ELSEVIER

Journal of Financial Economics

Global Business Networks☆

Christian Breitung *, Sebastian Müller

Technical University of Munich, TUM School of Management, Campus Heilbronn, Center for Digital Transformation, Am Bildungscampus 9, Heilbronn, 74076, Germany

ARTICLE INFO

JEL classification:
G10
G12
G14

Keywords:
Business network
Textual analysis
Natural language processing
GPT-3
Large language models

ABSTRACT

We leverage the capabilities of GPT-3 to generate historical business descriptions for over 63,000 global firms. Utilizing these descriptions and advanced embedding models from OpenAI, we construct time-varying business networks that represent business links across the globe. We showcase the performance of these networks by studying the lead–lag effect for global stocks and predicting target firms in M&A deals. We demonstrate how masking firm-specific details can mitigate look-ahead bias concerns that may arise from the use of embedding models with a recent knowledge cutoff, and how to differentiate between competitor, supplier, and customer links by fine-tuning an open-source language model.

1. Introduction

In recent decades, global economies have experienced a growing trend of specialization. Companies in advanced economies frequently focus on producing highly specialized products, which leads to considerable heterogeneity within industries. For instance, two automobile manufacturers may offer comparable products with distinct features, such as electric or gasoline engines, or cater to different market segments, such as luxury or budget consumers. Additionally, these companies may also differ in other domains, such as their digitization levels, supply chain resilience, and geographical locations. Previous research emphasizes that traditional sector and industry classifications may not effectively represent this within-sector variety (Hoberg and Phillips, 2016). Instead, each company has a unique network of affiliated competitors, suppliers, and customers, interconnected through economic ties.

In this paper, we are the first to generate time-varying global business networks by applying two state-of-the-art embedding models from OpenAI[1] as well as an open-source embedding model to historical business descriptions of more than 63,000 publicly traded firms across 67 countries. We test the accuracy of our global networks in various dimensions and find that they reveal value-relevant economic links.

We introduce an innovative method that enables us to generate historical business descriptions from annual reports. We obtain 10-K filings for US companies from EDGAR, a platform operated by the Security Exchange Commission (SEC) and international reports from the London Stock Exchange Group (LSEG). We then apply Artificial Intelligence (AI) to identify business-specific information and instruct the large language model (LLM) *GPT-3* to create concise business descriptions aligned to those of commercial data providers. In combination with a limited number of additional historical descriptions

☆ Toni Whited was the editor for this article. We thank the editor, an anonymous reviewer, Vincent Bogousslavsky, Charlie Cai, Gerard Hoberg, Christoph Merkle, Gordon Phillips, Jeffrey Pontiff, Oliver Rehbein, Ronnie Sadka, and Christoph Schiller for valuable comments and feedback. We further thank participants at the 2024 conference of the Financial Intermediation Research Society (FIRS), the AFA 2024 Annual Meeting, the EFMA 2024 Annual Conference, the EUROFIDAI-ESSEC Paris December 2023 Finance Meeting, the 29th Annual Meeting of the German Finance Association, the FinEML Conference 2023, and the TUM Global Technology Forum 2023, as well as seminar participants at the Aalto University, Carroll School of Management (Boston College), the University of Liverpool, the Poznan University of Economics and Business, and the TUM School of Management. We also express our gratitude to Hendrik Leber and Roland Schweter from Acatis Investment for providing a record of historical S&P Global business descriptions. Global business networks and AI-generated historical business descriptions are available at www.global-business-networks.com.

[1] We use the two embedding models *text-embedding-3-small* and *text-embedding-3-large* from OpenAI. These models represent text as numerical vectors, enabling the identification of textual similarity. For robustness, we include both the smaller and larger embedding models in the paper.

Received 10 July 2023; Received in revised form 14 January 2025; Accepted 18 January 2025

Available online 13 February 2025

obtained from external data providers, our business description dataset covers between 91.6% and 99.8% of the market capitalization in the US and between 79.9% and 98.3% internationally between 2000 and 2021.

To derive the global business networks, we represent these business descriptions as high-dimensional vectors using advanced embedding models and construct a cosine similarity matrix based on the embeddings. We then determine firm pairs exceeding the 99th percentile of the cosine similarity distribution as economically linked. The global business network thus includes all firm relations that surpass this threshold.

Applying OpenAI embeddings may introduce a look-ahead bias due to the training data of its foundational model GPT-3, which extends until September 2021. For example, if the model's training data includes information about *Procter & Gamble's* acquisition of *Gillette* in 2005 (SEC, 2005), the embedding model might produce highly similar embeddings for historical descriptions of both firms despite their distinctiveness before the acquisition. We address this bias by effectively masking company-identifying information to prevent the model from using additional company-specific information, similar to Glasserman and Lin (2023) and Kim et al. (2024).

Our networks can be useful for a variety of research questions, particularly those with a global focus on firm competition, industrial organization, and informational spillovers. For instance, similar to Hoberg and Phillips (2010), one can investigate to what extent business similarity may predict future mergers and acquisitions in international markets. Researchers could also analyze ESG-related spillover effects (Li et al., 2023), the relation between corporate investment and peer valuation (Foucault and Fresard, 2014), or the influence of recent IPOs on competition in the network (Aghamolla and Thakor, 2022). Additionally, the networks can be applied to study lead–lag effects in stock returns of economically related firms across countries or to build efficient global (factor) portfolios that remove industry-specific risks in line with Daniel et al. (2020). We add to the work of Frésard et al. (2020), who create a vertical link network, and show how to distinguish between competitor- and supplier-customer-relations in the networks. This may foster research on global supply chains, such as determinants of resilience during exogenous shocks like Covid-19.

We employ a multi-dimensional approach to assess the accuracy of our global business networks. Our first analysis focuses on the proportion of firms with peers in identical industries and countries, as firms typically share more similarities along these dimensions. We find that our embedding-based networks exhibit significantly higher industry and country congruence than a word-based network. However, even with state-of-the-art embedding models, only about half of the resulting firm relations are domestic, with notable differences across countries.

Next, we calculate and compare pairwise firm overlaps and return correlations across all business networks. The context-aware networks display significantly higher business link co-occurrences and enhanced return correlations. For example, around 58% of the relations in the smaller OpenAI embedding model are mirrored in its larger model. The average return correlation between competitor portfolios from both networks is as high as 0.85.

We then benchmark our networks' accuracy against a word-based network and the TNIC dataset (Hoberg and Phillips, 2010, 2016) by assessing the identification of US competition relations disclosed in annual reports, M&A filings, and FactSet Revere, which features competitor, supplier, and customer links. Regardless of the dataset, networks based on the OpenAI embedding models consistently outperform word-based networks and perform comparably to TNIC.

We next showcase two of the above mentioned potential applications for our business networks. This provides insights into their accuracy and the extent of a potential *LLM look-ahead bias* that might occur if we do not mask company-identifying information.

First, we revisit the documented lead–lag effect. As summarized by Ali and Hirshleifer (2020), a vast body of literature establishes predictive links for stock returns among firms grouped within the same industry (Moskowitz and Grinblatt, 1999; Hoberg and Phillips, 2018), sharing a similar geographic location (Parsons et al., 2020), related through supply chains (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010), or utilizing similar technologies (Lee et al., 2019). Moreover, Cohen and Lou (2012) find lead–lag effects from single-segment to multi-segment corporations from the same industry, and Müller (2019) identifies economic links between stocks with similar firm characteristics. While these studies focus exclusively on the US stock market, Huang (2015) utilizes international industry-level returns to forecast future returns of US multinational firms, and Finke and Weigert (2017) study lead–lag effects for multinational firms from 22 developed markets.

We extend these studies by examining the lead–lag effect for a comprehensive sample of global stocks covering both developed and emerging markets, as well as multinational and domestic firms. For each business network, we implement a strategy to go long (short) in stocks whose economically linked stocks performed best (worst) in the preceding month. The portfolio yielding the highest seven-factor alpha (controlling for the five factors of Fama and French (2015) plus momentum and short-term reversal) might best represent its economic links.

As expected, we observe positive alphas across all networks. In the US, portfolios based on context-aware networks outperform word-based networks by up to a statistically significant 27 bps per month. With alphas ranging between 119 and 146 bps per month, these portfolios perform similarly to a strategy based on TNIC (156 bps per month) and are highly statistically significant with t-statistics around six. In a global setting of US and international stocks, context-aware networks yield seven-factor alphas of up to 281 bps per month, outperforming a comparable strategy based on bag-of-words by up to 73 bps, with the performance difference being highly statistically significant. We also observe statistically significant value-weighted monthly alphas of up to 40 bps in the US and 74 bps globally. Employing a capped value-weighted approach to avoid a dominating influence of mega-caps on portfolio returns (Jensen et al., 2023), the alphas are up to 81 bps (165 bps) per month for the US (global) lead–lag strategy. These results change only minimally with masked networks, indicating that the *LLM look-ahead bias* does not seem to play a major role here.

In the second application, we examine our networks' ability to predict target firms in M&A deals, building on Hoberg and Phillips (2010), who show that target firms often operate in similar product markets. We find that around 50% of US target firms for US acquirers rank among the top 100 firms with the highest business description similarity to the acquiring firm, comparable to TNIC, which detects 58%. Internationally, we observe a similar effect.

Motivated by these results, we conduct a logistic regression analysis, controlling for industry membership, country, profitability, and other variables. We find that high textual similarity between two firms' business descriptions significantly increases the likelihood of an M&A deal. However, this predictive power is significantly lower when using masked business descriptions, though it remains statistically significant. This difference suggests a look-ahead bias in predicting future M&A deals with original descriptions, aligning with the example of *Procter & Gamble* and *Gillette*. However, our findings suggest that the *LLM look-ahead bias* can be significantly reduced through masking.

Our networks map global business relations without distinguishing whether companies are competitors or linked through customer–supplier relationships. However, depending on their research questions, researchers might be interested in focusing on specific types of relations. Our final analysis shows how a language model can be fine-tuned to discern the nature of potential business relations in our networks. This fine-tuning uses actual business relations documented in FactSet Revere. We test this approach and achieve an accuracy of 85.73% for a multiclass classification model trained on AI-generated descriptions to differentiate between competitors, suppliers, and customers.

This study introduces two main methodological advances in Finance and Economics. First, we enhance the field of textual analysis in Finance. While existing research applies textual analysis to identify competitors (Eisdorfer et al., 2021), gauge competition intensity (Li et al., 2013), and develop time-varying industry classifications (Hoberg and Phillips, 2010, 2016), our work demonstrates the effectiveness of embedding models in representing business information. Second, we show how generative AI can identify, summarize, and streamline information in corporate disclosures across a global spectrum of stocks. This enables us to offer the first global business networks, encouraging researchers to study global economic links.

While our networks cover the vast majority of stocks across the globe, there are certain limitations researchers should be aware of when using our data, and recommendations they should consider. First, smaller stocks are somewhat underrepresented in the networks, especially at the beginning of our sample period. Second, we observe a lower coverage in some Asian and African markets, which mainly stems from difficulties of the GPT-3 tokenizer in processing certain languages like Chinese. Third, to mitigate look-ahead bias concerns, researchers are advised to use our masked business networks as this limits the embedding models' abilities to consider other company-specific information. Fourth, researchers who wish to mask specific text parts in other research settings should use high-quality named entity recognition models to ensure a high masking accuracy. While we use the advanced model from the Python package "spaCy", we also find that simpler models can result in significantly lower masking accuracy. Finally, researchers that are solely interested in the US market may consider US-specific networks (see Hoberg and Phillips, 2010, 2016), as they are constructed on entire Item 1 sections rather than concise business descriptions, have a higher firm coverage, and are available for a longer sample period. In addition, in a recent work, Hoberg and Phillips (2025) also introduce an embedding-based network for US stocks and report a 20% improvement in the informativeness of their industry classification.

The structure of this paper is as follows. Section 2 details the construction of business networks and highlights potential biases inherent in this process. Section 3 presents a comprehensive overview of the business descriptions and the stock data sourced from LSEG and CRSP (Center for Research in Security Prices). In Section 4, we evaluate the efficacy of various business networks across multiple aspects before we showcase the performance of our networks in Section 5. Section 6 discusses how language models can be fine-tuned to distinguish between competitors, suppliers, and customers. We conclude in Section 7.

2. Methodology

2.1. Established ways of creating Business Networks (BNs)

To construct global business networks (BNs), it is crucial to identify firms with related business operations. In theory, various approaches can achieve this. The most straightforward approach is to define economically linked firms as those operating in the same industry. However, this assumes that all firms within an industry are economically linked, which does not necessarily have to be the case. Conversely, firms may also be linked with companies from other industries, such as suppliers or customers that operate in different industries.

As an alternative to industry membership, we could compare past stock returns to identify related firms, assuming that the most similar firms should have the highest return co-movements due to similar risk exposures. Gatev et al. (2006) show that this approach can be used to construct a profitable pairs-trading investment strategy. However, a disadvantage is that firms might randomly co-move and thus appear related even though they are not. While we could mitigate this effect by identifying the most similar firms within an industry, similar to De Franco et al. (2011), we would not be able to detect economic links across industries.

Researchers could also extract competitor, supplier, and customer information from public firms disclosures. For example, Eisdorfer et al. (2021) extract competitors from the business section of US annual reports (10-K filings). However, since firms have some flexibility in disclosing business relations, the resulting business network might lack essential links. Moreover, international annual reports do not share a harmonized structure, which complicates the extraction of competitor names, even if they are reported.

Another method of identifying related business models is introduced by Hoberg and Phillips (2010, 2016, 2025). They establish a time-varying text-based network industry classification for the US (*TNIC*) by comparing the *Item 1* business sections of 10-K filings. However, as most international firms neither provide 10-K filings nor disclose information on their business operations in a separate section, we cannot directly implement this approach internationally.

A feasible alternative for establishing *BNs* could be the use of standardized company descriptions from data vendors. Unfortunately, none of the major vendors could provide historical descriptions on a global scale. As a workaround, we collect historical descriptions for a subset of firms through SDC Platinum, a dataset containing information on stock repurchases, M&As, and recapitalization events. However, this dataset is incomplete and suffers from selection bias because only firms with such corporate events are included.[2] The inclusion of business descriptions from S&P Global as of September 2014, which we receive from the German asset manager *Acatis Investment*, mitigates this problem but does not fully alleviate it.

2.2. AI-generated business descriptions

We, therefore, turn to AI for generating historical business descriptions (*AI-Gen descriptions*). By providing LLMs with unstructured business-related information, we use their text-generation capabilities to describe a company's business model in a standardized way. To do so, we initially source US annual reports from EDGAR and international reports from the Refinitiv filings API provided by LSEG. We prioritize English reports but include non-English ones in case no English reports are available. We then implement an HTML parser to extract *Item 1* sections from 10-K filings. For international reports, we extract the whole text body using the Python package *fitz*, given the lack of a harmonized report structure for international markets. As a next step, we filter the extracted text to remove tables and organize it as a list of sentences.

In case we are unable to extract the Item 1 section from a 10-K filing, we download the PDF version of the US annual report from LSEG as well, if available. This was the case in approximately 10% of all cases. We process the file in the same manner as international reports.

Given that an average report in our dataset contains around 20,000 tokens, we need an appropriate model that can handle this token size as input. A suitable model for this task would be GPT-4-turbo, which is capable of processing up to 128,000 tokens. However, because of the large number of reports to be processed, we decided to limit the number of tokens to accommodate less expensive models like GPT-3.[3] Additionally, as Liu et al. (2023) note, language models with significant input contexts often "lose" information in the middle of a

[2] In written communication with officials from SDC Platinum, we were assured that the business descriptions presented refer to the description valid prior to the corporate event. Thus, for companies covered by SDC Platinum, we have a time series of historical business descriptions. However, this time series has significant temporal gaps for most companies, depending on the number of recorded corporate events.

[3] Using GPT-4-turbo, this would result in costs of around 0.2$ for the model input of one report alone. Given that we deal with hundreds of thousands of reports, this would quickly add up to costs of five-digit figures.

long document. Given the varying information order in international reports, a model with such large input capacity might hence not be ideal anyway.

For token reduction, we focus on sentences in the report that most likely contain business-relevant information. Specifically, we semantically compare all sentences in the reports to sentences extracted from LSEG's 2022 actual business descriptions using text embeddings.[4] We compare English sentences using the sentence transformer model *all-mpnet-base-v2* (Reimers and Gurevych, 2019) and non-English sentences using the *paraphrase-multilingual-mpnet-base-v2* (Reimers and Gurevych, 2020). Sentences that are not sufficiently similar (no sentence pair with cosine similarity above 0.5) to any exemplary business-related sentences are considered irrelevant to our purpose. To maintain the token count within GPT-3's capacity, we select only as many sentences of the business-related sentences with the highest cosine similarity such that the sum of tokens does not surpass the model's token input limit.

Using OpenAI's API, we next instruct GPT-3 to construct business descriptions from this masked business information, focusing on the business model, segments served, and products offered, using only the provided information. The full prompt is as follows:

> Based on the provided information on company X, generate an English business description that describes the main business model, the segments company X operates in and the products company X offers. The description should be written from an outsider's perspective. Do not use other information you may have on the company. The description should not exceed 200 tokens. Just provide the description, do not add further comments.

To better align with the general structure of LSEG's descriptions, we extend the *AI-Gen descriptions* by adding information on the headquarters location and founding year as long this is not already present in the description. Finally, we unmask company and product names in the generated descriptions.

2.3. Measuring firm similarity with embedding models

Choosing a suitable similarity measure is essential for identifying firms with similar business operations. One commonly adopted approach is *bag-of-words* (BOW), where text is encoded as a vector of its constituent words. For example, Hoberg and Phillips (2010, 2016) use BOW by first identifying a set of relevant words and then constructing binary high-dimensional vectors based on the presence of words in the text. By doing so, all word vectors possess a standard dimensionality which facilitates the parallelization of the cosine similarity calculations.[5]

Despite its widespread use, BOW has two main drawbacks. First, measuring document similarity by counting common words may not be accurate, as firms with dissimilar businesses may use the same words in different contexts (e.g., "security" could refer to cyber security, production security, or health security). This issue is exacerbated with short texts, where informative words are typically scarce. Second, BOW does not account for synonyms. For example, firms might describe their business as "selling cars" or "selling automobiles".

Researchers may circumvent these issues by leveraging the latest advances in NLP. These advances were sparked with the invention of the *transformer* (Vaswani et al., 2017). This architecture substantially improved the training speed and performance of deep learning models, leading to the development of several milesone models, notably BERT in 2018 (Devlin et al., 2019), RoBERTa in 2019 (Liu et al., 2019), and GPT-3 in 2020 (Brown et al., 2020). Since OpenAI presented GPT-4 (OpenAI, 2023), a substantially more powerful language model that might even possess "sparks of artificial general intelligence" (Bubeck et al., 2023), numerous other powerful models followed, including one from Meta (Touvron et al., 2023), Google (Reid et al., 2024) and the French startup Mistral AI (Jiang et al., 2024).

Language models can be fine-tuned to assess textual similarity using embeddings, which are high-dimensional vector representations that capture the semantic meaning of text. Unlike word frequency vectors from the bag-of-words approach, where each dimension corresponds to the frequency of individual words, embeddings collectively encode meaning across dimensions using real numbers. This produces a more nuanced and comprehensive representation of linguistic information. Regardless of input size, these models generate fixed-size vectors, enabling comparison through common similarity measures like cosine similarity or Euclidean distance.

For instance, Reimers and Gurevych (2019) introduce Sentence Transformer models, which generate embeddings for sentences or short paragraphs. Of the 38 pre-trained Sentence Transformer models that are currently available as open-source, we use $T5-XXL$ to gauge textual similarity between business descriptions. This model is trained explicitly on sentence similarity and yields the best average performance on a set of 14 diverse sentence similarity tasks.[6]

Besides open-source models, commercial solutions are available. For instance, OpenAI offers pre-trained models for sentence similarity via an API. The *text-embedding-3-small* ($OpenAI-S$) and *text-embedding-3-large* ($OpenAI-L$) models can generate vector representations for up to 8192 tokens. Using these two recently published embedding models, we investigate three state-of-the-art embedding models in this paper. This allows us to investigate if our approach to identifying economic links is robust to different models and to examine whether significant differences exist in the accuracy of open-source and commercial models.

We benchmark these methodologies against the *bag-of-words* approach we construct every year by assembling word-frequency vectors akin to Hoberg and Phillips (2010, 2016). Precisely, upon identifying the most recently available description for all firms, we extract all nouns that occur in at least two different business descriptions. We classify those words as nouns that are recognized as such by the tokenizers of the Python packages *spaCy* and *NLTK*. Finally, we construct word-frequency vectors with the dimension n_y, where n is the number of unique nouns in a given year y. If a noun is present in a description, we assign a value of one and zero otherwise.

2.4. Mitigating a potential look-ahead bias of LLMs

Applying an embedding model, that is based on a large language model, to business descriptions might introduce a look-ahead bias in our study. The reason is that the embeddings obtained from such a model might be influenced by information beyond the era of the historical business descriptions we analyze. For instance, Amazon's evolution from a bookseller to a diverse technology giant could be reflected in the embeddings. This may potentially skew the embeddings of Amazon's 2000 business description to show a heightened cosine similarity with the embeddings of Microsoft's description from the same period, thus creating anachronistic associations.

[4] Note that we do not consider the entire universe of sentences from LSEG's business descriptions, but only a random 10% sample to reduce computation time. All company and product names have been masked using a named entity recognition model from the Python package "spaCy".

[5] In contrast, Cohen et al. (2020) calculate pairwise similarities using a list of relevant words determined on a per-pair basis, resulting in a challenging parallelization problem due to the lack of a fixed vector size.

[6] A disadvantage of $T5-XXL$ is its limit of 256 tokens. We therefore discard excess tokens in a business description. For more information on all available models, their performance, and their primary objectives, see https://www.sbert.net/docs/pretrained_models.html. We also run a performance comparison between $T5-XXL$ and alternative models and provide the results in the Online Appendix.

The extent of this *LLM look-ahead bias* is not clear ex-ante and likely case-specific. In our setting, one may argue that business descriptions, typically phrased in general terms, rarely reference specific, time-bound events. This should lead to a lower exposure to the bias. Nonetheless, there could be research questions in which the *LLM look-ahead bias* becomes significant even when using relatively stable business descriptions.

Eliminating this bias, or at least estimating its magnitude, is challenging. Ideally, we would utilize a language model trained solely on data preceding our evaluation period. However, while the computational and data demands of developing such a custom model would be substantial, this approach would arguably also limit model performance by restricting the amount of available training data.

A more practical method involves masking company-specific identifiers, such as names and product terms, in descriptions before generating embeddings. This tactic limits the model's capacity to link future-relevant company-specific data with historical descriptions. For instance, Glasserman and Lin (2023) use this approach to investigate a *look-ahead bias* in stock return predictions based on sentiment analysis with GPT-3. While other information like industry or macroeconomic trends learned by the model over time could still influence embeddings, the likelihood of such biases affecting anonymized business descriptions seems low. Thus, masking should at least significantly reduce a look-ahead bias, ensuring a more accurate analysis of historical business relations.

In principle, there exist multiple ways to mask company-specific information. For example, Glasserman and Lin (2023) implement an anonymization algorithm that builds on a Google Knowledge Graph. However, we could not retrieve product information for roughly two-thirds of the firms in our dataset.[7] It seems that Google primarily lacks information on smaller and foreign firms, rendering this approach less suitable for us. As alternatives, we could apply named entity recognition (NER) models from the Python package *spaCy* to identify company and product names, or we could instruct general-purpose language models like GPT-3 or GPT-4 to mask company-specific information, but this approach may be costly, depending on the amount of data that has to be masked.

To test the performance of different masking approaches, we instruct GPT-3 to determine the firm name based on the masked business description. If the firm name is not correctly identified, we assume the masking was successful.

Fig. 1 provides an overview of the accuracies of different masking strategies. We find that the smaller spaCy NER model performs worst. The share of successful masking attempts is only 87.25%, which is in line with the observation of Glasserman and Lin (2023) that the model overlooks many entities. However, we find that the most potent transformer-based NER model (*en_core_web_trf*) achieves a substantially better performance. Here, the masking attempt fails in only 0.36% of the business descriptions, thus successfully masking company-specific information in 99.64% of the cases. This methodology even outperforms GPT-3 which yields a successful masking rate of 97.27%. Based on these findings, we decide to rely on the transformer-based NER model from spaCy to mask company-specific information, as it yields the best performance and does not induce API costs.

2.5. Defining business networks

After creating the AI-based business descriptions and applying the various embedding models to gauge textual similarity, we need to formalize the construction of the (global) *BNs*. This entails making two critical decisions. First, we must determine which description to use if historical business descriptions from commercial data providers and *AI-Gen descriptions* are available for a specific firm-year. Here, we decide to primarily use the AI-based descriptions and resort to the actual descriptions (primarily S&P Global and, secondarily, SDC Platinum) in the absence of *AI-Gen descriptions*. Considering that the descriptions of S&P Global are not available before September 2014 and that the SDC Platinum dataset has a selection bias, this prioritization seems sensible to us. To assess the quality of the *AI-Gen descriptions* and the resulting *BNs*, we also repeat critical analyses of our study excluding the descriptions of the commercial data providers, which can be found in the Online Appendix.

Second, we need to establish a threshold for identifying sufficiently similar business descriptions. Here, we follow Hoberg and Phillips (2016) and select a percentile threshold, as this also controls for differences in the cosine similarity distributions of different embedding models. Specifically, firms are deemed economically linked if their business description similarity ranks within the top 1% in a given model's cosine similarity matrix. As a consequence, some firms may reveal a lack of sufficiently similar companies. The concerned firms are excluded from our *BNs* in such instances. To ensure the robustness of our findings, we repeat key analyses using alternative thresholds to construct the *BNs*, which rely on the top 5% and top 0.1% of the similarity rankings. The results of these robustness tests are also shown in the Online Appendix.

Finally, due to the absence of an established global *BN* for benchmarking, we also construct US-only networks in a similar manner, allowing us to benchmark our results against TNIC (Hoberg and Phillips, 2010, 2016).

3. Data

3.1. Global stock sample

We gather stock market and accounting data from CRSP and Compustat for the US, and from LSEG (formerly Refinitiv) Datastream and Worldscope for non-US, i.e., international markets. Our global stock sample is confined to common equity from countries included in one of the major MSCI regional indices as of June 2021 (namely, MSCI North America, Europe, Pacific, Emerging Markets, and Frontier Markets). It covers the period from January 2000 to December 2021.

To eliminate any non-common equity, we apply several filters. In the US, we focus exclusively on NYSE, AMEX, and NASDAQ-listed stocks with CRSP share codes of 10 or 11. For other countries, where data comes from Datastream/Worldscope, we (i) select only securities that are classified as common equity, (ii) focus on the firm's major security if it has multiple securities, (iii) consider only primary exchange listings, and (iv) require all stocks to have a valid Worldscope identifier to further exclude non-common equity securities, such as ADRs. Furthermore, to avoid a survivorship bias, we select both "active" and "dead" equities from Datastream. In total, this selection strategy yields a comprehensive view of the global equity landscape consisting of 68,402 stocks with 9.28 million stock-month observations across 67 countries. 56,105 stocks with 8.13 million stock-months are from international markets, and 12,297 stocks with 1.15 million stock-months are from the US.

For US stocks, we use returns from CRSP, and for international stocks we use Datastream's total return index, including dividends (data variable: *RI*), to calculate monthly stock returns in US dollar. We clean stock data from LSEG as recommended in the literature (Griffin et al., 2010; Ince and Porter, 2006; Jacobs and Müller, 2020). Most notably, we winsorize returns at the 0.1% and 99.9% level to account for the presence of few data outliers, and we include delisted stocks in our analysis only up to the point of their actual delisting by using the methodology of Ince and Porter (2006) to detect stale prices.

To add potentially missing information about the founding year and headquarters location to the AI descriptions, we also use Datastream/Worldscope. The founding year is sourced from the foundation

[7] We randomly selected 1000 firms, queried Google's Knowledge Graph API and obtained information for only 34.4% of the firms.

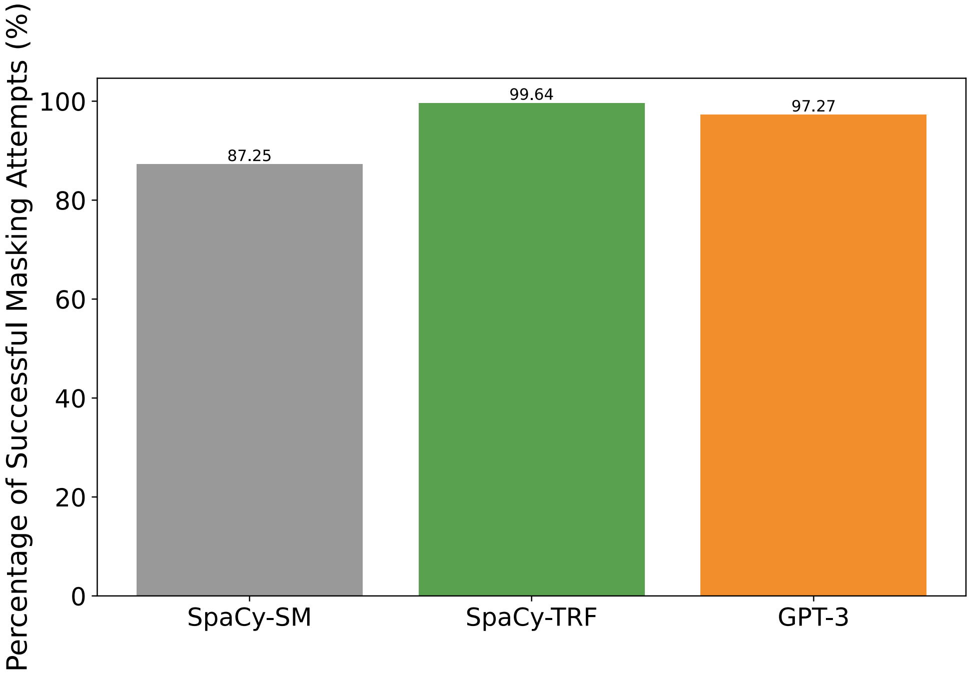


Fig. 1. Quality assessment of different masking techniques. This figure examines the accuracy of different masking methodologies. We mask the company and product names in the most recently available AI-generated business description for each firm. We then instruct GPT-3 to identify the company name given the masked description using the following prompt: "You will receive a business description. If you know to which company the description belongs, return the company name, otherwise return 'Unknown'". We rate a masking attempt as successful, if GPT-3 was unable to identify the correct company name.

or incorporation date (variables *WC18272* and *WC18273*) primarily, and the listing date (variable *BDate*) when the former are unavailable. The headquarters location is retrieved from data item *WC06023*.

We match annual financial reports to our sample firms by accessing SEC's EDGAR service for US stocks and the Refinitiv Filings API from LSEG for international stocks and those US stocks, for which we could not parse the Item 1 section. Data on M&A deals of US and international firms from 2001 until 2022 is obtained from SDC Platinum, which is also maintained by LSEG. We restrict our sample to deals in which publicly traded acquirers purchase publicly traded targets.

3.2. Commercial business descriptions

We collect global business descriptions for our stock sample from LSEG at the beginning of August 2022. The descriptions are written in English and contain 300 words or less. Although these brief descriptions are less comprehensive compared to the Item 1 section in 10-K filings, they cover the critical aspects of a firm's business model. The first sentence of the business description typically describes the core business model of a company, followed by an enumeration of the company's segments in the following sentences and a list of products offered. The description typically closes with information on the company's headquarters and founding year.

A boxplot is provided in Fig. 2, illustrating the 90% confidence interval for token counts[8] across different NYSE size deciles to guarantee adequate business description lengths for smaller firms. On average, the largest firms exhibit marginally higher character counts than those in the lowest decile. Nonetheless, over 95% of descriptions for firms in the lowest decile exceed fifty tokens, suggesting a representative coverage.

The LSEG business descriptions are stored in Worldscope data item *WC06092*. This is a so-called "static" variable, that is regularly overwritten. For this reason, we cannot use the descriptions directly to construct time-varying *BNs*, even though they prove to be a helpful template for GPT-3 to generate historical business descriptions from the annual reports, as described in Section 2.2.

In contrast, the business descriptions which we obtain from SDC Platinum and S&P Global are historical and partially time-varying. We can identify 51,129 companies in our sample that have at least one corporate event in SDC Platinum. We maintain the descriptions from SDC Platinum in the dataset until there is a new record with an updated description. For S&P Global, we are able to match 24,664 business descriptions to our sample. These descriptions refer to September 2014, and we carry them forward until the end of our sample period.

3.3. AI-generated business descriptions

AI-Gen descriptions are available for 47,853 firms (433,744 firm-years) in our sample. To evaluate the quality of these descriptions anecdotally, Table 1 shows the three most recent examples for automotive manufacturers *Ford Motor*, *Volkswagen*, and *Toyota Motor*. While the AI description for Ford uses Item 1 of the company's 10-K filing, the descriptions for Volkswagen and Toyota are based on annual reports obtained from LSEG.

The examples in Table 1 convey the impression that the descriptions adequately cover the companies' business models. They highlight the various business segments and, in some cases like *Volkswagen*, also address sales markets. While all descriptions list product categories, specific products or brand names are explicitly mentioned only in the description for *Toyota Motor*. The likely reason is that we mask product names in the information provided to GPT-3 to prevent the model from mapping additional information via the product name.

Overall, the output meets our prompt's specifications by containing information on the business model, the segments served, and the product categories offered. This was also the case for most other descriptions, which we manually reviewed. Furthermore, regarding structure and content, the AI-based descriptions seem remarkably similar to the actual descriptions from LSEG.

Fig. 3 illustrates the distribution of the number of tokens contained in *AI-Gen descriptions* across different NYSE size deciles. The boxplot

[8] We tokenize the descriptions with the cl100k_base tokenizer included in the Python package *tiktoken*.

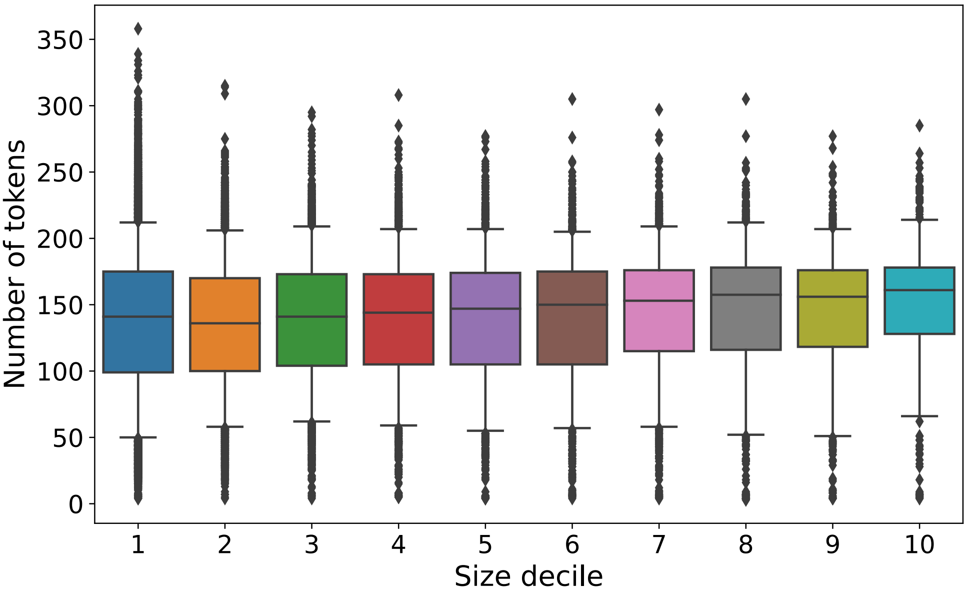


Fig. 2. Business description length (LSEG Refinitiv). This figure shows the distribution of the number of tokens in a business description from LSEG (Worldscope item *WC06092*) across different NYSE size deciles. Business descriptions are obtained in August 2022.

Table 1
Exemplary AI business descriptions as of 2021.

Ford Motor	
EDGAR	FORD MOTOR is a global automotive company that operates in various segments of the industry. The company is committed to sustainable development goals and has a strong focus on environmental responsibility. FORD MOTOR offers a range of products, including automotive components such as radiators, wiring, circuitry, and printed circuit boards. The company also provides management tools for fleet managers to analyze their vehicles and make informed decisions. FORD MOTOR is dedicated to improving fuel economy and reducing carbon emissions through the development of electric vehicles and hydrogen fuel cell vehicles. The company has a diverse supply chain that includes component suppliers and indirect suppliers of facilities, equipment, materials, and services. Additionally, FORD MOTOR is invested in creating a customer-centric experience and has a loyalty rewards program for its customers. The company prioritizes sustainability in its manufacturing processes by investing in state-of-the-art facilities and emerging technologies. Overall, FORD MOTOR aims to have a positive influence on the future by providing innovative and environmentally friendly mobility solutions. The company was founded in 1903. The company is based in DEARBORN, USA.
Volkswagen	
Filings API	VOLKSWAGEN is a global company operating in various segments, including dealer and customer financing, leasing, banking, insurance, fleet management, and mobility services. In the automotive segment, the company is involved in the development, production, and distribution of cars, motorcycles, and commercial vehicles, as well as the sale of original parts and components. The company's product portfolio ranges from small cars to luxury vehicles. The company has a strong presence in Asia-Pacific, Europe, South America, and Africa. The company also offers financial services such as dealer and customer financing, vehicle leasing, and insurance. The company has a diverse production network with numerous manufacturing sites worldwide. VOLKSWAGEN aims to promote sustainable growth and shareholder value through continuous innovation and investment in technology and cybersecurity. The company is also committed to advancing electric mobility and developing software and interfaces for electric vehicles. The company was founded in 1937. The company is based in WOLFSBURG, Germany.
Toyota Motor	
Filings API	TOYOTA MOTOR is a mid-size company that specializes in passenger vehicle planning, development, and production. The company offers a wide range of vehicles, including the Corolla, Coaster, Land Cruiser, Alphard, and more. The company's core technology, TOYOTA MOTOR, is used in all electrified vehicles and greatly impacts vehicle performance. TOYOTA MOTOR also provides after-sales services, such as insurance and maintenance, to ensure customer satisfaction. The company is actively involved in the development and provision of connected devices and services. Additionally, TOYOTA MOTOR collaborates with other companies to create innovative and differentiated products. The company is committed to improving fuel efficiency and reducing CO2 emissions. TOYOTA MOTOR also offers financial services to enable more customers to use their cars. The company is dedicated to building sustainable relationships with stakeholders and contributing to sustainable development. The company was founded in 1937. The company is based in TOYOTA-SHI, Japan.

This table contains the most recent *AI-Gen descriptions* of *Ford Motor*, *Volkswagen* and *Toyota Motor*, as well as information from which source the annual report used for the data generation process has been collected from.

generally does not reveal meaningful differences in the token counts across size buckets, with the exception that the AI descriptions are somewhat shorter for stocks in the lowest NYSE size decile. However, even for stocks in the smallest NYSE size decile, over 95% of *AI-Gen descriptions* exceed 170 tokens, ensuring sufficient information for large-cap stocks and the majority of small stocks.

Next, we conduct an empirical analysis to assess the quality of the AI-based descriptions. We determine the cosine similarity between the $T5-XXL$ embeddings of the actual descriptions from commercial providers and the AI descriptions for the same year. For example, we compare the 2014 business descriptions from S&P Global with the 2014 *AI-Gen descriptions*.[9] Fig. 4 presents the median cosine similarity between *AI-Gen* and commercial business descriptions (LSEG, S&P Global, and SDC Platinum).

[9] We compare *AI-Gen descriptions* from 2021 with LSEG descriptions from 2022, as we lack access to LSEG's 2021 descriptions.

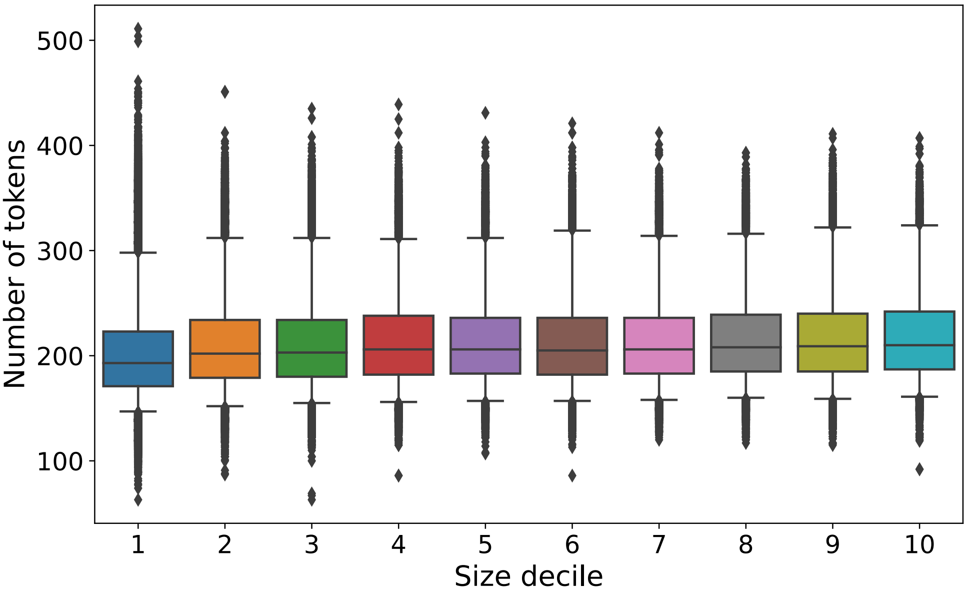


Fig. 3. Business description length (AI descriptions). This figure shows the distribution of the number of tokens in *AI-Gen descriptions* across different NYSE size deciles.

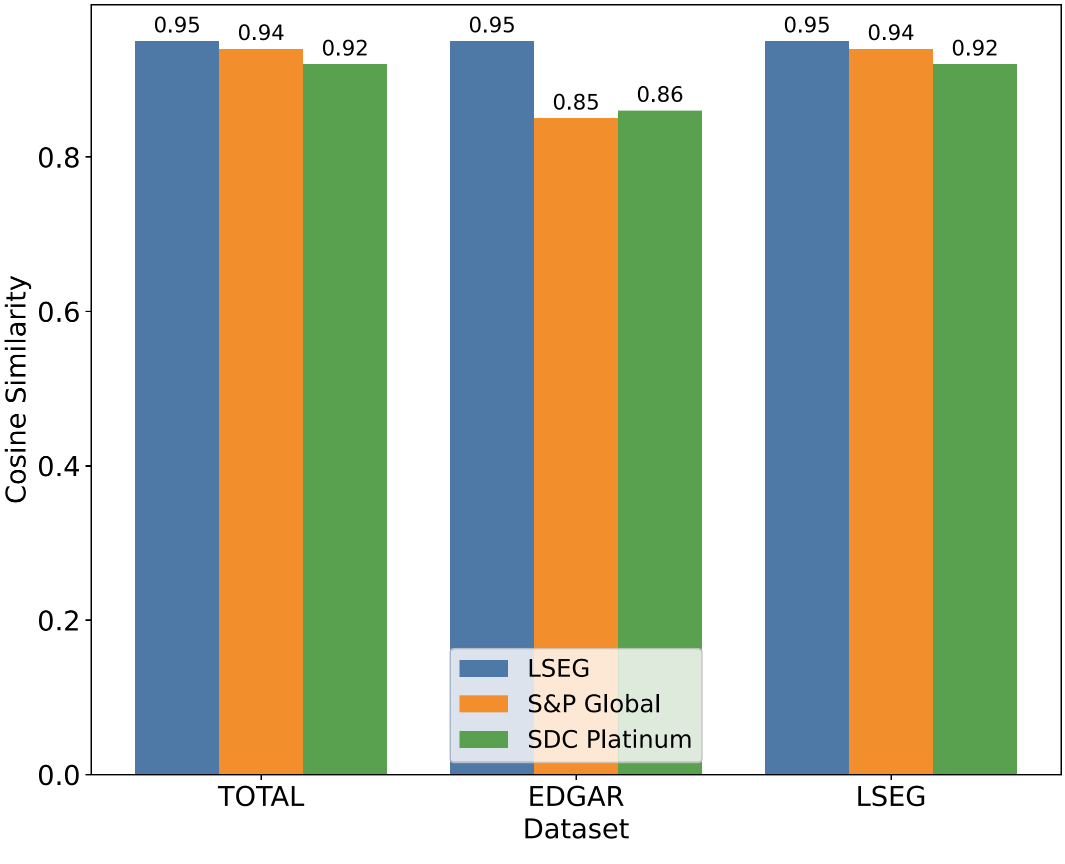


Fig. 4. Comparison of AI descriptions to the actual descriptions from data vendors. This figure shows the median cosine similarity of *AI-Gen descriptions* to the descriptions from LSEG (blue bars), S&P Global (orange bars), and SDC Platinum (green bars). We generate AI descriptions with GPT-3 using information from annual reports. For each company, we calculate the similarity between the LSEG (S&P Global, SDC Platinum) description and the *AI-Gen description* of the same firm on the basis of $T5-XXL$. We consider the AI description of the year the commercial description was generated. We display results for three subgroups (*EDGAR*, *LSEG*, and *TOTAL*, i.e., *EDGAR* + *LSEG*) to further investigate potential differences between the accuracy of descriptions constructed from US and international annual reports.

Overall, this figure indicates that *AI-Gen descriptions* are highly similar to commercial business descriptions. The median cosine similarity between a firm's AI business description and its actual LSEG description is 0.95. This indicates a high textual overlap, given that cosine similarities range between −1 and 1, and the upper limit implies a perfect semantic text match. The median cosine similarities of the AI description to the actual descriptions of S&P Global and SDC Platinum, respectively, are slightly lower at 0.94 and 0.92.

It might be more challenging for GPT-3 to generate business descriptions from unstructured annual reports as they might be less

Table 2
Availability of business descriptions over time.

Year	#Stocks	USA (CRSP)				#Stocks	Non-US (LSEG Refinitiv)			
		#AI	AI-MV	#All	All-MV		#AI	AI-MV	#All	All-MV
2000	7553	65.35	79.27	82.85	91.61	21 298	22.16	55.03	45.35	79.85
2001	7178	69.46	83.06	86.01	93.64	23 028	34.51	67.91	53.73	83.80
2002	6447	76.11	87.59	89.65	96.12	23 785	43.09	73.28	61.56	87.92
2003	5762	83.48	91.17	92.33	97.32	24 527	48.45	76.53	67.65	90.09
2004	5358	86.95	92.56	93.88	97.77	24 829	50.82	78.70	71.13	92.03
2005	5172	88.65	93.48	94.43	97.92	25 861	53.12	79.38	74.33	92.83
2006	5114	90.05	94.05	95.03	98.08	27 666	54.76	78.87	77.38	93.38
2007	5027	91.23	94.35	95.90	98.45	29 286	56.85	81.32	80.91	94.50
2008	4999	91.74	94.66	95.94	98.61	30 915	59.13	81.05	84.10	95.17
2009	4714	93.64	95.29	96.86	98.88	31 858	61.26	82.76	86.36	95.66
2010	4431	94.65	94.82	97.36	98.88	31 984	61.66	82.25	87.27	95.76
2011	4237	95.80	94.93	98.25	98.94	32 784	61.72	83.16	87.76	96.09
2012	4067	95.97	95.35	98.33	98.93	33 539	62.55	82.95	88.75	96.03
2013	3914	96.17	95.27	98.36	99.09	33 633	63.20	84.19	89.73	96.58
2014	3879	96.29	94.62	98.50	99.09	33 484	64.21	84.38	97.20	99.25
2015	3963	96.52	95.86	98.44	98.32	33 885	65.00	82.23	97.49	99.33
2016	3981	96.66	97.55	98.44	99.33	34 615	64.89	77.88	97.04	99.09
2017	3880	96.75	97.72	98.32	99.25	35 018	65.10	80.02	96.90	98.88
2018	3819	96.83	98.21	98.25	99.38	36 024	64.86	81.55	96.46	98.76
2019	3809	96.95	99.43	98.37	99.83	36 786	65.07	83.89	96.49	98.74
2020	3803	96.98	99.61	98.29	99.86	37 116	65.10	83.19	96.59	98.71
2021	3911	95.45	99.68	96.86	99.83	37 466	64.67	80.00	96.39	98.26

This table outlines the extent of coverage (in %) provided by our dataset of *AI-Gen descriptions*, alongside an expanded version that incorporates commercial descriptions from S&P Global and SDC Platinum. Coverage is assessed annually by identifying the most recent *AI-Gen description*. For the US, we relate the number of stocks with an AI description to the number of all US stocks covered by CRSP at any of the major stock indices NYSE, NASDAQ and AMEX in the previous year (#AI). We also report coverage using all available historical descriptions including AI and commercial descriptions (#All). The one-year lag is included to control for new listings where we lack annual reports to construct AI business descriptions from. For stocks outside the US, we compare against the total number of actively traded stocks in the previous year as reported by the data provider LSEG Refinitiv. We also calculate the relative market capitalization coverage in a similar manner (AI-MV and All-MV).

informative than the *Item 1* sections of 10-K filings. Addressing this concern, we also analyze the textual overlap separately for US firms relying on 10-K filings from EDGAR and AI descriptions that are derived from annual reports. Measuring the textual match with the actual LSEG business descriptions, we do not observe a difference in the median cosine similarity, which is 0.95 for both firm groups. Comparable values for both groups are also observed if we measure the similarity of the AI text to the descriptions from SDC Platinum and S&P Global. Overall, we conclude that the *AI-Gen descriptions* are qualitatively well-suited for constructing global business networks.

3.4. Availability of business descriptions

Our business description dataset contains 514,389 *AI-Gen* and commercial business descriptions. For 63,486 out of the 68,402 firms in our stock return dataset we have at least one business description, representing a coverage of 92.8%. To better understand the availability of business descriptions across the dataset, we examine the number and the equally weighted and value-weighted coverage per year from 2000 to 2021. We report coverage separately for US stocks (from CRSP) and international stocks (from LSEG) (see Table 2).

In 2000, our dataset covers 65.35% of all common US stocks, translating to a value-weighted coverage of 79.27%. Over time, both the equally weighted and value-weighted shares of covered US stocks increase. By 2006, our dataset includes AI descriptions for 90.05% of US stocks, representing 94.05% of total US market capitalization. By 2021, over 95% of all US stocks are covered, accounting for more than 99% of the total market capitalization.

Outside the US, AI descriptions are available for 22.16% of firms in 2000, covering 55.03% of the non-US market capitalization. The lower initial international coverage results from the prevalence of scanned, non-electronically readable annual reports in earlier years.[10]

As electronically readable reports become more prevalent, international coverage improves. By 2003, the dataset covers over 76.53% of the international market capitalization, and the equally weighted share of covered non-US stocks increases to 48.45%. From 2004 onwards, AI descriptions are generated for more than 50% of international stocks, reaching an equally weighted (value-weighted) coverage of 65% (over 82%) by 2015, and remaining stable in subsequent years. The higher value-weighted coverage indicates that AI descriptions are primarily missing for smaller stocks.

In summary, our method allows us to successfully create AI business descriptions for a large number of stocks across the globe. However, if we solely relied on *AI-Gen descriptions*, we would miss a meaningful fraction of stocks in the global *BNs*, particularly for international stocks in the early years of the sample period. Therefore, we also consider the historical descriptions from *S&P Global* and *SDC Platinum* for those firm years without an AI description. We carefully assign only those descriptions to firm-year observations available at a particular time. As a result, we observe a substantial increase in coverage in the US and internationally.

Internationally, the inclusion of *S&P Global* and *SDC Platinum* descriptions results in an approximate 23 percentage point increase in stock coverage in 2000. This extension leads to a value-weighted international coverage of 79.9%. Since 2003, our dataset consistently covers over 90% of the market capitalization of non-US stocks. Moreover, equally and value-weighted coverage increases over the years to 96.4% and 98.3% in 2021. The notable surge in coverage in 2014 can be largely attributed to the incorporation of S&P Global descriptions from that year.

The higher value-weighted coverage in comparison to the equally-weighted coverage suggests that our dataset still lacks business descriptions for smaller stocks, especially at the beginning of our sample period, which should be considered by researchers using our data. Nevertheless, we conclude that by utilizing all available business descriptions, we can create comprehensive time-varying business networks that include most global stocks.

We next investigate coverage across different countries in our AI and full dataset of business descriptions. Table 3 displays the average

[10] To ensure high input quality for the business descriptions, we avoid using OCR tools due to frequent errors in scanned annual reports.

Table 3
Average availability of business descriptions across countries.

Country	#	AI	All	AI_{MV}	All_{MV}	Country	#	AI	All	AI_{MV}	All_{MV}
Argentina	80	80.6	89.2	79.1	96.6	Malaysia	918	96	98.4	97	99.1
Australia	1607	89.6	93.4	89.3	93.6	Mauritius	42	56	59.4	56.8	59.7
Austria	92	89.5	91.9	97.5	98	Mexico	133	77.9	85.2	80.1	82.9
Bahrain	40	81.4	87.8	94.9	96.9	Morocco	64	77.2	84.8	89	95.1
Bangladesh	94	54.8	58.4	62.7	64.2	Netherlands	146	90.5	95.7	93.6	99.3
Belgium	146	86.8	90	97.1	98.4	New Zealand	128	92.1	96.7	94.9	97.3
Brazil	166	83.2	87.5	91	93.6	Nigeria	144	68.2	80.6	79.7	95.9
Bulgaria	153	16.5	49.3	26.7	73.1	Norway	224	81.3	92.8	89.6	95.3
Canada	2750	42.1	81.9	68.6	93	Oman	103	92.1	92.9	93.8	94.2
Chile	182	86.2	88.3	94.6	97.4	Pakistan	335	85.2	86.9	85.9	93.4
China	2229	10.1	76.8	33.1	82.3	Peru	132	72.8	79.7	74.3	90.4
Colombia	57	57.9	71.8	68.7	84.3	Philippines	232	88	93.8	93.1	98.1
Croatia	98	41.3	67.7	80.2	90.8	Poland	410	82.6	90.3	94.1	97.5
Czech Rep.	31	66.5	78.8	96.4	97.6	Portugal	64	87.1	89.2	94.1	97.4
Denmark	180	93.5	95.9	97.1	98.1	Qatar	41	94.5	96.2	94.8	97.3
Egypt	178	16.3	66.9	56.3	86.5	Romania	111	24	43.6	77.3	85.7
Estonia	14	98.1	98.1	99.8	99.8	Russia	266	50.9	76.4	93.3	98.1
Finland	142	95.9	97.6	99.6	99.7	Serbia	58	23.2	42.3	51.8	71
France	866	81.1	88.4	95.9	98.9	Singapore	640	89.5	95.1	94	97.9
Germany	873	82.4	86.5	96.3	97.8	Slovenia	17	79.9	97.8	84.5	99.5
Greece	255	53.1	78.9	85.3	94.4	South Africa	359	80.3	88.5	84.2	94.6
Hong Kong	1275	92.2	96.5	80.5	85.7	Spain	179	45.9	84.1	64.4	96.8
Hungary	43	50.1	77.2	93.8	98.5	Sri Lanka	218	77.8	82.9	79.6	87.3
India	2540	80.2	84.8	90.1	94.8	Sweden	485	91.9	95.2	96.8	99.3
Indonesia	421	80.4	85.8	87.6	95.3	Switzerland	252	92.8	95.8	86.3	93
Ireland	60	80	89.2	74.9	86.1	Taiwan	1472	8.4	51.8	50	84.2
Italy	295	88.5	95.7	91.2	97.5	Thailand	556	78	88.4	76.7	95.8
Japan	3738	18.1	79.7	77.8	95.4	Tunisia	51	96	97	95.3	95.5
Jordan	210	13.5	69.7	71	91.5	Turkey	317	96.4	97.3	97.4	98.8
Kazakhstan	30	80.9	88.7	90.5	95.7	USA	4774	90.5	95.5	94	98.1
Kenya	51	60	72.2	77.2	86	Ukraine	60	28.6	68.2	63.1	82.6
Korea	1707	4.1	65.2	44.4	89.4	UK	1712	87.1	93.4	92.7	95.9
Kuwait	164	34.3	73.8	62.7	85.7	Vietnam	788	70.2	83.1	73.5	95.2
Lithuania	33	75.3	88	81.2	89.1						

This table shows the average number of publicly traded companies per country between 2000 and 2021 (#). We further provide the average share (in %) of those firms included in our AI (*AI*) and full business description dataset (*All*), which supplements AI descriptions with actual historical descriptions from *S&P Global* and *SDC Platinum*. We also provide the average market value coverage per country over the same time horizon for our AI (*AI-MV*) and full description dataset (*All-MV*).

coverage at the country level between 2000 and 2021.

According to Table 3, our AI dataset encompasses over 80% of stocks in most countries. The value-weighted coverage is even higher and often exceeds 90%. However, certain countries are substantially underrepresented. For instance, only 10% of Chinese firms are included in our AI business description dataset, covering roughly 33% of the Chinese stock market. This shortfall is attributed to the limitations of the GPT-3 tokenizer in processing Chinese text. A similar underrepresentation is observed for other Asian countries such as Japan and South Korea.

This issue can be mitigated by incorporating commercial descriptions from S&P Global and SDC Platinum. By doing so, coverage in China increases by more than 65 percentage points to 76.8%. In Japan, coverage rises to 79.7%, and in South Korea, it improves to 65.2%.

Nevertheless, smaller countries such as Romania (43.59%), Taiwan (51.80%), Bangladesh (58.35%), and Mauritius (59.45%) continue to exhibit underrepresentation, even when external business descriptions are additionally considered. This can be attributed to the lower availability of annual reports, language barriers, and fewer corporate events, particularly M&A deals, recorded in SDC Platinum. Despite these gaps, our dataset provides a good to very good coverage for the vast majority of global stock markets.

4. Evaluation of the business networks

4.1. Anecdotal evidence

In this section, we evaluate the global *BNs*, starting with some anecdotal evidence. For this purpose, we report the five most similar firms to the car manufacturer Ford according to various networks as of 2021 in Table 4. To ensure that the companies are known to the readers, we restrict the lists to stocks in NYSE size decile eight or above.

In Panel A, we focus on US competitors as identified by TNIC (Hoberg and Phillips, 2010, 2016), the open-source model ($T5-XXL$), and the large embedding model from OpenAI ($OpenAI-L$). According to *TNIC*, *General Motors* is the most similar firm, a result also observed in $T5-XXL$ and $OpenAI-L$ networks.

In Panel B, considering global firm relations, BOW, $T5-XXL$, and $OpenAI-L$ also list *General Motors* as the most similar firm. For the two context-aware embedding models, all other similar firms also operate in the automotive sector. $T5-XXL$ lists *Great Wall Motor Co.*, *Maruti Suzuki India*, *Toyota Motor*, and *SAIC Motor*. $OpenAI-L$ includes *Honda Motor*, *Hyundai Motor*, *Nissan Motor*, and *Toyota Motor* among the five most similar firms. BOW, besides *General Motors*, lists only one other automotive company, *Great Wall Motor Co.*, among the most similar firms.

To assess the stability of business relations over time, we display the trajectories of textual similarity measured by the open-source model ($T5-XXL$) for selected global competitors of *Ford* in Fig. 5.

We find relative stability for most cosine similarities between *Ford's* business descriptions and those of its competitors. For example, *General Motors* consistently had the highest similarity since 2010. Similarities are also relatively stable and high for other peers like *Volkswagen*, *Daimler*, and *Hyundai*. For *Tesla*, the cosine similarities are slightly lower, fitting the expectation that its business operations differ more from *Ford's*.

To understand how our global *BNs* are populated by domestic and foreign firms, we calculate the share of foreign relations (*foreign-share*). This is the number of foreign firms in the network divided by the total number of business relations for a given firm-year. We then calculate

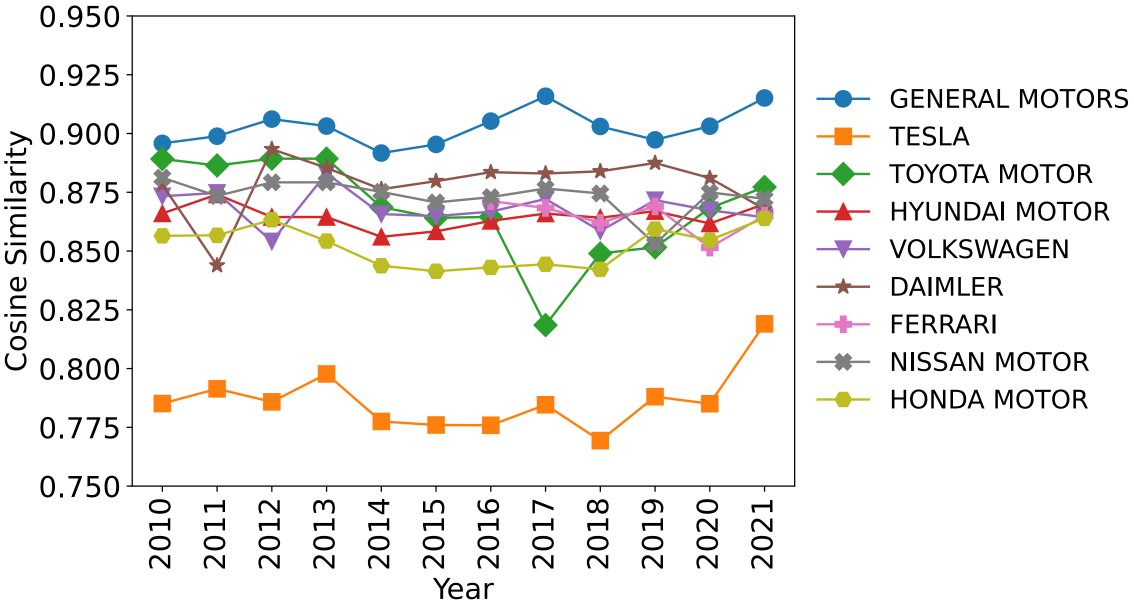


Fig. 5. Textual similarity to Ford Motor for selected competitors over time. This figure shows the similarity of the business description of *Ford* with the descriptions of other car producers over time. We display the similarity based on the $T5 - XXL$ embeddings.

the median foreign share based on the $OpenAI - L$ network for 23 countries and globally aggregated for the starting (2000) and ending year (2021) of our sample and provide the results in Fig. 6. This figure highlights the importance of a global perspective on business networks and shows significant cross-country differences in *foreign-share*. Smaller countries like Denmark, Switzerland, and Singapore have the highest *foreign-share*, ranging between 75% and 80% in 2000, due to their relatively low number of domestic firms. In contrast, larger economies like the United States, Japan, and Germany show substantially lower *foreign-share* values. Over time, there is a general increase in the number of foreign firms in business networks. This trend is observed globally, with *foreign-share* rising from 39% in 2000 to 49% in 2021. The US also sees a significant increase, with *foreign-share* rising by 38 percentage points from 23% in 2000 to 61% in 2021.

4.2. Performance evaluation

4.2.1. Network summary and similarities

We aim to comprehensively and systematically understand the differences between the constructed *BNs* across multiple dimensions, including industry and country congruence, based on the hypothesis that business networks should be more homogeneous in these regards. We also quantify to what extent networks overlap and correlate and provide the results in Table 5.

Panel A of reports the average and median number of firm relations across different global business networks. To enhance interpretability, we also include a network with random firm pairs ("Random"). The average values show minimal variation, ranging from 398 to 404, but the median values reveal more pronounced differences. The *BOW* network has a median of 188, compared to 281 for the smaller *OpenAI* network, indicating fewer firms with disproportionately many relations in the latter. This variation is due to differences in the concentration of high cosine similarities across networks.

Panel A also shows that context-aware networks, compared to *BOW*, have more relations with domestic firms. For instance, 40.22% of *BOW*'s relations are domestic, while the larger *OpenAI* model has 54.71%. A similar trend is observed concerning industry membership, measured by SIC codes. A random pair of firms shares the same four-digit SIC code in only 0.86% of cases, while this share is 5.93% for the word-based network. However, context-aware networks yield even higher shares, between 10.12% and 13.13%. This suggests that context-aware networks better identify relevant business relations than bag-of-words.

In Panel B, we analyze the overlap percentage, which is the share of relations occurring in two networks. The average overlap between a network of randomly selected firm pairs and the other networks is below 2%, serving as a baseline. This is substantially smaller than the 15% to 17% overlap between word-based and context-aware networks. The overlap between the two OpenAI models is significantly higher (around 59%), and there is also a notable overlap of around 40% to 46% between $T5 - XXL$ and the OpenAI models. This indicates that word-based networks differ substantially from context-aware networks.

To further assess the similarity of the different business networks, we correlate the stock returns of peers from different networks. For each firm and model, we calculate the average return of all peers in its global *BN*. We then correlate the average peer returns across networks and report the average firm-level correlation in Panel C. While correlations are about 0.6 for *BOW* and the embedding models, return correlations among context-aware networks, particularly between the two OpenAI models, are markedly higher, reaching up to 0.85. This suggests that context-aware business networks share more economically similar peers.

4.2.2. Identification of US competition links

We next validate our business networks by assessing how many "actual" competitors (as marked in alternative datasets) are included in our business networks. Due to the lack of a global competition dataset, we evaluate US-only *BNs*, allowing us also a comparison with the TNIC dataset.

First, we benchmark against the list of competitors disclosed by firms themselves, adopting the approach of Eisdorfer et al. (2021) and

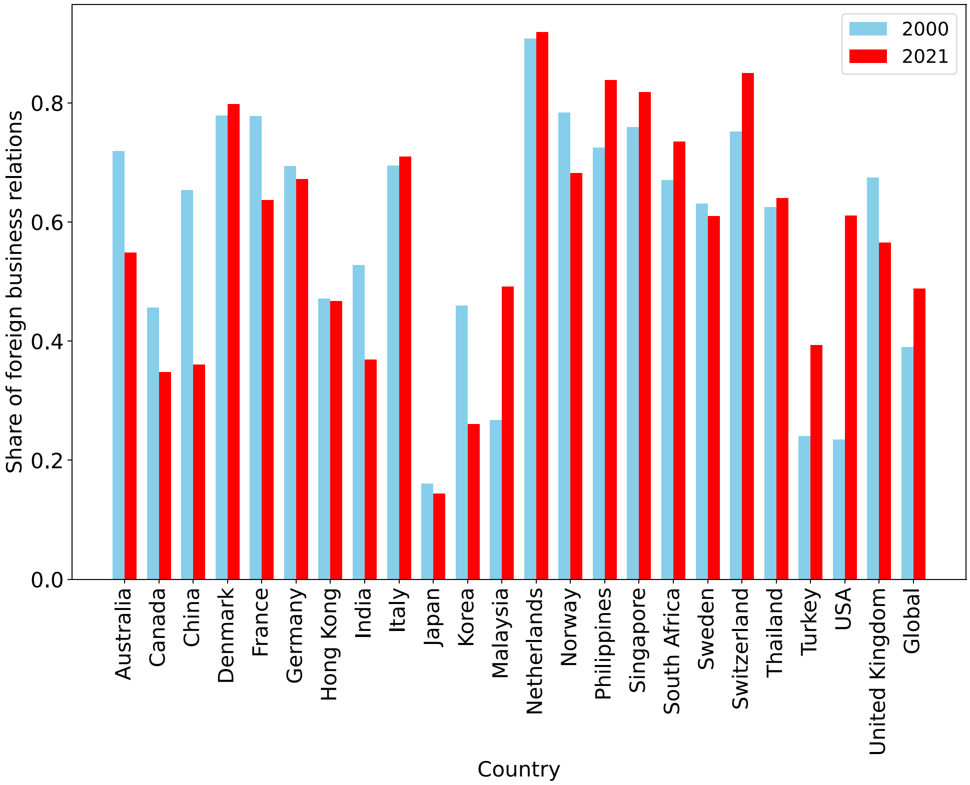


Fig. 6. Average share of foreign relations per country. The figure shows the median *foreign-share* in our global *BNs* at the country-level. We present these values for 23 countries and globally aggregated (the two right-most bars) for the years 2000 and 2021. Global *BNs* consist of firms with a business description similarity ranking in the top 1% of all similarities, according to the large embedding model from OpenAI ($OpenAI - L$). We calculate *foreign-share* as the number of foreign firms in the network divided by the total number of business relations for a given firm-year. We only consider firms with at least 30 business relations and countries with at least 100 firms fulfilling our criteria.

extract competitor names from the Item 1 "competition" subsection of US 10-K filings.[11] Second, following Guo et al. (2023), we extract economic peers from merger and acquisition filings submitted to the SEC in 2022, focusing on the "opinion of the financial advisor" section of M&A documents like PREM14 A, DEFM14 A, and S-4/A. Third, we compare our network against US firm competition relations included in *FactSet Revere*.[12]

Table 6 summarizes the results of these validation tests by presenting the recall scores for various business networks, each fixed at a distinct count of most similar firms (10, 50, 100), in the context of competitor identification.

We find that the $TNIC$ dataset, when restricted to the top 100 most similar firm relations, identifies 52.60% of the competitor relations reported in 10-K filings (Panel A). In contrast, a word-based network based on our business descriptions has a notably lower recall score of 32.83%. A network using the larger embedding model from OpenAI significantly surpasses the BOW but underperforms TNIC by around five percentage points. The $T5 - XXL$ network falls between the two OpenAI models.

We replicate the analysis with networks based on masked business descriptions to determine if the superior performance of LLM-based embedding models is due to potential biases. This approach unexpectedly raises recall scores across all embedding models. For networks containing the 100 most similar firms, the recall score of the $T5-XXL$ network increases by approximately three percentage points, while the OpenAI models show increases of up to twenty-one and seven percentage points, respectively. These results suggest that the presumed upward bias associated with the look-ahead character of LLMs does not exist in this context.

Extended analyses using competitor lists from merger and acquisition filings (Panel B) and FactSet Revere (Panel C) corroborates the initial findings. Embedding models from OpenAI consistently outperform the word-based approach. Notably, using masked business descriptions, higher recall scores are observed for the *OpenAI* networks. This highlights the importance of text anonymization in such applications.

In conclusion, the validation tests suggest that commercial embedding models from OpenAI demonstrate superiority over traditional word-based methods, underscoring the robustness and quality of the business networks generated by these advanced models.

5. Applying the business networks

5.1. The lead-lag effect

The well-documented lead–lag effect (see e.g., Hou, 2007; Cohen and Frazzini, 2008; Menzly and Ozbas, 2010; Cohen and Lou, 2012; Huang, 2015; Hoberg and Phillips, 2018; Ali and Hirshleifer, 2020) indicates cross-predictability in stock returns, suggesting a gradual information diffusion. These informational spillover effects have been primarily documented for economically linked stocks. Following this argument, we should detect spillover effects using our *BNs* and compare the generated alphas to assess network accuracy.

To do so, we construct equally-weighted (value-weighted) calendar-time portfolios from 2001 to 2021. For every stock, we calculate the

[11] We use the Python package "spaCy" and a transformer-based entity recognition model to identify company names from text, matching recognized organizations with the SEC's EDGAR database, the CRSP master file, and the LSEG dataset.

[12] Although this dataset contains some international firm relations, most links are among US firms, so we refrain from using it to evaluate our global network.

Table 4
Most similar firms for *Ford Motor* in 2021.

Name	Country	Sector
Panel A: US business networks		
TNIC		
GENERAL MOTORS	USA	Automobiles and Parts
TESLA	USA	Automobiles and Parts
LKQ	USA	Automobiles and Parts
PACCAR	USA	Industrial Engineering
ON SEMICONDUCTOR	USA	Technology hardware and equipment
T5 − XXL		
GENERAL MOTORS	USA	Automobiles and parts
PACCAR	USA	Industrial engineering
BORGWARNER	USA	Automobiles and parts
GENUINE PARTS	USA	Automobiles and parts
COPART	USA	General retailers
OpenAI − L		
GENERAL MOTORS	USA	Automobiles and parts
BORGWARNER	USA	Automobiles and parts
ALLY FINANCIAL	USA	Financial services (Sector)
GENERAL ELECTRIC	USA	General industrials
MICROSOFT	USA	Software and computer services
Panel B: Global business networks		
BOW		
GENERAL MOTORS	USA	Automobiles and parts
GREAT WALL MOTOR CO.	China	Automobiles and parts
BASF	Germany	Chemicals
ZHONGSHENG GP.	Hong Kong	General retailers
BOEING	USA	Aerospace and defense
T5 − XXL		
GENERAL MOTORS	USA	Automobiles and parts
GREAT WALL MOTOR CO.	China	Automobiles and parts
MARUTI SUZUKI INDIA	India	Automobiles and parts
TOYOTA MOTOR	Japan	Automobiles and parts
SAIC MOTOR	China	Automobiles and parts
OpenAI − L		
GENERAL MOTORS	USA	Automobiles and parts
HONDA MOTOR	Japan	Automobiles and parts
HYUNDAI MOTOR	Korea	Automobiles and parts
NISSAN MOTOR	Japan	Automobiles and parts
TOYOTA MOTOR	Japan	Automobiles and parts

This table presents the five most similar firms with a sufficiently similar market capitalization (the company should be at least in the 8th NYSE size decile) of the car manufacturer *Ford Motor* as of 2021. We consider US-only networks as well as global networks.

Table 5
Global relations, overlaps, and correlations of business networks (2021).

	Random	BOW	T5-XXL	OpenAI-S	OpenAI-L
Panel A: Summary					
Mean #Relations	398	400	404	398	398
Median #Relations	266	188	213	281	266
Same country (%)	5.24	40.22	42.85	51	54.71
Same SIC4 (%)	0.86	5.93	13.13	10.12	12.25
Same SIC3 (%)	1.46	9.19	17.89	13.92	16.40
Same SIC2 (%)	3.59	15.56	27.98	23.39	26.95
Panel B: Overlap (%)					
Random	100	1.15	1.75	1.69	1.90
BOW	1.13	100	17.27	15.09	16.45
T5-XXL	1.76	17.53	100	40.34	45.81
OpenAI-S	1.69	15.26	40.21	100	58.79
OpenAI-L	1.90	16.65	45.69	58.84	100
Panel C: Return correlation					
Random	1	0.39	0.38	0.42	0.42
BOW	–	1	0.60	0.62	0.62
T5-XXL	–	–	1	0.76	0.77
OpenAI-S	–	–	–	1	0.85
OpenAI-L	–	–	–		1

We analyze the similarities of different global business networks on the basis of the most current descriptions available in 2021. Next to the average number of relations and the number of relations within a country and industry, we also identify how similar the relations are. Therefore, we calculate how many of the firm relations co-occur in another network in Panel B. For example, 16.65% of all stocks in the *BOW* network are also included in the network based on *OpenAI − L*, and conversely, 16.45% of all stocks in the *OpenAI − L* network are also found in the *BOW* network. In Panel C, we report the correlation between peer returns identified across the different networks for the same firm. Next to the different business networks, we also construct a network that contains random firm relations (*Random*). For a given stock, we identify the number of peers as identified by the OpenAI-L network and randomly select the same number of (fictitious) peers.

average past month's return of peer firms at the start of each month.[13] We then pursue long (short) investments in the 20% of stocks whose most similar firms performed best (worst) in the previous month. We evaluate these portfolios using the five factors from Fama and French (2015) plus momentum and short-term reversal. Next to US-only portfolios, we also construct portfolios for global markets using global factor data, as US factor data is arguably insufficient for explaining international stock returns.[14]

According to Panel A in Table 7, which shows equally-weighted seven-factor alphas, the US *BOW* portfolio exhibits the lowest monthly alpha of 119 bps. In contrast, *OpenAI − L* generates a 146 bps alpha with a t-statistic of 6.72. The 27 bps increase compared to the word-based network is significant at the 1% level.[15] This is slightly lower than the 156 bps generated by *TNIC*, but higher than the 132 bps of *T5 − XXL* and 124 bps for *OpenAI − S*. When controlling for the *LLM look-ahead bias* via masking, there is a 16 bps decrease for *OpenAI − L*, but an increase of 14 bps for *OpenAI − S*.

Globally, Panel A illustrates that *BOW* achieves an alpha of 208 basis points (bps). However, context-aware networks exhibit higher seven-factor alphas ranging from 265 to 281 bps, with t-statistics above 10, and significantly outperforming *BOW* at the 1% level. Although alphas for masked networks are lower, they still significantly exceed those of *BOW*, again suggesting that the *LLM look-ahead bias* might not have a strong presence in this application.

Panel B reveals that value-weighted portfolios can also achieve highly significant seven-factor alphas. For *OpenAI − L*, global portfolios yield highly significant alphas up to 74 bps, with statistical significance at the 1% level. Considering masked networks, we even observe a 88 bps for (*OpenAI − S*). Notably, there is greater variation in the value-weighted alphas of US portfolios. A portfolio based on *OpenAI − L* generates a monthly seven-factor alpha of 40 bps (t-statistic of 2.54), whereas the alpha from a portfolio based on *OpenAI − S* is not statistically significantly larger than zero. This discrepancy could be attributed to the presence or absence of mega-cap stocks in the long or short portfolios, which may have a strong influence of the value-weighted portfolio returns. We therefore follow the methodology of Jensen et al. (2023) and present capped value-weighted portfolio alphas in Panel C. Specifically, we cap market capitalizations at the 80th percentile within each quintile to avoid a dominating influence of few extremely large

[13] Peer firms are determined considering the BNs of the previous year to avoid using future information.

[14] The factors are calculated by following the methodology mentioned on the website of Kenneth R. French as closely as possible. For additional information about the construction of global asset pricing factors, see Huber et al. (2023), who analyze the suitability of competing asset pricing models for global stock markets.

[15] We test statistical significance by regressing the return difference of the two portfolios on the seven factors of the asset pricing model and observe the t-statistic of the intercept. Table 2 in the Online Appendix reports statistical significance levels for all return differences between the word-based and context-aware networks.

Table 6
Detection rate of US competition links.

	(1)	(2)	(3)
	100	50	10
Panel A: Disclosed 10-K (%)			
TNIC	52.60	44.46	**25.65**
BOW	32.83	23.37	10.44
T5-XXL	43.57	33.53	16.01
OpenAI-S	34.65	24.52	10.07
OpenAI-L	47.66	37.71	17.89
T5-XXL-Masked	46.42	34.48	15.64
OpenAI-S-Masked	**56.43**	**45.85**	24.38
OpenAI-L-Masked	54.67	45.10	24.55
Panel B: Comparable company analysis (%)			
TNIC	44.16	34.79	16.69
BOW	28.43	21.76	9.69
T5-XXL	39.60	31.64	14.12
OpenAI-S	30.30	23.04	9.18
OpenAI-L	43.20	32.67	13.67
T5-XXL-Masked	38.83	30.04	12.71
OpenAI-S-Masked	45.70	37.03	16.94
OpenAI-L-Masked	**47.88**	**38.13**	**18.36**
Panel C: FactSet revere (%)			
TNIC	46.44	36.30	16.83
BOW	30.02	20.96	8.23
T5-XXL	42.64	30.40	12.83
OpenAI-S	34.67	23.60	8.50
OpenAI-L	45.51	33.90	13.85
T5-XXL-Masked	43.86	32.88	13.35
OpenAI-S-Masked	**53.70**	**41.33**	**18.99**
OpenAI-L-Masked	52.28	40.29	18.33

We construct different US business networks on the basis of the most current descriptions available in 2021. Next to the TNIC dataset from Hoberg and Phillips (2010, 2016), we evaluate a word-based network (*BOW*) and multiple embedding-based networks ($T5-XXL$, $OpenAI-S$, and $OpenAI-L$). To directly compare the accuracy of our networks to the TNIC dataset, we modify the business networks so that they are restricted to US firms that are also included in the TNIC dataset. We then calculate in Panel A and B how many of the US competition links disclosed in the business section (Item 1) of 10-K filings and in M&A acquisition files (PREM14A, DEFM14A, and S-4/A) may be found in the networks (recall score). We repeat the analysis in Panel C using the competition links reported in FactSet Revere, a dataset containing competition, supplier-customer, and partnership links. The recall scores are presented in percentage points. The highest recall scores in a column and panel are shown in bold.

Table 7
Lead–lag effect: US and global evidence for different business networks.

Business network	TNIC	BOW	T5-XXL	OpenAI-S	OpenAI-L
Panel A: Equally-weighted					
US	1.56***	1.19***	1.32***	1.24***	1.46***
	(6.33)	(6.38)	(6.61)	(5.94)	(6.72)
US-Masked	–	–	1.34***	1.38***	1.3***
	–	–	(6.26)	(6.69)	(6.46)
GLOBAL	–	2.08***	2.65***	2.68***	2.81***
	–	(9.01)	(10.7)	(10.02)	(10.23)
GLOBAL-Masked	–	–	2.4***	2.6***	2.68***
	–	–	(10.02)	(10.69)	(10.54)
Panel B: Value-weighted					
US	0.32*	0.15	0.38**	0.15	0.4**
	(1.75)	(1.12)	(2.49)	(0.95)	(2.54)
US-Masked	–	–	0.21	0.35**	0.23
	–	–	(1.18)	(2.07)	(1.44)
GLOBAL	–	0.45**	0.71***	0.67***	0.74***
	–	(2.55)	(3.56)	(3.05)	(3.25)
GLOBAL-Masked	–	–	0.84***	0.88***	0.77***
	–	–	(3.92)	(4.19)	(3.69)
Panel C: Capped value-weighted					
US	0.93***	0.62***	0.72***	0.68***	0.81***
	(4.34)	(4.24)	(4.37)	(4.26)	(4.68)
US-Masked	–	–	0.76***	0.78***	0.68***
	–	–	(4.45)	(4.8)	(4.2)
GLOBAL	–	1.14***	1.56***	1.53***	1.65***
	–	(4.43)	(5.65)	(5.14)	(5.52)
GLOBAL-Masked	–	–	1.49***	1.59***	1.55***
	–	–	(5.59)	(6.22)	(5.84)

We study the lead–lag effect in the US and globally by constructing calendar-time portfolios that are rebalanced every month. We go long (short) in the 20% stocks whose most similar firms showed the best (worst) return in the previous month. In Panel A, we report seven-factor alphas (Fama and French, 2015 five-factor model plus momentum and short-term reversal) for equally weighted portfolios. Panel B, we display alphas for value-weighted portfolios. We also provide *capped* value-weighted alphas as suggested by Jensen et al. (2023) in Panel C. We consider several networks, including the TNIC dataset (available for the US only), a word-based network (BOW), an open-source Sentence Transformer model ($T5-XXL$) as well as a small and a large embedding model from OpenAI ($OpenAI-S$ and $OpenAI-L$) that are based on our full dataset of business descriptions (AI-Gen and commercial descriptions). We also show seven-factor alphas for portfolios based on masked business networks to explore the effect of a potential *LLM look-ahead bias*. We use Newey–West standard errors with two lags and denote the corresponding t-statistics of the coefficients in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

stocks. We observe highly significant capped value-weighted alphas between 68 and 81 bps in the US, and 149 and 165 bps globally for the embedding-based networks.

We also examine the lead–lag effect using (Fama and MacBeth, 1973) regressions at the stock level to control for other known cross-sectional return predictors, including alternative past month peer returns (e.g., based on industry membership). These results, reported in Table 8 in the Online Appendix, suggest our global business networks reveal novel business links.

Overall, our findings suggest the potential for developing profitable trading strategies with our context-aware *BNs*. The results are robust to a potential *LLM look-ahead bias* and can be reproduced with alternative cosine similarity thresholds (see Online Appendix). Importantly, the high strategy alphas indicate that our *BNs* consider economically closely intertwined stocks.

5.2. Predicting M&A deals

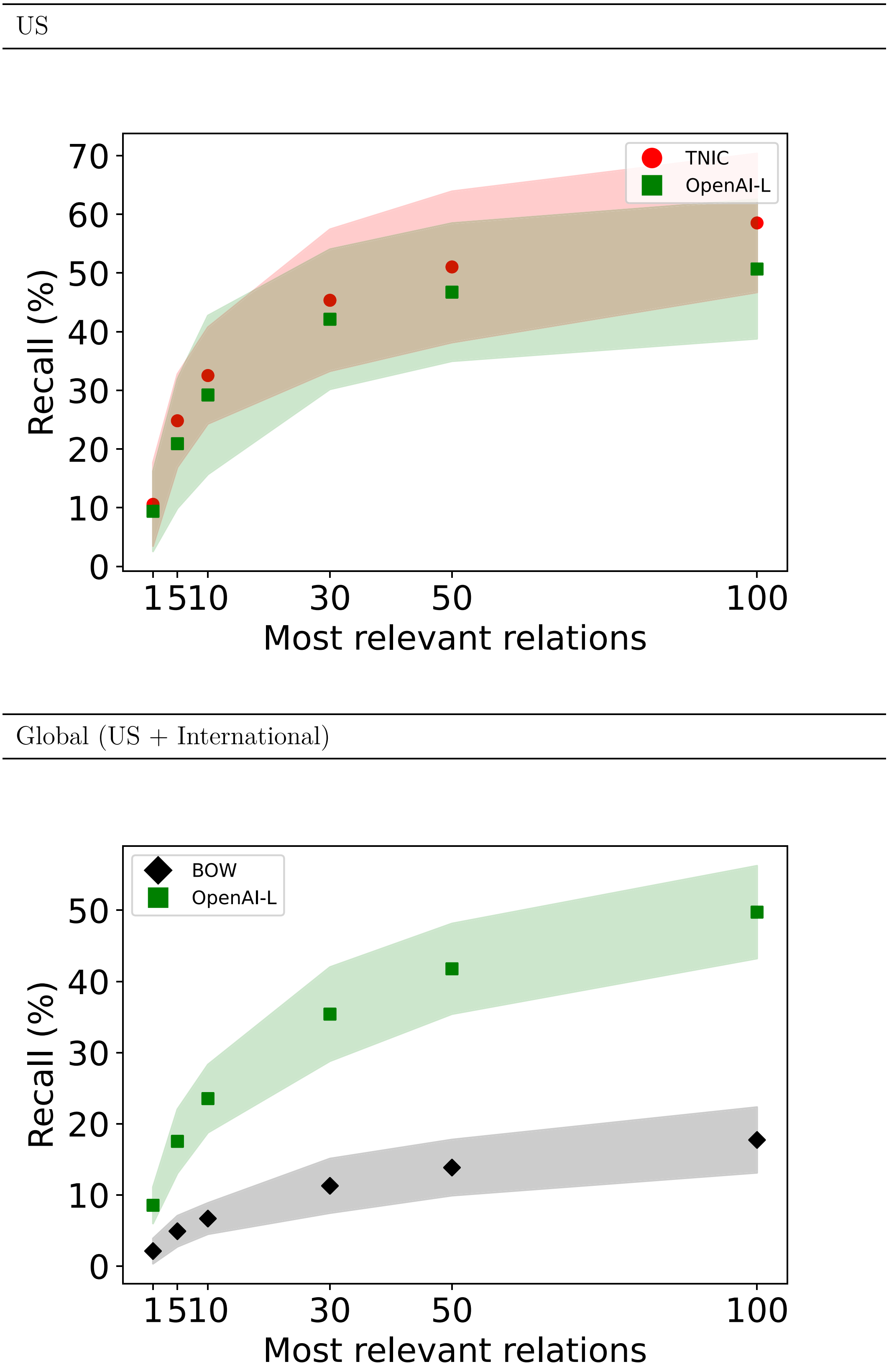
We continue evaluating our business networks' accuracy by testing their ability to predict takeover targets. This analysis is inspired by Hoberg and Phillips (2010), who find that firms tend to acquire highly similar firms due to potential synergies. We compare the accuracy of different business networks by calculating how many target firms are included among the most similar 10, 30, 50, and 100 firms.

We gather data from SDC Platinum on public firms that acquired publicly traded companies from 2000 to 2021. We calculate the recall score by counting how many target firms of M&A deals are included in our business networks.

According to Fig. 7, we observe a recall score of 29.22% for $OpenAI-L$, which is smaller (although not statistically significantly) than the 32.52% for $TNIC$, if we restrict our M&A dataset to US firms and consider the ten most similar firms. Considering the 100 most similar firms from $OpenAI-L$, we observe a recall score of 50.68%, compared to the 58.53% for $TNIC$. However, this difference is not statistically significant.

If we consider global M&A deals, we find that up to 23.54% of the acquired firms were among the ten firms with the most similar business description, according to $OpenAI-L$, in contrast to roughly 6.68% for BOW. This difference is statistically significant at the 1% level. We observe a similar pattern if we consider a larger number of firms. For example, among the 100 most similar firms, $OpenAI-L$ achieves a recall score of 49.73%, whereas a word-based approach significantly underperforms with a recall score of 17.73%.

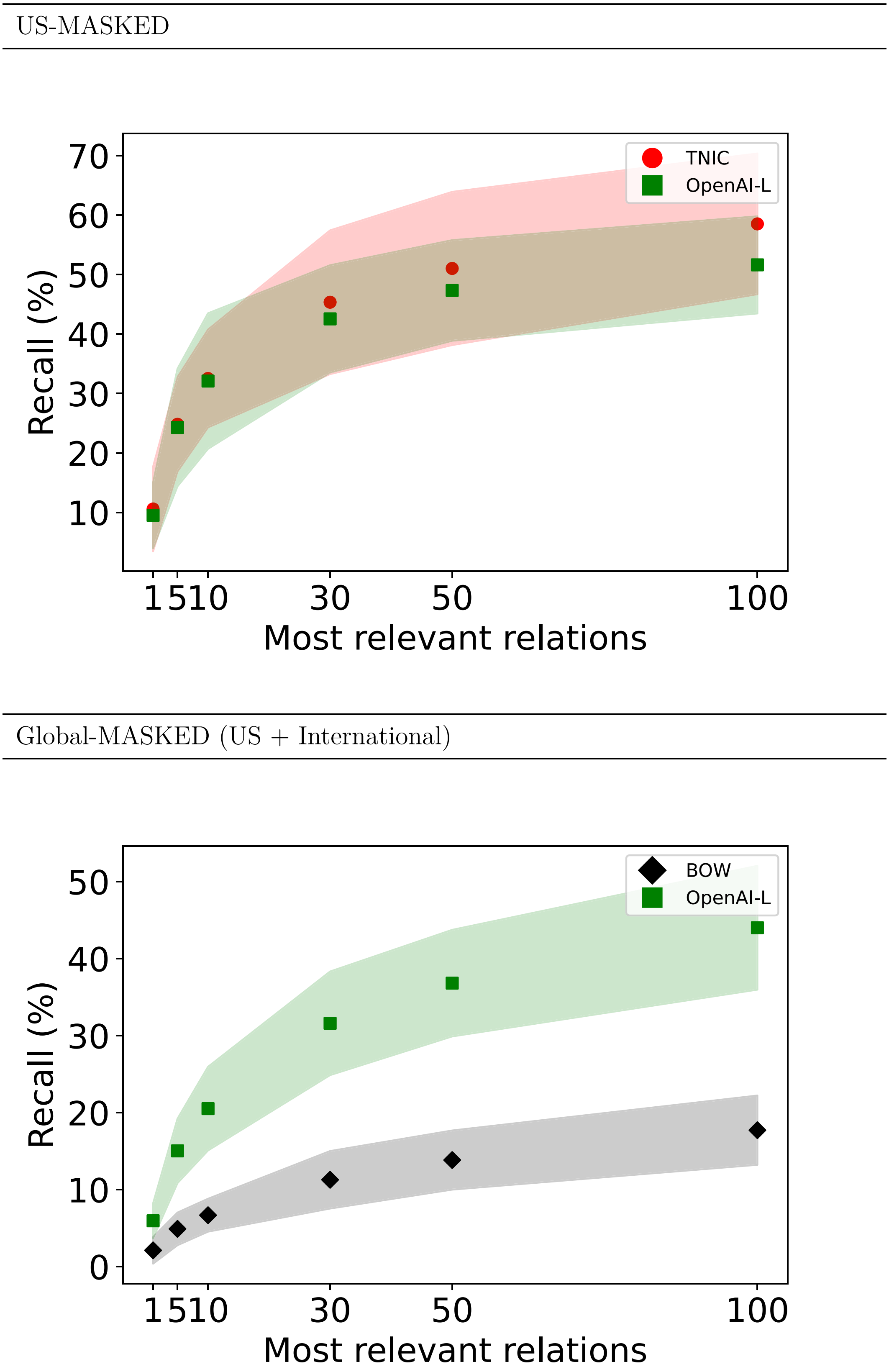
We repeat our analysis of M&A cases using masked business networks to control for a potential *LLM look-ahead bias* and present the results in Fig. 8. Here, we observe different effects for the US and



US

Global (US + International)

Fig. 7. M&A target firm detection. These figures examine the proportion of M&A target firms that are included in the different US (global) business networks (recall scores). The shadow areas in our visualizations represent the 95% confidence intervals of the data.



US-MASKED

Global-MASKED (US + International)

Fig. 8. M&A target firm detection (Masked networks). These figures examine the proportion of M&A target firms that are included in the different US (global) masked business networks (recall scores). The shadow areas in our visualizations represent the 95% confidence intervals of the data.

Table 8
Predicting M&A deals with logistic regressions.

	(1)	(2)	(3)	(4)
Same SIC4	39.63***	4.805***	10.69***	4.478***
	(23.83)	(8.68)	(11.57)	(8.10)
Same country	28.03***	4.789***	9.770***	4.476***
	(32.87)	(17.93)	(20.14)	(16.71)
Debt ratio	1.082***	1.078**	1.100***	1.083**
	(3.65)	(2.19)	(3.47)	(2.28)
ROE	1.060***	1.067**	1.062**	1.067**
	(2.84)	(2.11)	(2.38)	(2.07)
Cash	1.014	1.062**	1.051**	1.067**
	(0.58)	(2.26)	(2.28)	(2.48)
Similarity		1.228***		
		(47.41)		
Masked similarity			1.159***	1.240***
			(34.03)	(43.26)
Similarity difference				1.211***
				(44.81)
Pseudo-R2	0.338	0.554	0.441	0.556
N	666 297	666 297	666 297	666 297

This table presents the results of logistic regressions that examine the relation between business description similarity and the likelihood of an M&A. We use M&A data from SDC and randomly select 100 times as many non-merger firm pairs. We restrict the number of deals per acquirer and year to one. In total, we consider 6598 deals, 1600 deals with an US acquirer and 4998 deals with a non-US acquirer. The cosine similarity of the business descriptions is measured using the embedding model OpenAI-L and is displayed in percentage points. To further control for a potential LLM look-ahead bias, we split the similarity into two parts: The similarity based on masked descriptions and the difference between the two similarity measures (unmasked minus masked). We control for other relevant factors, such as whether the firm pair shares the same four-digit SIC code and country. Additionally, we consider fundamental information of the (fictitious) target firms, such as their profitability (ROE), cash amount (Cash), and debt share (Debt Ratio). All non-categorical variables other than the similarity measures are grouped into quintiles. We report odds ratios to ease the interpretation of the results. We cluster standard errors at the industry level. *indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

globally. In the US, the recall score for $OpenAI - L$ increases by about three percentage points when focusing on the top 10 most similar firms according to a masked network. Conversely, the recall score decreases by about three percentage points in a global setting. Importantly, masked embedding-based networks still outperform word-based networks, suggesting that a potential look-ahead bias does not explain the outperformance of these models.

Next, we establish a logistic regression framework to account for the likelihood that companies are also more inclined to acquire firms in the same industry and country. We control for the debt ratio, return on equity, and cash amount of targets to determine the additional explanatory power of textual similarity of business descriptions beyond these known variables. We map 6598 deals from SDC Platinum to our dataset, comprising 1600 deals with a US acquirer and 4998 deals with a non-US acquirer. For each M&A deal, we randomly select 100 non-merger firm pairs. The dependent variable in the logistic regressions is a dummy variable "acquired" set to one if a firm pair represents an M&A deal and zero otherwise. Our findings are presented in Table 8.

Table 8 confirms that firms in the same industry and country are more likely to be targeted, with odds ratios significantly larger than 1 and t-statistics of 23.83 and 32.87, respectively (see column 1). Acquiring firms also tend to purchase firms with higher debt ratios, though with smaller t-statistics. Even after controlling for these factors, the similarity between business descriptions remains statistically significant, with an odds ratio of 1.228 and a t-statistic of 47.41 (see column 2). The coefficient indicates that a one standard deviation increase in description similarity leads to a 83.22% increase in merger probability.

To control for a potential *LLM look-ahead bias*, we run a regression on the cosine similarity of masked descriptions in column (3). The odds ratio for textual similarity is significant but lower at 1.159, with a t-statistic of 34.03. In column (4), we add the difference between masked and unmasked cosine similarity as a bias proxy. We find a highly significant odds ratio for the *Similarity Difference*, which indicates that a *LLM look-ahead bias* might play a role here. Nevertheless, we still observe a significant odds ratio for the masked cosine similarity in column (4), suggesting that the *LLM look-ahead bias* does not fully explain our initial results. Based on these findings, we conclude that firms target peers in M&A deals with more similar businesses after controlling for other factors.

6. Classification of business relations

Up to this point, our networks do not distinguish between the exact types of firm relations, as they may include potential competitors, suppliers, or customers. However, some applications may require networks exclusively focused on one type of relationship. While databases like FactSet Revere provide insights into firm relations, particularly in the US, they do not cover all global stocks.

We propose fine-tuning a language model on the relationship data from FactSet Revere to address this gap. Specifically, we train a three-class prediction model on the business descriptions of the involved firms (input data) and the relationship categories (*competitor*, *supplier*, and *customer*). As a next step, we identify all active firm relations in FactSet Revere and align the previous year business descriptions of the corresponding firms. We then subdivide the dataset into distinct training and test sets. In the test set, we include various relations: some involving firms not present in the training set and others featuring firms that appear in different relations within the training set. We adopt an oversampling strategy to achieve a balanced representation of competition, supplier, and customer relations in the training data, ensuring equal distribution across these three categories.

The next phase involves fine-tuning a "Longformer" (Beltagy et al., 2020), a BERT-style open-source transformer model capable of processing up to 4096 tokens. As a robustness test, we further train a similar model using masked descriptions to identify to what extent the performance is biased by company or product names.[16]

We calculate the accuracy, precision, recall, and F1 scores for the model based on the full dataset of business descriptions in Panel A of Table 9. The overall three-class accuracy is 78.14%, significantly better

[16] Note that we refrain from training a BERT (Devlin et al., 2019) or RoBERTa (Liu et al., 2019) model, because the concatenation of two business descriptions can exceed their input limits of 512 tokens.

Table 9
Business relation classifier.

Relation	Accuracy	Precision	Recall	F1
Panel A: All descriptions				
Unmasked descriptions (original)				
Three class dataset	78.14			
Competitor vs. non-Competitor	83.33	84.38	76.21	80.09
Supplier vs. non-Supplier	87.06	52.04	73.76	61.02
Customer vs. non-Customer	85.88	84.50	81.56	83.01
Masked descriptions				
Three class dataset	73.03			
Competitor vs. non-Competitor	78.79	81.14	67.48	73.68
Supplier vs. non-Supplier	85.64	48.26	62.72	54.55
Customer vs. non-Customer	81.63	76.24	82.16	79.09
Panel B: AI descriptions				
Unmasked descriptions (original)				
Three class dataset	79.23			
Competitor vs. non-Competitor	83.56	80.96	82.32	81.63
Supplier vs. non-Supplier	88.93	55.84	72.07	62.92
Customer vs. non-Customer	85.96	87.52	78.20	82.60
Masked descriptions				
Three class dataset	85.73			
Competitor vs. non-Competitor	88.55	89.29	84.29	86.72
Supplier vs. non-Supplier	92.81	67.08	88.07	76.15
Customer vs. non-Customer	90.09	89.86	86.50	88.15

This table provides a detailed comparative analysis of four multi-class classification models, that differentiate between competition, supplier, and customer relations. Panel A displays the accuracy, recall, precision, and F1 score of a model that is trained on *AI-Gen* and commercial business descriptions. Additionally, the multi-class problem is transformed into three distinct binary classification tasks. For example, relationships are classified as either competitor or non-competitor (incorporating both supplier and customer categories), followed by the calculation of accuracy, precision, recall, and F1 score for these binary classifications. A similar approach is applied to the customer and supplier relationships. To control for a potential look-ahead bias, we also train and evaluate a similar model that is based on the masked versions of the same business descriptions. In Panel B, we evaluate the performance of two similar models, which they are trained on AI descriptions only.

than a random guess (33.33%), demonstrating the model's predictive capabilities. Transforming the multi-class problem into a binary classification task, the model also shows high accuracy (83.33%), precision (84.38%), recall (76.21%), and an F1 score of 80.09% in identifying competitors. Similar high accuracies are observed for identifying suppliers (87.06%) and customers (85.88%), with reasonable F1 scores. The performance metrics for masked descriptions are similar, indicating the model does not solely rely on firm names and other identifiers.

Panel B reports the performance for a model trained only on *AI-Gen descriptions*. Here, we observe an accuracy of 79.23% (85.73% for masked descriptions), which is higher than in Panel A. This increase may be due to the longer AI descriptions providing more critical information than the often shorter commercial descriptions from SDC Platinum and S&P Global.

Overall, we show that state-of-the-art language models can effectively derive the likely nature of business relations from pairs of business descriptions. Thus, our *BNs* can be used for research tasks focusing on specific types of business relations, like competitor links, provided the prediction model's errors are tolerable.

7. Conclusion

We generate a novel dataset of historical business descriptions, covering 91.6% to 99.8% of the US and 79.9% to 98.3% of international stock market capitalization from 2000 to 2021. These descriptions are constructed using business-relevant information extracted from annual reports and harmonized with GPT-3. For firm years where AI descriptions cannot be generated, we supplement with historical descriptions from vendors like S&P Global and SDC Platinum.

We are the first to apply open-source and proprietary embedding models to construct time-varying global business networks. Unlike word-based methods, our approach handles synonyms and word context to determine the similarity of business descriptions. Due to a lack of global networks, we evaluate the accuracy of our networks by calculating the number of competitor relations of US firms identified by our networks with those identified by TNIC and obtain a comparable performance.

We demonstrate the usability of our context-sensitive networks in two dimensions. First, we revisit the lead–lag effect and obtain significant equally weighted and value-weighted seven-factor alphas in the US and globally. Second, we predict M&A deals with firm similarity and find a significant relation in the US and globally, even when controlling for other characteristics.

Our study also investigates a potential *LLM look-ahead bias* when applying LLMs and their derivative embedding models in Finance and Economics. By repeating the competitor identification, lead–lag effect, and M&A prediction tasks using business networks based on masked descriptions, we investigate the size of this *LLM look-ahead bias* in different applications. While our findings indicate that the extent of the bias depends on the specific task, it seems generally advisable to anonymize information processed by LLMs to minimize the bias.

Finally, we demonstrate how open-source language models can be fine-tuned to distinguish between competitor, supplier, and customer relations, allowing us to refine our business networks further. This approach enables a more focused analysis depending on the specific research question.

Our networks, with their global reach and extensive coverage, uncover insights into worldwide economic connections and enable more precise, firm-specific assessments of competitor, supplier, and customer dynamics. To promote future research, we make the *AI-Gen descriptions* and the resulting global *BNs* available.

CRediT authorship contribution statement

Christian Breitung: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sebastian**

Müller: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT to improve the readability of the paper. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aghamolla, C., Thakor, R.T., 2022. IPO peer effects. J. Financ. Econ. 144: (1), 206–226.

Ali, U., Hirshleifer, D., 2020. Shared analyst coverage: Unifying momentum spillover effects. J. Financ. Econ. 136 (3), 649–675.

Beltagy, I., Peters, M.E., Cohan, A., 2020. Longformer: The long-document transformer. arXiv:2004.05150.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M.T., Zhang, Y., 2023. Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv:2303.12712.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. J. Finance 63 (4), 1977–2011.

Cohen, L., Lou, D., 2012. Complicated firms. J. Financ. Econ. 104 (2), 383–400.

Cohen, L., Malloy, C., Nguyen, Q., 2020. Lazy prices. J. Finance 75 (3), 1371–1415.

Daniel, K., Mota, L., Rottke, S., Santos, T., 2020. The cross-section of risk and returns. Rev. Financ. Stud. 33 (5), 1927–1979.

De Franco, G., Kothari, S.P., Verdi, R.S., 2011. The benefits of financial statement comparability. J. Account. Res. 49 (4), 895–931.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, pp. 4171–4186.

Eisdorfer, A., Froot, K., Ozik, G., Sadka, R., 2021. Competition links and stock returns. Rev. Financ. Stud. 35 (9), 4300–4340.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116 (1), 1–22.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. J. Polit. Econ. 81, 607–636.

Finke, C., Weigert, F., 2017. Does foreign information predict the returns of multinational firms worldwide? Rev. Financ. 21 (6), 2199–2248.

Foucault, T., Fresard, L., 2014. Learning from peers' stock prices and corporate investment. J. Financ. Econ. 111 (3), 554–577.

Frésard, L., Hoberg, G., Phillips, G.M., 2020. Innovation activities and integration through vertical acquisitions. Rev. Financ. Stud. 33 (7), 2937–2976.

Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G., 2006. Pairs trading: Performance of a relative value arbitrage rule. Rev. Financ. Stud. 19 (3), 797–827.

Glasserman, P., Lin, C., 2023. Assessing look-ahead bias in stock return predictions generated by GPT sentiment analysis. Available At SSRN: https://ssrn.com/abstract=4586726.

Griffin, J.M., Kelly, P.J., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Rev. Financ. Stud. 23 (8), 3225–3277.

Guo, F., Liu, T., Tu, D., 2023. Neglected peers in merger valuations. Rev. Financ. Stud. 36 (8), 3257–3310.

Hoberg, G., Phillips, G.M., 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. Rev. Financ. Stud. 23 (10), 3773–3811.

Hoberg, G., Phillips, G.M., 2016. Text-based network industries and endogenous product differentiation. J. Polit. Econ. 124 (5), 1423–1465.

Hoberg, G., Phillips, G.M., 2018. Text-based industry momentum. J. Financ. Quant. Anal. 53 (6), 2355–2388.

Hoberg, G., Phillips, G.M., 2025. Scope, scale and concentration: The 21st century firm. J. Finance 80 (1), 415–466.

Hou, K., 2007. Industry information diffusion and the lead-lag effect in stock returns. Rev. Financ. Stud. 20 (4), 1113–1138.

Huang, X., 2015. Thinking outside the borders: Investors' underreaction to foreign operations information. Rev. Financ. Stud. 28 (11), 3109–3152.

Huber, D., Jacobs, H., Müller, S., Preissler, F., 2023. International factor models. J. Bank. Financ. 150, 1–18.

Ince, O.S., Porter, R.B., 2006. Individual equity return data from Thomson Datastream: Handle with care!. J. Financ. Res. 29, 463–479.

Jacobs, H., Müller, S., 2020. Anomalies across the globe: Once public, no longer existent? J. Financ. Econ. 135 (1), 213–230.

Jensen, T.I., Kelly, B., Pedersen, L.H., 2023. Is there a replication crisis in finance? J. Finance 78 (5), 2465–2518.

Jiang, A.Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D.S., Casas, D.d.l., Hanna, E.B., Bressand, F., et al., 2024. Mixtral of experts. arXiv:2401.04088.

Kim, A., Muhn, M., Nikolaev, V.V., 2024. Financial statement analysis with large language models. Available At SSRN: https://ssrn.com/abstract=4835311.

Lee, C.M., Sun, S.T., Wang, R., Zhang, R., 2019. Technological links and predictable returns. J. Financ. Econ. 132 (3), 76–96.

Li, J., Lian, G., Xu, A., 2023. How do ESG affect the spillover of green innovation among peer firms? Mechanism discussion and performance study. J. Bus. Res. 158, 1136–1148.

Li, F., Lundholm, R., Minnis, M., 2013. A measure of competition based on 10-K filings. J. Account. Res. 51 (2), 399–436.

Liu, N.F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., Liang, P., 2023. Lost in the middle: How language models use long contexts. arXiv:2307.03172.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692.

Menzly, L., Ozbas, O., 2010. Market segmentation and cross-predictability of returns. J. Finance 65 (4), 1555–1580.

Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? J. Finance 54 (4), 1249–1290.

Müller, S., 2019. Economic links and cross-predictability of stock returns: Evidence from characteristic-based 'Styles'. Rev. Financ. 23 (2), 363–395.

OpenAI, 2023. GPT-4 technical report. arXiv:2303.08774.

Parsons, C.A., Sabbatucci, R., Titman, S., 2020. Geographic lead-lag effects. Rev. Financ. Stud. 33 (10), 4721–4770.

Reid, M., Savinov, N., Teplyashin, D., Lepikhin, D., Lillicrap, T., Alayrac, J.-b., Soricut, R., Lazaridou, A., Firat, O., Schrittwieser, J., et al., 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv:2403.05530.

Reimers, N., Gurevych, I., 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, pp. 3982–3992.

Reimers, N., Gurevych, I., 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. EMNLP, Association for Computational Linguistics, pp. 4512–4525.

SEC, 2005. EX-99.1: P&g reaffirms gillette acquisition financial impacts. URL: https://www.sec.gov/Archives/edgar/data/80424/000095015205007960/l16247aexv99w1.htm.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al., 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv:2307.09288.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.