

Summary of Global Business Networks (Zheng Zhou ; October 19, 2025)

1) The **research question**: How can AI and large language models be leveraged to generate historical business descriptions to construct time-varying business networks that represent business links across the globe.

2) **Motivation**: To enhance the prediction of global business activities through global business networks constructed by artificial intelligence .

3) **Marginal contributions to the literature**: 1. Use AI to generate historical business descriptions, thereby providing—for the first time—a global business network. 2. Mask firm-specific details to mitigate look-ahead bias concerns arising from the use of embedding models with a recent knowledge cutoff. 3. Differentiate between competitor, supplier, and customer links by fine-tuning an open-source language model.

4) **Hypothesis**: The global business networks demonstrate strong predictive capability, particularly in studying the lead-lag effect for global stocks and predicting target firms in M&A deals.

a) This hypothesis follows from the research question and b) it follows from the large language models. The **logic of the hypothesis** is shown below(**visualization**)

business descriptions -> high-dimensional vectors -> a cosine similarity matrix -> firm pairs surpassing a threshold -> business networks

5) **Sample(Data)**: Mask firm-specific details in the descriptions to generate business networks

6) **Independent variable**: Business networks **Dependent variable**: Lead-lag effects(alpha)/Merger and Acquisition (M&A) transactions(0&1)/ Business relationships(competitor, supplier, and customer) in order to test the accuracy of the global business networks in different dimensions.

7) **Model**: large language models -> BNs; multi-factor model -> Lead-lag effects; logistic regressions -> M&A; fine-tune the large language models -> Business relationships

8) **Difficulties** arise in drawing inferences from the empirical work: Look-ahead bias arising from the use of embedding models with a recent knowledge cutoff.

9) **Publishable and feasible extension** of this research: This approach can be extended to other types of firm descriptions, such as product and service descriptions.

Summary of Can Generative AI Help Identify Peer Firms? (Zheng Zhou ; October 19, 2025)

- 1) The **research question**: How well can generative AI identify product market competitors?
- 2) **Motivation**: Construct a better peer identification system to enhance the prediction capability for corporate behaviors.
- 3) **Marginal contributions to the literature**: 1. Evaluate the potential of generative large language models in identifying peer firms. 2. Apply generative LLMs as an information aggregator to identify peer firms.
- 4) **Hypotheses**: **1)** Machine-generated peers(LLM) have high correlations with the focal firm in stock returns, sales growth, and gross profit margin in the subsequent year, **2)** and exhibit higher homogeneity among themselves, compared with peers identified by analyzing the similarity of business descriptions in annual reports(TNIC) or by using members in the focal firm's SIC industry(SIC).
 - a) These hypotheses follow from the research question, b)and follow from generative large language models. The **logic of the hypotheses(visualization)**: LLMs that are pre-trained on a large amount of data -> set the temperature to 0 -> enter prompts in Bard to generate a focal firm's peers -> Manually process the machine-generated peers
- 5) **Sample selection**: Require the Levenshtein distance of the closest match for a peer name to be within 0.4. Exclude focal firm-years that do not have any peer matches or only have focal firm itself.
- 6) **Independent variable**: peers generated by LLMs(key)/TNIC/SIC
Dependent variable: Correlations with the focal firm in stock returns, sales growth, and gross profit margin/ Homogeneity among themselves
- 7) Regression **model specification**:
$$\text{Ret}_{i,t+1,m} = \alpha_{i,t} + \beta_{i,t}\text{PortRet}_{i,t+1,m} + \varepsilon_{i,t+1,m} \quad (1)$$
$$\text{AFP}_{i,t+1} = \alpha_t + \beta_t\text{PortAFP}_{i,t+1} + \varepsilon_{i,t+1} \quad (2)$$
- 8) Difficulties arise in drawing **inferences from the empirical work**: LLM peers exhibit an upward bias in compensation benchmarking.
- 9) Publishable and **feasible extension** of this research: use generative AI to identify other types of business relationships, such as product market suppliers and customers.