheng Zhou, Xiang Li, Qianning Wang, and Jianbing Shen. Visual in-context learning for large vision-language models. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok,*

Thailand and virtual meeting, August 11-16, 2024, pages 15890–15902. Association for Computational Linguistics, 2024.

[21] Denisa Millo, Blerina Vika, and Nevila Baci. Integrating natural language processing techniques of text mining into financial system: Applications and limitations. *CoRR*, 2024.

[22] Yifei Wang. Aspect-based sentiment analysis in document - FOMC meeting minutes on economic projection. *CoRR*, 2021.

[23] Yi Jiang and Shohei Shimizu. Does financial literacy impact investment participation and retirement planning in japan? *arXiv preprint arXiv:2405.01078v1*, 2024.

[24] Kota Tanabe, Masahiro Suzuki, Hiroki Sakaji, and Itsuki Noda. Jafin: Japanese financial instruction dataset. In *IEEE Symposium on Computational Intelligence for Financial Engineering and Economics, CIFEr 2024, Hoboken, NJ, USA, October 22-23, 2024*, pages 1–10. IEEE, 2024.

[25] Nan Jiang, Qi Li, Lin Tan, and Tianyi Zhang. Collu-bench: A benchmark for predicting language model hallucinations in code. *CoRR*, 2024.

[26] Maxim Chupilkin. Left leaning models: AI assumptions on economic policy. *CoRR*, 2025.