# Generating Exposures with Large Language Models: Insights into M&A Activity*

Artem Streltsov§

**Abstract**

I propose a Large Language Model to generate time-varying exposure of firms to arbitrary user-defined risks using conference calls. This approach respects language complexity, does not require an expert to generate an extensive list of keywords, produces exposure measures with respect to any query and is directly applicable to any textual input. Leveraging this methodology, I generate firm-level exposures to geopolitical risk, population aging, climate change and demonstrate how exposure to such risks affects M&A activity. Geopolitical risk makes firms more likely to be an acquirer and less a target. It is propagated through supply chains and makes firms acquire targets in the U.S. with higher investment irreversibility driving vertical integration. Population aging increases aggregate M&A activity doubling the effect from labor shortage on M&A activity. Managers pay larger premia for low climate change risk targets, while investors react negatively to deal announcements of targets undergoing structural change. Finally, I describe other potential applications of the methodology in finance.

**Keywords:** Large Language Models, Machine Learning, Textual Analysis, Mergers and Acquisitions

*I am forever grateful to my advisers, Lin William Cong, Murillo Campello, Maureen O'Hara and Thorsten Joachims, for their support and invaluable guidance. Additionally, I am thankful to Richard Townsend, Daniel Neuhann, Boris Vallee, Amiyatosh Purnanandam, Tarun Chordia, Valentin Haddad, Scott Yonker, Filippo Mezzanotti and Kyle Bradbury for their helpful suggestions and comments.

§Cornell University, Johnson Graduate School of Management. E-mail address: as3923@cornell.edu

# 1 Introduction

Mergers and acquisitions (M&As) are some of the most important corporate decisions that can have a paramount effect on a firm's operations and outlook. The aggregate deal value, albeit down from its record high of $2.7trln, is still at an astonishing $1.3trln as of 2024 and the magnitudes of many M&A deals make such events a high-stake bet. Prior literature has made tremendous strides in explaining variation of M&A activity and outcomes over time, but much of this variation still remains unaccounted for.

The economic environment is rapidly evolving with the prominence of some risks such as inflation or geopolitical risks drastically changing over time and brand-new risks such as cybersecurity or cryptocurrency arising. These risks, however, affect each firm to a different extent even within the same product space or industry having a variable impact on the firm's financial and M&A decision-making. Ability to reliably measure firm's exposure to any risk, however immeasurable, over time is invaluable.

In this paper I propose a flexible Large Language Model to estimate time-varying firm-level exposure to any user-defined risk using conference call data. I further apply my model to M&A events studying how geopolitical risks, population aging and climate change affect firm-level M&A decisions.

The model is applied to, but never trained on conference calls and is directly applicable to any other textual input such as company websites or contract descriptions. The pipeline is inspired by Generative Adversarial Networks (or GANs, Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, and Bengio (2014)) and combines two transformer-based models: a topic-modelling transformer for sequence processing and topic ex-

traction and a text-generating transformer, a Large Language Model (LLM) tasked with generation of synthetic human-like text given a query of interest.

The query in question represents whatever risk one might be interested in quantifying exposures to. For each such query the LLM generates a synthetic mini corpus of text passages on the topic of interest. The topic-modelling transformer then learns to distinguish this new topic of interest from its knowledge base, the New York Times Annotated Corpus. As an advanced supervised algorithm explicitly trained to detect topics, the topic-modelling transformer is able to generate exposures reliably, fast, and free-of-cost. The model is further applied to conference calls to produce firm-level probabilities for all topics in the updated knowledge base. These probabilities represent exposure of each firm to specific risk or topic at time $t$.

I concretely demonstrate superiority of my approach over commonly used in finance word counting and "beta" exposure estimation methods as well as over proprietary LLMs such as ChatGPT in Section 4. In the next section I also provide an extensive literature review on textual analysis in finance explaining how my methodology contributes to that strand of literature joining Cong, Liang, Zhang, and Zhu (2019) and Hanley and Hoberg (2019) as the only papers combining some form of word embedding with topic modelling.

Geopolitical risks are increasingly prominent in today's world, with tensions at the highest level since the Cold War. These risks, however, have so far been understudied in the context of M&A. In a recent survey by Ernst&Young 81% of executives reported altering their investment strate-

gies in response to geopolitical risks. Rising geopolitical tensions disrupt international supply chains and customer-supplier relationships. Firms exposed to such tensions are likely to start looking to hedge their geopolitical risk exposure, build these relationships elsewhere and M&A is one common way to do so. At the same time such tensions create uncertainty, making firms vulnerable to such risks unattractive for acquisition.

In support of this, I find that a higher measure of geopolitical risks makes U.S. firms on average 3.5% more likely to acquire another U.S. firm the following year, but 58% less likely to be an acquisition target themselves. I confirm these findings by leveraging the exogenous nature of sanctions against Russia imposed in 2014. I use my LLM exposures to assign firms to treatment with matched controls. In a differences-in-differences setting I demonstrate that exposed firms were 25% more likely to acquire another firm and 100% less likely relative to counterfactual to be acquired themselves following the spike in geopolitical risk.

Economic uncertainty makes firm values and merger synergies less predictable, directly affecting M&A decision making (Garfinkel and Hankins (2011); Bhagwat, Dam, and Harford (2016); Bonaime, Gulen, and Ion (2018)). I confirm that uncertainty depresses M&A activity with financial uncertainty having by far the largest effect on M&A activity, twice as large as the Economic Policy Uncertainty studied in Bonaime et al. (2018). This result aligns with my finding that geopolitical risks induce uncertainty making exposed firms much less likely to be a target of an acquisition, but it goes against my

other finding that such firms are also more likely to acquire other U.S. firms.

To uncover the driving forces behind this puzzling result I hypothesize the following economic channels. First, geopolitical risks are likely realized through supply chain disruptions or a perceived threat thereof making firms with pre-existing relationships with riskier countries more vulnerable. Second, uncertainty abroad increases the value of the option to delay investment in such countries (Lee and Shin (2000); Bloom (2009)), while decreasing the relative value of the option to delay investments within the U.S. making firms more likely to acquire within the U.S. Finally, higher levels of geopolitical risks could lead to U.S. firms using M&A as an operational hedge for risk management purposes (Garfinkel and Hankins (2011)).

Using Factset's Supply Chain dataset, I demonstrate that the heightened acquisition probability from geopolitical risks is attributable to supply chain disruptions. Firms with a pre-existing supply chain relationship with Russia and China are instead 9.5% more likely to be an acquirer in response to heightened geopolitical risks (compared to 3.5% for all firms). Absent geopolitical risk such supply chain relationships make U.S. mergers less likely, but in presence of geopolitical risks they significantly increase acquisition probability within the U.S.

The real option channel largely posits that the value of the option to delay an investment drives M&A decisions with uncertainty increasing the value of this option and decreasing M&A activity in the affected countries. This option to delay foreign investment can, however, decrease the relative

value of the option to delay investment within the U.S. The value of the option depends to a large extent on the degree of investment irreversibility with more irreversible investments exhibiting a greater effect. Using target's capital intensity ratio as a proxy for investment irreversibility in a Heckman's two-stage sample selection model framework, I demonstrate that geopolitical risks are associated with more irreversible M&A deals in the U.S.

Vertical mergers are commonly used as an operational hedge. It has been shown that firms facing a heightened cash flow volatility are more likely to utilize vertical mergers (Garfinkel and Hankins (2011)). U.S. targets are much less likely to be exposed to the same geopolitical risks as targets abroad making their acquisition a viable risk management strategy helping acquirers diversify their geopolitical risk exposure and potentially reduce costs. I use Input-Output Make and Use tables from the Bureau of Economic Analysis and classify mergers as vertical if inputs from an industry account for at least 1% of outputs of another industry following classification strategies from Fan and Goyal (2006), Garfinkel and Hankins (2011) and Ahern and Harford (2014). In a Heckman's two-stage sample selection model framework, I show that geopolitical risks make vertical mergers more likely relative to non-vertical demonstrating that geopolitical risks drive vertical integration within the U.S.

In addition to geopolitical risks, I also study how population aging and climate change affect M&A decisions. Population aging is a major challenge that most developed countries face. It can affect M&A activity through a

variety of channels. Older entrepreneurs may wish to sell their businesses as they retire, increasing M&A activity. Additionally, an aging population increases demand for healthcare, travel, pharmaceuticals and social services, increasing M&A activity in these sectors further. Labor shortages induced by mass retirement may force acquisitions of firms stuffed with younger skilled workers. On the other hand, aging may also be associated with a decreased economic growth (Maestas, Mullen, and Powell (2016)) potentially decreasing M&A activity. I find that population aging increases aggregate M&A activity both in terms of aggregate deal values and number of deals. I attribute this effect to the labor shortage channel whereas population aging doubles the effect from labor shortage on M&A activity. To my knowledge, this is the first paper to study how population aging affects M&A activity.

Climate change is a pressing issue, and many managers have started incorporating ESG into their decisions. The relationship between climate change or the E-pillar of ESG and M&A activity has so far received little to no attention in finance literature. According to a Deloitte survey, over 70% of companies have abandoned some of their acquisitions due to ESG concerns. Additionally, 83% of respondents in that survey indicated that they would be ready to pay a premium of at least 3% for a high ESG profile target. I have found that, indeed, a decrease in target's climate change exposure by one standard deviation is associated with a 23.3% higher buyout premium. Interestingly enough, my measure of climate change exposure is highly statistically significantly (t-statistic of -14.5) and negatively correlated with the

E-pillar of the ESG score and Environmental R&D expenditures. In other words, managers of companies faring worse in terms of sustainability actually spend more time talking about it than their peers.

Finally, markets disapprove of targets undergoing a structural change as defined by a larger Euclidean distance of the latest confidence call representation from its 1-year moving average representation and respond more negatively to deal announcements with such firms as targets.

More broadly, my model is highly flexible and capable of estimating exposures with respect to any topic of interest, is applicable to any textual input – be it company websites, 10-Ks, conference calls, newsfeeds, patent or contract descriptions – with no model retraining required, and has potential applications in a wide range of finance subfields ranging from banking to asset pricing to institutional ownership.

# 2 Literature on Textual Analysis in Finance and Methodological Contribution

Textual analysis is widely used in finance research. One of the first papers to apply textual analysis in a finance setting, Das and Chen (2007), proposes a methodology to construct small investor sentiment from Yahoo stock forums. In an influential paper by Loughran and McDonald (2011) the authors develop the so-called LM positive and negative word lists that are used to analyze the tone of text. These word lists are used in a plethora of follow-up

papers such as Jegadeesh and Wu (2013) and Jiang, Lee, Martin, and Zhou (2019) where authors use market reaction to sign the words in each word list and construct a managerial sentiment index respectively.

The theme of using textual data to construct novel measures and indices is particularly topical. Baker, Bloom, and Davis (2016) introduce their economic policy uncertainty (EPU) index constructed from newspaper articles. Hassan, Hollander, van Lent, and Tahoun (2019); Hassan, Hollander, Lent, and Tahoun (2024) use conference calls to construct measures of political sentiment and risk, quantify exposure of firms to Brexit uncertainty, Covid-19, SARS, H1N1, Ebola, Zika, and MERS. Sautner, Van Lent, Vilkov, and Zhang (2023) and Li, Mai, Shen, and Yan (2021) estimate climate change exposures from conference call transcripts.

Most textual analysis in finance, however, is limited to a simple keyword search or a bag-of-words representation. This rudimentary approach, while intuitive and often powerful, has several well-known weaknesses. First, it requires an expert to generate for each query of interest an extensive list of smart keywords to search for and the ultimate exposure measure may be quite sensitive to this choice of keywords. Secondly, words are considered out of context and word order is ignored. In theory, "we operate an automobile business" and "we have abandoned our automobile business", "we are a major U.S. bank" and "our hotel is located on the bank of the Cayuga Lake" could all erroneously yield the same positive exposure to automobiles and banking respectively. Thirdly, all keywords are generally given the same

importance in computing the final score, but some words are more important than others and the context in which each word appears usually determines this importance. Finally, language is a complex living organism. It evolves with us making the task for a keyword search approach particularly hard. Machine learning has the capability to address these issues.

Word embeddings such as Word2Vec proposed in the seminal paper by Mikolov, Chen, Corrado, and Dean (2013) quickly gained popularity and eventually left their mark in the finance literature as well. Unlike the keyword or bag-of-words approaches, Word2Vec captures local semantic relationships with representations reflecting word meanings and relatedness. It can also handle out-of-vocabulary words by essentially imputing their representations from nearby words improving out-of-sample performance.

The finance literature employing word embeddings is rapidly growing. Cong et al. (2019) introduce a novel methodology to construct "textual factors" by combining a word embedding based on Word2Vec and topic modeling using Locality-Sensitive hashing. Li et al. (2021) uses word embedding to measure corporate culture. Wu (2023) proposes a measure of supply chain risk exposure based on a word embedding model trained on conference call transcripts. Chen, Kelly, and Xiu (2022) extract Word2Vec representations of the Wall Street Journal (WSJ) articles to measure the intensity of institutional investors' participation in news production. Hoberg and Phillips (2021) apply a related to Word2Vec Doc2Vec embedding model (Mikolov et al. (2013)) to 10-K filings and demonstrate that U.S. firms have substan-

tially expanded their scope of operations in the last 30 years, increasing their valuations and R&D activity. Caskurlu, Hoberg, and Phillips (2023) leverages patent embeddings from the Google Cloud Public Database constructed to predict a patent's CPC code from its text to produce a measure of technology sectoral disruption.

The main weakness of Word2Vec and of most word embeddings used in finance literature is that these embeddings are static, context-independent and generally based on a shallow neural network. This implies that if one were to apply a pretrained Word2Vec to the following two sentences, the representation for "bank" would still be the same: "we are a major U.S. bank" and "our hotel is located on the bank of the Cayuga Lake". Bidirectional Encoder Representations from Transformers or BERT (Devlin, Chang, Lee, and Toutanova (2018)) opened a new chapter in Natural Language Processing. BERT representations are bidirectional, dynamic, context-dependent and based on a deep transformer neural network capable of (or comes close to) uncovering the true meaning of each word based on its context.

Literature using such transformer models in finance is still scarce. Huang, Wang, and Yang (2023) propose their FinBERT model that is fine-tuned on the masked-language-modelling task on corporate filings, earnings conference calls and analyst reports demonstrating substantial improvement over traditional methods such as the widely used Loughran and McDonald dictionary. Cao, Jiang, Yang, and Zhang (2023) study the feedback effect on corporate disclosure in response to technology and show that firms with higher AI talent

supply reduce the frequency of negative sentences as measured by FinBERT after the introduction of BERT in 2018.

Computational and memory complexity are some of the greatest challenges faced by transformer models. Having a complexity of $O(N^2)$ with $N$ being the sequence length, BERT is limited to a maximum of 512-word tokens substantially limiting its useability in finance research. A new generation of transformers, however, has emerged (Linformer, BigBird, Longformer, Reformer, among others) approximating the self-attention matrix and reducing the complexity to linear in sequence length. Such models are capable of processing much longer sequences with often little loss in accuracy. Applications of such more advanced models in finance research are even harder to find, but one related working paper is Acikalin, Caskurlu, Hoberg, and Phillips (2022) that applies Longformer to model the probability of a patent becoming weakened following the Alice Corp. vs. CLS Bank International Supreme Court decision. One component of my methodology is the topic-modelling transformer that is also based on the Longformer encoder, but it is trained to detect topics instead in a multilabel classification setting.

A separate, just as scarce, but related to transformers and the present manuscript strand of literature, applies Large Language Models (LLMs) such as ChatGPT or Meta Llama to a finance setting. For example, Chen, Hoberg, and Zhang (2024) leverage these LLMs to extract contextualized representations of news text for predicting stock returns and demonstrate significant predictive power in 16 global equity markets. Another paper by Lopez-Lira

and Tang (2023) directly use ChatGPT to predict stock price movements and demonstrate substantial outperformance over traditional methods. The other component of my methodology is the text-generating transformer based on the Meta Llama 3 LLM, an open-source alternative to ChatGPT, that is tasked with generation of synthetic corpora on any topic of interest that the topic-modelling transformer further learns to detect.

Most modern topic-modelling analysis in finance is realized via Latent Dirichlet Allocation (LDA), a generative Bayesian model proposed by Blei, Ng, and Jordan (2003). Bybee, Kelly, and Su (2023) propose a narrative factor pricing model based on LDA and WSJ articles demonstrating strong out-of-sample performance. In a similar vein, Bybee, Kelly, Manela, and Xiu (2024) estimate a topic model to measure the state of the economy using WSJ articles successfully forecasting macroeconomic dynamics. Bellstam, Bhagat, and Cookson (2020) apply LDA to analyst reports and construct a novel measure of innovation. Zhang and Luo (2022) use LDA on Yelp reviews to model restaurant survival. Bae, Belo, Li, Lin, and Zhao (2022) analyze the relationship between information content, complexity of 10-K filings and uncertainty dynamics using LDA. Lowry, Michaely, and Volkova (2020) analyze communications between the SEC and firms prior to IPOs using LDA analysis. Adams, Ragunathan, and Tumarkin (2021) look at proxy statements and infer information gathering and decision-making responsibilities of boards and committees using LDA. Hanley and Hoberg (2019) propose an LDA-based methodology to detect emerging risks in the financial sector.

Huang, Lehavy, Zang, and Zheng (2017) demonstrate that analysts play the role of information intermediaries discovering information beyond corporate disclosures by applying LDA to analyst reports. Zhang, Yang, Zhang, and Palmatier (2022) analyze online text reviews and propose a model to extract consumer opinions.

Unsupervised machine learning methods such as the Latent Dirichlet Allocation (LDA) so popular in finance research have several limitations. They generally require selecting the number of topics to model ahead of time. In most cases, to capture the specific risk one might be interested in, the number of topics would have to be prohibitively large and would require substantial human effort to recognize and label each constructed topic. Topics constructed by LDA are also often ambiguous and hard to interpret (Aleti and Bollerslev (2024)) which is especially true when the number of topics is large. In addition, the model constructs uncorrelated topics and may therefore be unable to capture many of the risks of interest. LDA commonly struggles with short documents or those lacking relevant terms. Such models also do not allow evolution of topics over time and make most of the same linguistic assumptions as the keyword search or bag-of-words ignoring word order and syntax.

My approach circumvents these weaknesses, respects language complexity, does not require an expert to come up with an extensive keyword list for every risk of interest and produces exposure measures with respect to any query including one that is not in the original training corpus and one that fo-

cuses on an unseen topic. It marries a topic-modelling transformer that uses Longformer as its encoder and predicts topic probabilities with Meta Llama 3, an open-source Large Language Model, tasked with generation of synthetic corpora on any risk of interest. In Section 4 I provide concrete evidence of superiority of my "bottom-up" and "top-down" approaches to exposure generation over commonly used in finance word counting and "beta" exposure estimation methods respectively.

As Cong, Liang, Yang, and Zhang (2020) highlighted, the main advantage of combining word embedding and topic modeling lies in the ability to tackle linguistic complexity, while maintaining computational scalability and interpretability. My study thus joins Cong et al. (2019) and Hanley and Hoberg (2019) in this dimension, the only two papers combining some form of word embedding with topic modelling. Unlike these two papers, however, my approach (1) is based on more advanced, contextualized embeddings; (2) is based on embeddings that are trained for and adapted to topics of interest; (3) produces exposure measures with respect to any topic of interest; (4) is more powerful at topic detection since it is based on an advanced supervised algorithm; (5) generates and learns from synthetic corpora for the first time in finance research.

# 3 Transformers as Exposure Generators

The proposed pipeline relies on concepts of transfer learning, few-shot learning, learning from synthetic data and Large Language Models (LLMs). Transfer learning is a technique in which knowledge obtained from an auxiliary task is used to improve performance on a typically harder task of actual interest, especially when this harder task has less training data. My model is first trained on Wikipedia on a masked-language-modelling task. It then learns to distinguish topics in the New York Times Annotated Corpus. The knowledge of the trained model is then applied to the finance domain.

Use of synthetic data to augment the training corpus is a common trend in recent years in machine learning research. For example, Shermeyer, Hossler, Etten, Hogan, Lewis, and Kim apply AI.Reverie's simulation platform to generate 50,000 synthetic satellite images to aid in airplane detection. Xu, Huang, Luo, Bradbury, and Malof (2022) propose a model to generate large quantities of synthetic overhead imagery with a focus on few-shot and zero-shot learning applications. I use Large Language Models to generate synthetic data to expand the knowledge and capabilities of my model to topics previously unknown to the model, but of particular interest to the user. To my knowledge, this is the first paper in finance proposing to learn from synthetically generated corpora.

Few-shot learning is concerned with generalizing a large model to a new modified task that includes a new class with only a few representative exam-

ples. Few-shot learning is relevant to this present paper in that my model learns to detect a new topic by contrasting a small synthetic corpus on this topic of interest with the knowledge base (the New York Times corpus). My pipeline therefore consists of two transformer-based networks, a topic-modelling transformer, tasked with producing probabilities of a certain topic being present in a passage, and a text-generating transformer, tasked with synthetic corpus generation given a topic of interest.

Large Language Models have revolutionized Natural Language Processing (NLP). These are commonly based on some transformer architecture selected based on application. At the heart of such transformers-based models is the self-attention mechanism that produces a contextualized representation of each word token based on words in the same text passage. This mechanism bypasses the sequential dependency inherent in recurrent neural networks, the state-of-the-art in NLP prior to transformers, allowing large-scale parallelization and consequently much greater model complexity, i.e. ever larger language models, and ability to process very large datasets in practical timeframes. These two virtues paired with the self-supervised nature of language gave rise to a plethora of powerful transformers-based pretrained models and have brought transformers to the forefront of the NLP research.

A major setback of vanilla transformers is that the memory and computational complexity of the full self-attention mechanism is quadratic in the sequence length making them computationally intractable in many applications requiring concurrent analysis of longer text passages such as conference

calls. Most conference call transcripts are considerably longer than 512-word tokens, the maximum input length for popular transformer-based models such as BERT. In the case of conference calls, the average number of word tokens is 1200 and 1700 for introduction and Q&A sections respectively, while roughly 99% of sections, introduction and Q&A each, have less than 4096 tokens. I take advantage of the latest developments in NLP and opt for a next-generation transformer architecture, Longformer, proposed by Beltagy, Peters, and Cohan (2020) that approximates the self-attention matrix reducing the quadratic dependency to linear. Longformer is able to support up to 4096-word tokens in a single sequence which is enough to fit full conference call sections without losing too much information.

Longformer is used as the encoder in the topic-modelling transformer. It is first trained on a masked language modeling task on English Wikipedia whereas some word tokens are removed from text passages at random and the model is asked to fill in the blanks. I further transfer this knowledge obtained from Wikipedia articles to the space of topic detection in a multi-label classification setting using the New York Times Annotated Corpus. This corpus contains over 1.5 million articles published by the NY Times between January 1, 1987, and June 19, 2007. Each article is manually assigned a list of topics touched upon with 1057 topics in total.

After training on this corpus, the topic-modelling transformer is able to produce the probability that one of the 1057 topics is discussed in a text sequence. I further transfer this accumulated knowledge from Wikipedia and

NYT corpora to conference call transcripts producing probabilities of each of the 1057 topics that were mentioned in the NYT corpus using the Refinitiv Conference Call transcript dataset spanning from 2003 to 2021.

[Figure 1 about here.]

Although NYT corpus covers a wide range of topics, many topics of interest are likely missing from this list. The full labelled sample ends in September 2004, while the training set, I use for topic modelling ends on January 1, 2000. Such concepts as, for instance, cryptocurrencies and cybersecurity risks are then unrepresented in the corpus. To circumvent this fact and the weakness of supervised ML of having a fixed set of classes (topics in this case) I propose the following solution based on Large Language Models (LLMs) schematically presented in Figure 1 where N represents the number of topics known to the model so far.

Given any risk or topic of interest, a pretrained LLM such as Meta Llama 3 (an open-source version of ChatGPT by Meta) can be used to generate a mini corpus of text passages or essays on the topic. Unlike with ChatGPT, given that Llama 3 is open source, all text generation is possible free-of-charge. The previously trained topic-modelling transformer network based on Longformer can then predict which other pre-learned topics are also present in the new synthetic passages apart from this new topic of interest if any. The topic-modelling transformer is further trained on this generated mini corpus and a small sample from the original NYT dataset, the Knowledge

Base, in a matter of minutes. The probabilities corresponding to the topic of interest that the new neural network assigns to each conference call are then their exposure to the arbitrary risk in question. The format of the python prompts passed to the Meta Llama 3 text generator is presented in Figure 2. In generating a synthetic corpus, I first create a list of titles to base the generation on and then ask the LLM to generate an article on the topic in question with each pre-generated article title in a loop. This allows for a diverse and representative synthetic corpus.

Notice in Figure 1 that before producing topic probabilities the topic-modelling transformer generates a 24-dimensional representation of the input sequence, the Bottleneck representation, that is further used to construct topic probabilities. We thus have a low-dimensional representation of each conference call or input text sequence in the topic modelling space, the space in which over 1000 topics are as separable as possible. Given the quarterly frequency of conference calls, this representation can be tracked over time for each firm and information gains relative to previous conference calls or the respective industry can be estimated.

[Figure 2 about here.]

# 4 Generating Exposure Measures

I propose two distinct approaches to firm-level exposure measure generation: top-down and bottom-up. The top-down approach consists in choosing an

economy-wide index of interest and projecting it onto the topic-modelling subspace spanned by the 24 features learned by the topic-modelling transformer. I do this as follows. First, I aggregate individual firm-level 24-dimensional conference call representations by taking a quarterly average. This produces an economy-wide conference call-implied representation. I further run a linear regression with the index of interest as the dependent variable and the 24 features as predictors. This essentially finds an optimal combination of the 24 features encoding economy-wide conference call representation to mimic the index of interest as closely as possible. Finally, I apply the fitted linear regression model to individual firm-level conference call representations to produce a firm-level time-varying exposure measure.

This approach is similar in spirit to the common "beta" approach in finance whereas industry or stock excess returns are regressed onto an index of interest with the corresponding beta indicating industry or stock exposure. Confounding variables and omitted variable bias, sector or industry bias, low signal-to-noise ratio commonly observed in stock returns reduce the power of this approach yielding potentially biased exposures. A low correlation between aggregate returns and the index of interest coupled with poor ability of aggregate returns to predict contemporaneous index values provides room for such inaccuracies and some evidence in favor of exposures being less than accurate and against using the traditional "beta" methodology in estimating exposures.

[Table 1 about here.]

I consider a wide variety of macroeconomic indicators including VIX, consumer confidence index, Economic Policy Uncertainty (Baker, Bloom & Davis, 2016) and Economic Financial Uncertainty (Jurado et al, 2015) among others. In Table 1 I report adjusted $R^2$, Root Mean Squared Error (RMSE) and correlations indicating quality of fit of economy-wide indicators with the conference call representations returned by my LLM model. The correlations are substantial ($70\% - 94\%$). Using these fitted models I further construct firm-level measures of exposure to EPU, ERU, US CCI and other indicators.

In the same table I also compare the quality of index reconstructions using the full LLM 24-dimensional representation, just one best LLM feature, equal-weighted and value-weighted returns. I find that for 10 out of 12 indices (TEDRATE and recession are the two exceptions) even just one LLM feature performs better than value- or equal-weighted index with predictability gains of up to 41% in terms of RMSE (for ERU index). The approach using all 24 LLM features, on the other hand, demonstrates complete superiority over stock returns with gains in terms of RMSE ranging from 25% to 66%. The topic-modelling subspace constructed by my model is much richer than returns alone and is able to estimate whatever economy-wide index with higher precision.

For illustrative purposes, in Figures 3 and 4 respectively I juxtapose the estimated conference call implied EPU and EFU indices (in blue) with their true counterparts (in red, dashed). Within each pair of time series, the indices tend to co-move and have a correlation of over 83%. My top-down

approach is thus a valid alternative to the commonly used "beta" approach in estimating firm-level exposures.

The bottom-up approach, on the other hand, consists in the text-generating transformer producing a synthetic corpus on the topic of interest and the topic-modelling transformer learning to detect it as in Figure 1. The main advantage of this approach is that the model constructs quarterly firm-level exposure measures to any arbitrary risk of interest from conference calls. In principle, conference calls could be replaced with any other textual data source that fits any specific application better such as newsfeeds, company websites or 10-Ks.

The natural competitor to bottom-up firm-level exposure generation is word counting, also commonly used in finance research. In Table 2 I demonstrate concrete evidence of superiority of my LLM methodology to word counting. The New York Times Annotated Corpus features over 1mln articles with topics discussed hand-labelled by library experts. Since each article has ground truth, a topic assigned by a library specialist, I can directly showcase the edge of my model over word counting by comparing how well each method detects topics out-of-sample. I choose the top 20 economics-related topics by number of articles on each topic and split the sample into training (944,488 articles, for 1987-1999), validation (141,582 articles, 2000-2001) and test sets (168,119 articles, 2002-2004).

I proxy for word counting using a bag-of-words representation. Bag-of-words (BoW) performance is an upper bound for word counting commonly

used in finance because unlike word counting it allows for a flexible, data-driven keyword list and each word is assigned an optimal weight rather than being equal-weighted. I train a separate BoW model for each topic. The performance of my LLM model, however, is its lower bound. The model is not separately trained to focus on any one topic and instead constructs a 24-dimensional "jack of all trades" representation of each textual input capable of detecting over 1000 topics at once from the NYT corpus.

I compare performance in terms of F1-score (a harmonic average of precision, the fraction of articles labelled by a model as discussing a topic actually covering that topic, and recall, fraction of all articles actually featuring a topic detected by a model) and Average Precision (area under the Precision-Recall curve). These are the best metrics for rare event detection and any one topic is a rare event in such a rich corpus with over 1000 topics in total.

My model shows significant improvement over Bag-of-Words for all topics. Average gain from my LLM model over Bag-of-Words is 81% and 113% in terms of F1-score and Average Precision respectively. Additional training on any one topic of interest, as proposed in my pipeline in Figure 1 using synthetic or same NYT article data, or even training a separate model for each topic of interest, is straightforward in deep learning and bound to improve performance for any specific topic of interest further.

[Table 2 about here.]

The original version of my model was developed before the advent of pro-

prietary LLMs such as ChatGPT. In view of recent developments, however, it is only natural to expect a comparison between my LLM and such models. Meta Llama-3 70b is a cheaper alternative to other state-of-the-art proprietary LLMs, but is comparable to them in terms of performance on most tasks. In Table 3 I provide evidence that my LLM is a strong competitor to the best proprietary LLMs represented by Meta Llama-3 70b statistically significantly outperforming them on 17 and 18 out of 20 topics in terms of Average Precision and Accuracy respectively. The average performance gain is 14%, 17% and 6% in terms of F1-score, Average Precision and Accuracy respectively.

Apart from the better performance, my model is also significantly faster, is highly flexible, lightweight and is free-of-charge. GPT-4 Turbo, the most popular proprietary LLM provided by OpenAI, is charged at $10 and $30 per 1mln input or output tokens. Given a dataset of 300,000 conference calls with around 4,000 tokens on average each, the estimated cost per one full run is $15,000. It is customary to have to rerun the model more than once increasing the costs further. My model thus provides very tangible benefits to the research community and leads the way towards democratization of LLMs in finance research.

[Table 3 about here.]

[Figure 3 about here.]

Next, I use my LLM to construct exposure measures with respect to the

following topics using the bottom-up approach: "geopolitical risk", "population aging", "climate change", "skilled labor shortage", "artificial intelligence", "technology", "cybersecurity risks", "financial uncertainty".

[Figure 4 about here.]

In Figures 5 - 8 I juxtapose a moving average of my LLM measure with Google Trends for such terms as "cybersecurity risks", "artificial intelligence", "technology" and "climate change" (all normalized). In all cases, the time series tend to co-move with correlations ranging from 35% to 63%. Interestingly enough, both my LLM measure of "technology" and Google Trends have been on the decline since the early 2000s (Figure 6).

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

# 5 Geopolitical Risk and M&A Activity

M&As are some of the most important corporate decisions for any firm involved. Firms, however, get involved in M&A for a variety of reasons. The finance literature has uncovered the following reasons for an acquisition: potential for operational and financial synergy, access to bank finance, human

capital and R&D, retention of private benefits of control, market power acquisition and prevention of market entry by competitors.

In particular, Devos, Kadapakkam, and Krishnamurthy (2009) estimate that average synergy amounts to 10% of combined equity value of the two companies. Beaumont, Hebert, and Lyonnet (2022) show that firms acquire when their existing workforce is too costly and lacks the skills needed for entry, while Cornaggia and Li (2019) find evidence that greater access to bank finance increases targets' attractiveness and incentivizes merger implying that targets contribute to financial synergy too.

Another key aspect is R&D and patents. Phillips and Zhdanov (2013) demonstrate that small firms tend to innovate anticipating a buyout, while larger firms instead of participating in the R&D race with the smaller firms prefer to simply buy them out. Corporate governance and agency problems play a significant role in M&A as well. For instance, Gorton, Kahl, and Rosen (2009) show that managers anticipate merger waves and execute defensive acquisitions to retain their private benefits of control. Similarly, Harford, Humphery-Jenner, and Powell (2012) find evidence that entrenched managers avoid private targets, tend to overpay, and choose low synergy targets in the first place.

Finally, market power acquisition and entry prevention are important considerations that acquirers take into account as well. Bonaime and Wang (2024) analyze M&A activity in the pharmaceutical industry and conclude that M&A events lead to higher prices within drugs belonging to consolidat-

ing markets than within matched control drugs, while Cunningham, Ederer, and Ma (2021) show that incumbent firms tend to acquire innovative targets to discontinue their projects and preempt future competition (they refer to these as "killer acquisitions").

Geopolitical risks are particularly topical at present and have not received enough attention in the context of M&A so far. In this section, I analyze how geopolitical risk affects M&A firm decisions and through what channels. I start by modelling the probability of a firm being an acquirer or a target of an acquisition the following year as a function of firm's previous year geopolitical risk exposure as estimated by my LLM and a barrage of controls. Firm-level controls include past stock returns and return volatility, company size, sales growth, book leverage, cash to total assets and market-to-book ratios, ROA. Potential economy-level confounders include capital availability, quality of investment opportunities, levels of uncertainty and market valuations.

Harford (2005) shows that industry-level economic shocks affect M&A activity. In the same vein, I construct a measure of Fama-French 48 industry-level economic shocks by taking the first principal component for each industry separately of 7 variables representing absolute changes in sales growth, ROA, profitability, R&D spending, capital expenditures, asset turnover and employee growth.

To proxy for investment opportunities and expectations about future economic conditions, I take the first principal component of (1) the Chicago Fed National Activity Index, (2) the University of Michigan consumer confidence

index, (3) average of one-year GPD growth forecasts from the Livingston survey.

Economy-wide uncertainty has been shown to affect M&A activity (for example, Bhagwat et al. (2016); Duchin and Schmidt (2013)) – a result that I confirm as well using conference call measures. To control for economy-wide uncertainty I include the first principal component of common uncertainty measures such as (1) cross-sectional standard deviation of year-on-year sales growth (2) and of past 3-month cumulative returns; (3) the VIX index of implied volatility and (4) the monthly index of macroeconomic uncertainty by Jurado, Ludvigson, and Ng (2015). I control for overall market valuation levels with Shiller's cyclically adjusted price earnings (CAPE) ratio. Following Garfinkel and Hankins (2011), I also include the spread between Baa-rated bonds and the federal funds rate as a liquidity control and Fama-French 48 industry medians of 3-year cumulative returns and Tobin's Q as an industry-level valuation control.

I start by running a separate logistic regression for firm-level measures of geopolitical risk, financial uncertainty and Economic Policy Uncertainty (EPU) as estimated by my LLM. Each regression includes time and industry fixed-effects with robust standard errors clustered at the firm-year level. As follows from Table 5, one standard deviation higher exposure to geopolitical risk is associated with a 3.5% higher probability of being an acquirer the following year. Geopolitical risk induces uncertainty and from corporate finance theory we should expect uncertainty to depress M&A activity with

firms delaying irreversible investment. Prior research such as Bonaime et al. (2018) demonstrated that the Economic Policy Uncertainty (EPU) index is negatively related to M&A activity - a result that I confirm using my own firm-level EPU index. Additionally, I also find that financial uncertainty is twice as important as an uncertainty measure as EPU for M&A activity and also predicts lower M&A activity. Most importantly, this established negative relationship between uncertainty and fixed investment cannot explain the positive relationship between geopolitical risk and U.S. acquisition probability.

Next, I demonstrate in Table 6 that geopolitical risk substantially reduces the probability of being a target of an acquisition by 58.1% on average per one standard deviation in measure. This result is substantial both economically and statistically. As expected, uncertainty continues to depress M&A activity with economic uncertainty reducing the probability of being a target of an acquisition by 32.2%. The effects of geopolitical risk and uncertainty thus align for targets, but not for acquirers - a puzzle worthwhile further exploration.

Before trying to pinpoint the economic channels driving the results behind geopolitical risk I confirm my findings using an exogenous shock. On March 6, 2014, President Obama imposed sanctions against Russia in response to the annexation of Crimea by the Russian Federation precluding U.S. companies from trading and collaborating with major Russian businesses. This marked the start of a new geopolitically turbulent era directly

affecting some U.S. firms and making others at least somewhat concerned. The degree of this concern is captured by my LLM from contemporaneous conference calls. Being unanticipated at a yearly data frequency, this is an exogenous event that could help me demonstrate the effect of geopolitical risks on M&A decisions.

I use my LLM to estimate exposure of each firm to geopolitical risks in 2014 and assign firms in top 20% of firms in terms of this measure to treatment and bottom 20% to potential control. Following Campello, Graham, and Harvey (2010) and Almeida, Campello, Laranjeira, and Weisbenner (2012), I further find one matching control for each treated firm by matching exactly on year, industry (sic2) and credit rating and by minimizing the deviation in terms of Tobin's Q, cashflow, size and cash holdings. I consider 2 years before and after the event (2012-2016) with either being an acquirer or a target as the outcome. Additionally, I repeat the experiment using the 2003-2007 window with 2005 as a placebo event.

I run a differences-in-differences regression with treated units being those with high exposure to geopolitical risk, matched controls having low exposure and the event set to 2014. Regression results are reported in Table 4 and parallel trends are confirmed. The differences-in-differences coefficient is statistically significant at 1% for both acquirer and target probabilities. Heightened geopolitical risk caused exposed firms to be 0.044 (or 25%) more likely to acquire another firm and be 0.004 (or 100%) less likely to be acquired. The percentage effects are computed relative to counterfactuals as

estimated upon fitting the differences-in-differences regression. The placebo test of a similar shock happening in 2005 yields insignificant results.

[Table 4 about here.]

To uncover the driving forces behind the result that geopolitical risk makes exposed U.S. firms more likely to acquire other U.S. firms, I hypothesize the following economic channels. First, geopolitical risks are likely realized through supply chain disruptions or a perceived threat thereof making firms with pre-existing relationships more vulnerable. Second, uncertainty abroad increases the value of the option to delay investment in such countries (Lee and Shin (2000); Bloom (2009)), while decreasing the relative value of the option to delay for investments within the U.S. making firms more likely to acquire within the U.S. Finally, higher levels of geopolitical risk could lead to U.S. firms using M&A as an operational hedge for risk management purposes (Garfinkel and Hankins (2011)).

I start by estimating the extent to which each U.S. firm has a supply chain relationship with geopolitically risky countries such as Russia and China. Using Factset's Supply Chain dataset, I compute the share of these countries in supply chain of each firm. I further test if geopolitical risk is transmitted through supply chains and whether firms with a pre-existing supply chain relationship with Russia and China experience a particularly high likelihood of being a U.S. acquirer. I report the relevant coefficients in Table 7 Panel (1). First, I find that controlling for supply chain relationships geopolitical

risk alone is no longer statistically significant supporting my hypothesis that the risk is propagated through supply chains. Secondly, notice that in absence of geopolitical risks pre-existing supply chain relationships with Russia and China actually make a U.S. acquisition 2.8% less likely which makes sense as such firms are likely to look to acquire in Russia, China, somewhere close to these or nowhere. Finally, in presence of geopolitical risks firms with a pre-existing relationship with risky countries are instead 9.5% more likely to acquire in the U.S. We have thus confirmed that geopolitical risk is propagated through supply chains.

The real option channel largely posits that uncertainty would reduce more irreversible fixed investment and increase R&D spending instead. In the context of M&A and geopolitical risks the value of the option to delay an investment drives M&A decisions with uncertainty increasing the value of this option and decreasing M&A activity in the affected countries. This delay to invest abroad can, however, decrease the relative value of the option to delay investment within the U.S. The value of the option depends to a large extent on the degree of investment irreversibility with more irreversible investments exhibiting a greater effect. I use target's capital intensity ratio (PP&E to total assets) as a proxy for investment irreversibility. In a Heckman's two-stage sample selection model framework with the first stage being the same as the logistic regressions for acquirers above, I find that geopolitical risks are associated with more irreversible M&A deals in the U.S. (Table 8).

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

Vertical mergers are commonly used as an operational hedge. It has been shown that firms facing a heightened cash flow volatility are more likely to utilize vertical mergers (Garfinkel and Hankins (2011)). U.S. targets are much less likely to be exposed to the same geopolitical risks as targets abroad making their acquisition a viable risk management strategy helping acquirers diversify their geopolitical risk exposure and potentially reduce costs. I use Input-Output Make and Use tables from the Bureau of Economic Analysis and classify mergers as vertical if inputs from an industry account for at least 1% of outputs of another industry following classification strategies from Fan and Goyal (2006), Garfinkel and Hankins (2011) and Ahern and Harford (2014). In a Heckman's two-stage sample selection model framework, I show that geopolitical risks make vertical mergers more likely relative to non-vertical demonstrating that geopolitical risks drive vertical integration within the U.S. (Table 8).

To sum up, geopolitical risk makes U.S. firms much more likely to acquire other U.S. firms and much less likely to be acquired themselves. Uncertainty alone cannot explain these results. I show that geopolitical risk is propagated through supply chains. Geopolitical risk decreases relative value of the

option to delay investment within the U.S. leading to less reversible mergers. Geopolitical risks drive vertical integration in the U.S.

# 6 M&A Activity and Population Aging

Population aging is a major challenge that most developed countries face. It can affect M&A activity through a variety of channels. Older entrepreneurs may wish to sell their businesses as they retire, increasing M&A activity. Additionally, an aging population increases demand for healthcare, travel, pharmaceuticals and social services, increasing M&A activity in these sectors further. Labor shortages induced by mass retirement may force acquisitions of firms stuffed with younger skilled workers. On the other hand, aging may also be associated with a decreased economic growth (Maestas et al. (2016)) potentially decreasing M&A activity.

To start with, I investigate how a shock to economy-level measure of population aging affects M&A activity using impulse response function analysis. Similarly to Bonaime et al. (2018), I model a VARX(2) model with a constant, a time trend and aggregate cash holdings from the annual Compustat database as exogenous control. The equation looks as follows:

$$\mathbf{Y_t} = \mathbf{c} + \mathbf{A_1 Y_{t-1}} + \mathbf{A_2 Y_{t-2}} + \mathbf{B X_t} + \mathbf{u_t}$$

Exogenous variables $\mathbf{X_t}$ are the linear time trend and aggregate cash holdings. The endogenous variables $\mathbf{Y_t}$ include (1) natural logarithm of

the economy-wide aggregate of the risk of interest (population aging in this case); (2) VIX as a measure of economic uncertainty; (3) return on the value-weighted market index as an economic conditions control; (4) a measure of market liquidity, spread between the Baa rate and the federal funds rate; (5) Robert Shiller's cyclically adjusted price earnings ratio (CAPE) and (6) natural logarithm of the aggregate deal value or number of M&A events. The frequency of all variables is quarterly to match the quarterly nature of earnings conference calls. I aggregate the LLM measures to economy-level by taking an average within each quarter.

After fitting the model, we are interested in the orthogonalized impulse response to study how a shock to the risk of interest affects M&A activity. Order matters for an orthogonalized impulse response and I use the same order as the one specified above for the endogenous variables to mimic propagation of a positive shock to the measure of interest through the economy and finally having a real effect on M&A activity.

My findings indicate that population aging has a positive effect on M&A activity in terms of both deal values and number of deals (Figure 9). Population aging expands demand for healthcare, travel, insurance, pharmaceutical and social services. These are the firms with highest LLM exposures to population aging and, indeed, drive the increase in M&A activity (Figure 10). Aging also makes companies look to replace their skilled labor and M&A is one way to do so.

[Figure 9 about here.]

[Figure 10 about here.]

I confirm that higher economy-wide exposure to most uncertainty measures as recovered from conference calls such as Economic Financial Uncertainty (EFU) and Economic Policy Uncertainty (EPU) decreases M&A activity. This result is consistent with Bonaime et al. (2018) with investment irreversibility and a higher target negotiating power during uncertainty largely behind these findings. On the contrary, I find that higher exposure to business and consumer confidence increases M&A activity.

Further, I establish the channel behind the effect of population aging on M&A activity. It is commonly believed that M&A is one way to gain access to human capital (Beaumont et al. (2022); Chen, Hshieh, and Zhang (2023)). I use my LLM to construct a firm-level measure of skilled labor shortage and test if population aging amplifies the effect from labor shortage on M&A activity. In Table 7 Panel (2) I report the respective coefficients. Absent population aging, labor shortage makes firms 16% more likely to acquire another firm per one standard deviation in measure. Firms exposed to population aging, however, experience an effect twice as large as that from labor shortage alone.

# 7 Climate Change, Structural Changes and M&A Outcomes

Climate change is a pressing concern, and many managers have started incorporating ESG into their decisions. According to Deloitte, over 70% of companies have abandoned some of their acquisitions due to ESG concerns. High exposure to climate change could affect M&A activity through several channels. The frequency and intensity of climate-related disasters is increasing over time and their potential to cause damage to acquiring firms affects M&A decisions. Acquisition of climate change exposed firms may require compliance with regulation that already exists or is anticipated in the future in the climate space. Changing market conditions and consumer preferences may also increase demand for greener firms making other firms riskier in the long-run. Climate change also creates opportunities for firms to innovate and expand to new markets increasing competition to traditional firms. ESG concerns are becoming a priority for investors as well forcing some companies to prioritize sustainability in their M&A decisions.

I find that my measure of climate change exposure is highly statistically significantly (t-statistic of -14.5) and negatively correlated with the E-pillar of the ESG score and Environmental R&D expenditures. In other words, managers of companies faring worse in terms of sustainability actually spend more time talking about it than their peers.

In the same Deloitte survey 83% of respondents indicated that they would

be ready to pay a premium of at least 3% for a high ESG profile target and I find that there indeed is a statistically significant relationship between buyout premia and climate change exposure.

I follow Deng, Kang, and Low (2013) and Guo, Hotchkiss, and Song (2011) in constructing controls for the regressions in this section. More specifically, acquirer-level controls include firm size, leverage, Tobin's Q, cash to total assets, change in leverage prior to deal announcement, EBITDA to sales prior to deal announcement, total assets to sales 3 years post-merger and total acquisitions in the 3 years post-merger. Additionally, I control for industry median of EBITDA to sales prior to announcement and the following deal-specific attributes: log of deal size, indicators whether the target is a private firm and whether acquisition is hostile, percentage of cash and percentage of stock in the deal composition. All regressions include industry and time fixed effects, standard errors are heteroskedasticity robust (HC3).

In Table 9 I report regression coefficients for a regression of buyout premia on acquirer and target climate change exposures as estimated by my LLM and a barrage of controls. I find that a target's lower climate change exposure that directly correlates with a higher E-pillar of the ESG score predicts a 23.3% higher buyout premium per one standard deviation in measure.

[Table 9 about here.]

Finally, I compute a measure of information gain or structural change as the Euclidian distance between a firm's latest conference call 24-dimensional

LLM representation and its moving average over 1 year prior. Large distance indicates a structural change in the business environment that the firm is undergoing. I find that an announcement of an acquisition of a target with one standard deviation greater information gain measure is associated with a 4.2 percentage point more negative Fama-French 4 market reaction (significant at 5%, t-statistic is 2.4 as is reported in Table 9). Markets disapprove of targets undergoing structural changes prior to a deal announcement.

# 8 Conclusions

In this paper I introduce a methodology to automatically generate time-varying firm-level exposure to arbitrary user-defined risks. I demonstrate how my model can be used to reconstruct well-known economy-wide indices such as Economic Policy Uncertainty (EPU) or Economic Real Uncertainty (ERU) producing their firm-level exposure counterparts (the top-down approach) or how it can directly produce firm-level exposure measures with respect to any topic of interest (the bottom-up approach).

Using my model, I produce firm-level exposures to geopolitical risk, population aging and climate change studying how such risks affect M&A activity. Geopolitical risk makes firms more likely to acquire, but much less likely to be an acquisition target. This results cannot be explained by uncertainty induced by geopolitical risk alone. I demonstrate that the effect is propagated through supply chain relationships. Geopolitical risk reduces relative

value of the option to delay investment within the U.S. incentivizing firms to undertake more irreversible investments in the U.S. Finally, I show that geopolitical risk drives vertical integration in the U.S.

I find that population aging increases M&A activity. I attribute this fact to heightened labor shortages as skilled workers retire. Population aging doubles the effect from labor shortage on M&A activity. Managers prefer targets with a low climate change exposure and pay a significantly greater buyout premium for such targets. Additionally, investors dislike targets undergoing structural change and react more negatively to such announcements.

My model is highly flexible and capable of estimating exposures with respect to any topic of interest, is applicable to any textual input – be it company websites, 10-Ks, conference calls, newsfeeds, patent or contract descriptions – with no model retraining required, and has potential applications in a wide range of finance subfields ranging from banking to asset pricing to institutional ownership.

# References


Utku Acikalin, Tolga Caskurlu, Gerard Hoberg, and Gordon M. Phillips. Intellectual property protection lost and competition: An examination using large language models. Working Paper 4023622, Dartmouth University, Tuck School of Business, 2022. doi:10.2139/ssrn.4023622. https://ssrn.com/abstract=4023622. Available at SSRN.

Renee Adams, Vanitha Ragunathan, and Robert Tumarkin. Death by committee? an analysis of corporate board (sub-) committees. *Journal of Financial Economics*, 141(3):1119–1146, 2021.

K. R. Ahern and J. Harford. The importance of industry links in merger waves. *The Journal of Finance*, 69:527–576, 2014. doi:10.1111/jofi.12122.

S. Aleti and Tim Bollerslev. News and asset pricing: A high-frequency anatomy of the sdf. *The Review of Financial Studies*, hhae019, 2024. doi:10.1093/rfs/hhae019.

Heitor Almeida, Murillo Campello, Bruno Laranjeira, and Scott Weisbenner. Corporate debt maturity and the real effects of the 2007 credit crisis. *Critical Finance Review, now publishers*, 1(1):3–58, January 2012.

J. W. Bae, Frederico Belo, Jun Li, Xiaoji Lin, and Xiaofei Zhao. The opposing effects of complexity and information content on uncertainty dynamics: Evidence from 10-k filings. *Management Science*, 69(10):6313–6332, 2022. doi:10.1287/mnsc.2022.4615.

Scott Baker, Nicholas Bloom, and Steven J. Davis. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, November 2016. doi:10.1093/qje/qjw024.

P. Beaumont, C. Hebert, and V. Lyonnet. Build or buy? human capital and corporate diversification. Working Paper, Ohio State University, 2022.

G. Bellstam, Sanjai Bhagat, and J. Anthony Cookson. A text-based analysis of corporate innovation. *Management Science*, 67(7):4004–4031, 2020. doi:10.1287/mnsc.2020.3682.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. 2020.

V. Bhagwat, Robert Dam, and Jarrad Harford. The real effects of uncertainty on merger activity. *The Review of Financial Studies*, 29(11):3000–3034, November 2016. doi:10.1093/rfs/hhw061.

D. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(2003):993–1022, 2003.

N. Bloom. The impact of uncertainty shocks. *Econometrica*, 77:623–685, 2009. doi:10.3982/ECTA6248.

A. Bonaime and Y. Wang. Mergers, product prices, and innovation: Evidence from the pharmaceutical industry. *J Finance*, 79:2195–2236, 2024. doi:10.1111/jofi.13321.

Alice Bonaime, Huseyin Gulen, and Mihai Ion. Does policy uncertainty affect mergers and acquisitions? *Journal of Financial Economics*, 129(3): 531–558, 2018. doi:10.1016/j.jfineco.2018.05.007. ISSN 0304-405X.

L. Bybee, Bryan Kelly, and Yinan Su. Narrative asset pricing: Interpretable systematic risk factors from news text. *The Review of Financial Studies*, 36(12):4759–4787, December 2023. doi:10.1093/rfs/hhad042.

L. Bybee, B. Kelly, A. Manela, and D. Xiu. Business news and business cycles. *J Finance*, 79:3105–3147, 2024. doi:10.1111/jofi.13377.

Murillo Campello, John R. Graham, and Campbell R. Harvey. The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics*, 97(3):470–487, 2010. doi:10.1016/j.jfineco.2010.02.009. ISSN 0304-405X.

S. Cao, Wei Jiang, Baozhong Yang, and Alan L. Zhang. How to talk when a machine is listening: Corporate disclosure in the age of ai. *The Review of Financial Studies*, 36(9):3603–3642, September 2023. doi:10.1093/rfs/hhad021.

Tolga Caskurlu, Gerard Hoberg, and Gordon M. Phillips. New technology sectoral disruptions. doi:10.2139/ssrn.4866566. https://ssrn.com/abstract=4866566. Available at SSRN. Tuck School of Business Working Paper, 12 2023.

A. J. Chen, Gerard Hoberg, and Miao Ben Zhang. Institutional participation in information production and anomaly returns. doi:10.2139/ssrn.4207132. https://ssrn.com/abstract=4207132. Available at SSRN. USC Marshall School of Business Research Paper Sponsored by iORB, 07 2024.

J. Chen, S. Hshieh, and F. Zhang. Hiring high-skilled labor through mergers and acquisitions. *Journal of Financial and Quantitative Analysis. Published online*, pages 1–37, 2023. doi:10.1017/S0022109023000856.

Yifei Chen, Bryan T. Kelly, and Dacheng Xiu. Expected returns and large language models (november 22. https://ssrn.com/abstract=4416687. Available at SSRN, 2022.

L. W. Cong, T. Liang, B. Yang, and X. Zhang. Analyzing textual information at scale. In *Information for Efficient decision making*, pages 239–271. Information for Efficient decision making: big data, blockchain and relevance. World Scientific, Singapore, Singapore, 2020.

Lin Cong, Tengyuan Liang, Xiao Zhang, and Wu Zhu. Textual factors: A scalable, interpretable. *and Data-Driven Approach to Analyzing Unstructured Information* (September*, 1, 2019. doi:10.2139/ssrn.3307057. https://ssrn.com/abstract=3307057. Available at SSRN.

Jess Cornaggia and Jay Yin Li. The value of access to finance: Evidence from m&as. *Journal of Financial Economics*, 131(1):232–250, 2019.

C. Cunningham, Florian Ederer, and Song Ma. Killer acquisitions. *Journal of Political Economy*, 129(3):649–702, 2021.

Sanjiv R. Das and Mike Y. Chen. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9):1375–1388, 2007.

Xin Deng, Jun-koo Kang, and Buen Sin Low. Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, 110(1):87–109, 2013.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. URL http://arxiv.org/abs/1810.04805. cite arxiv:1810.04805Comment: 13 pages.

Erik Devos, Palani-Rajan Kadapakkam, and Srinivasan Krishnamurthy. How do mergers create value? a comparison of taxes, market power, and efficiency improvements as explanations for synergies. *The Review of Financial Studies*, 22(3):1179–1211, 2009.

Ran Duchin and Breno Schmidt. Riding the merger wave: Uncertainty, reduced monitoring, and bad acquisitions. *Journal of Financial Economics*, 107(1):69–88, 2013. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2012.07.003. URL https://www.sciencedirect.com/science/article/pii/S0304405X12001535.

Joseph P. H. Fan and Vidhan K. Goyal. On the patterns and wealth effects of vertical mergers. *The Journal of Business*, 79(2):877–902, 2006. doi:10.1086/499141.

Jon A. Garfinkel and Kristine Hankins. The role of risk management in mergers and merger waves. *Journal of Financial Economics*, 101(3):515–532, 2011.

I. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, pages 2672–2680, 2014.

Gary Gorton, Matthias Kahl, and Richard Rosen. Eat or be eaten: A theory of mergers and firm size. *Journal of Finance*, 64(3):1291–1344, 2009.

S. Guo, E. S. Hotchkiss, and W. Song. Do buyouts (still) create value? *The Journal of Finance*, 66:479–517, 2011. doi:10.1111/j.1540-6261.2010.01640.x.

K. W. Hanley and Gerard Hoberg. Dynamic interpretation of emerging risks in the financial sector. *The Review of Financial Studies*, 32(12):4543–4603, December 2019. doi:10.1093/rfs/hhz023.

J. Harford. What drives merger waves? *Journal of Financial Economics*, 77: 529–560, 2005.

J. Harford, M. Humphery-Jenner, and R. Powell. The sources of value destruction in acquisitions by entrenched managers. *Journal of Financial Economics*, 106:247–261, 2012.

Tarek Hassan, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4):2135–2202, November 2019. doi:10.1093/qje/qjz021.

Tarek Hassan, S. Hollander, L. V. Lent, and A. Tahoun. The global impact of brexit uncertainty. *Journal of Finance*, 79:413–458, 2024. doi:10.1111/jofi.13293.

Gerard Hoberg and Gordon M. Phillips. Scope, scale and concentration: The 21st century firm (april 1. *Journal of Finance*, 2021. doi:10.2139/ssrn.3746660. https://ssrn.com/abstract=3746660. Available at SSRN forthcoming.

A. H. Huang, Reuven Lehavy, Amy Y. Zang, and Rong Zheng. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, 64(6):2833–2855, 2017. doi:10.1287/mnsc.2017.2751.

A. H. Huang, H. Wang, and Y. Yang. Finbert: A large language model for extracting information from financial text*. *Contemp Account Res*, 40: 806–841, 2023. doi:10.1111/1911-3846.12832.

Narasimhan Jegadeesh and Di Wu. Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3):712–729, 2013. doi:10.1016/j.jfineco.2013.08.018. ISSN 0304-405X.

Fuwei Jiang, Joshua Lee, Xiumin Martin, and Guofu Zhou. Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1):126–149, 2019. doi:10.1016/j.jfineco.2018.10.001. ISSN 0304-405X.

Kyle Jurado, Sydney C. Ludvigson, and Serena Ng. Measuring uncertainty. *American Economic Review*, 105(3):1177–1216, 2015.

Jaewoo Lee and Kwanho Shin. The role of a variable input in the relationship between investment and uncertainty. *American Economic Review*, 90(3): 667–680, 2000.

K. Li, Feng Mai, Rui Shen, and Xinyan Yan. Measuring corporate culture using machine learning. *The Review of Financial Studies*, 34(7):3265–3315, July 2021. doi:10.1093/rfs/hhaa079.

Alejandro Lopez-Lira and Yuehua Tang. Can chatgpt forecast stock price movements? return predictability and large language models (april 6. doi:10.2139/ssrn.4412788. https://ssrn.com/abstract=4412788. Available at SSRN, 2023.

T. Loughran and B. McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66:35–65, 2011. doi:10.1111/j.1540-6261.2010.01625.x.

M. Lowry, Roni Michaely, and Ekaterina Volkova. Information revealed through the regulatory process: Interactions between the sec and companies ahead of their ipo. *The Review of Financial Studies*, 33(12):5510–5554, December 2020. doi:10.1093/rfs/hhaa007.

Nicole Maestas, Kathleen Mullen, and David Powell. The effect of population aging on economic growth, the labor force and productivity, no 22452, nber working papers, national bureau of economic research, inc. 2016.

T. Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *ICLR*, 2013.

Gordon Phillips and Alexei Zhdanov. R&d and the incentives from merger and acquisition activity. *The Review of Financial Studies*, 26(1):34–78, 2013.

Z. Sautner, L. Van Lent, G. Vilkov, and R. Zhang. Firm-level climate change exposure. *J Finance*, 78:1449–1498, 2023. doi:10.1111/jofi.13219.

J. Shermeyer, Thomas Hossler, Adam Van Etten, Daniel Hogan, Ryan Lewis, and Daeil Kim. Rareplanes: Synthetic data takes flight. https://arxiv.org/abs/2006.02963.

D. Wu. Text-based measure of supply chain risk exposure. *Management Science*, 70(7):4781–4801, 2023. doi:10.1287/mnsc.2023.4927.

Y. Xu, Bohao Huang, Xiong Luo, Kyle Bradbury, and Jordan M. Malof. Simpl: Generating synthetic overhead imagery to address zero-shot and few-shot detection problems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 2022.

M. Zhang and Lan Luo. Can consumer-posted photos serve as a leading indicator of restaurant survival? evidence from yelp. *Management Science*, 69(1):25–50, 2022. doi:10.1287/mnsc.2022.4359.

Z. Zhang, Kejia Yang, Jonathan Z. Zhang, and Robert W. Palmatier. Uncovering synergy and dysergy in consumer reviews: A machine learning approach. *Management Science*, 69(4):2339–2360, 2022. doi:10.1287/mnsc.2022.4443.

# Tables

Table 1: Index reconstruction using LLM features vs stock returns

| Index | LLM 24 features | | | LLM 1 feature | | Equal-weighted return | | Value-weighted return | |
|---|---|---|---|---|---|---|---|---|---|
| | Corr | Adj $R^2$ | RMSE | Adj $R^2$ | RMSE | Adj $R^2$ | RMSE | Adj $R^2$ | RMSE |
| VIX | 79 | 45 | 0.62 | 14 | 0.93 | 8 | 0.96 | 13 | 0.93 |
| TEDRATE | 73 | 31 | 0.7 | 8 | 0.97 | 14 | 0.94 | 15 | 0.93 |
| US CCI | 89 | 70 | 0.44 | 22 | 0.86 | -1 | 0.98 | -1 | 0.97 |
| US BCI | 75 | 37 | 0.61 | 7 | 0.88 | 4 | 0.9 | 5 | 0.89 |
| CPI | 70 | 26 | 0.6 | 5 | 0.81 | 4 | 0.82 | 4 | 0.82 |
| EXP INF 1YR | 79 | 45 | 0.39 | 17 | 0.58 | 1 | 0.63 | 0 | 0.63 |
| EXP INF 5YR | 87 | 66 | 0.25 | 35 | 0.42 | 3 | 0.51 | 5 | 0.5 |
| EPU | 88 | 68 | 0.37 | 49 | 0.56 | -1 | 0.79 | -1 | 0.79 |
| ERU | 94 | 84 | 0.39 | 65 | 0.68 | 1 | 1.14 | 0 | 1.14 |
| EMU | 91 | 75 | 0.47 | 46 | 0.83 | -1 | 1.13 | -1 | 1.13 |
| EFU | 83 | 54 | 0.58 | 35 | 0.83 | 0 | 1.04 | 2 | 1.02 |
| recession | 74 | 34 | 0.61 | 14 | 0.84 | 13 | 0.84 | 18 | 0.82 |

Note. Using my LLM model trained on NYT articles, I produce an economy-wide 24-dimensional representation for each quarter by averaging firm-level representations of conference calls. I further regress the indices of interest such as Economic Financial Uncertainty index or Economic Policy Uncertainty index onto the 24 features. This table reports adjusted $R^2$ (%), RMSE and correlations (%) indicating quality of fit of economy-wide indicators using LLMs and conference calls. This performance of using full LLM 24-dimensional representation is compared to using just one best LLM feature, equal-weighted and value-weighted returns. Better fit (higher adjusted $R^2$, correlation and lower RMSE) indicates potential for a more accurate "beta" proxying for firm-level exposure, the approach that I term the "top-down" approach to compute firm-level exposures. LLM features are therefore much more appropriate than returns in constructing "beta" exposures to risks. This Table presents performance metrics between the true value of index in year-quarter $t$, $y_t$, and its contemporaneous prediction $\hat{y_t} = \hat{\alpha} + \hat{\beta} X_t$ where $X_t$ is either the equal- or value-weighted return, 1 feature from my LLM model or all 24 features (in which case both $\hat{\beta}$ and $X_t$ are vectors).

Table 2: My LLM vs Bag of Words (word counting upper bound) in topic detection

| **Topic** | **F1-score** | | | **Average Precision** | | |
|---|---|---|---|---|---|---|
| | **BoW** | **LLM** | **Gain (%)** | **BoW** | **LLM** | **Gain (%)** |
| Appointments and Executive Changes | 0.15 | 0.58 | 280 | 0.06 | 0.37 | 560 |
| Mergers, Acquisitions and Divestitures | 0.4 | 0.78 | 96 | 0.21 | 0.61 | 192 |
| Economic Conditions and Trends | 0.18 | 0.4 | 119 | 0.08 | 0.21 | 152 |
| International Relations | 0.2 | 0.59 | 200 | 0.17 | 0.43 | 152 |
| Accidents and Safety | 0.3 | 0.61 | 105 | 0.16 | 0.4 | 146 |
| Credit | 0.23 | 0.42 | 85 | 0.1 | 0.24 | 142 |
| Diseases and Conditions | 0.32 | 0.63 | 95 | 0.2 | 0.44 | 123 |
| Medicine and Health | 0.43 | 0.75 | 74 | 0.31 | 0.61 | 95 |
| Finances | 0.37 | 0.7 | 89 | 0.29 | 0.55 | 93 |
| Reform and Reorganization | 0.06 | 0.1 | 59 | 0.01 | 0.01 | 88 |
| Labor | 0.32 | 0.55 | 71 | 0.2 | 0.38 | 86 |
| Automobiles | 0.45 | 0.69 | 55 | 0.28 | 0.52 | 86 |
| Elections | 0.57 | 0.82 | 42 | 0.4 | 0.68 | 71 |
| Politics and Government | 0.4 | 0.73 | 85 | 0.37 | 0.63 | 69 |
| Computers and the Internet | 0.48 | 0.7 | 45 | 0.32 | 0.53 | 63 |
| Taxation | 0.46 | 0.57 | 24 | 0.29 | 0.4 | 36 |
| Airlines and Airplanes | 0.27 | 0.34 | 29 | 0.18 | 0.24 | 29 |
| Immigration and Refugees | 0.36 | 0.44 | 23 | 0.21 | 0.26 | 27 |
| Budgets and Budgeting | 0.39 | 0.46 | 19 | 0.24 | 0.3 | 26 |
| Advertising and Marketing | 0.5 | 0.57 | 15 | 0.32 | 0.39 | 20 |

Note. NYT Annotated Corpus includes over 1mln articles with topics discussed hand-labelled by library experts. In this Table I compare performance in topic detection of my LLM and a bag-of-words (BoW) logistic regression out-of-sample using 2002-2004 NYT articles as test set (168,119 articles in total) and 1987-1999 articles for training (944,488 articles). I choose the top 20 economics-related topics by number of articles featuring each topic. Bag-of-words performance is an upper bound for word counting commonly used in finance because unlike word counting it allows for a flexible, data-driven keyword list and each word is assigned an optimal weight rather than being equal-weighted. A separate BoW model is trained for each topic. The measures compared are F1-score (a harmonic average of precision, the fraction of articles labelled by a model as discussing a topic actually covering that topic, and recall, fraction of all articles actually featuring a topic detected by a model) and Average Precision (area under the Precision-Recall curve). The entries are sorted by the last column, the gain in Average Precision from using my LLM model over BoW. Average gain from my LLM model over BoW is 81% and 113% in terms of F1-score and Average Precision respectively.

Table 3: My LLM vs Meta Llama-3 70B in topic detection

| **Topic** | **Accuracy** | | | **Average Precision** | | |
|---|---|---|---|---|---|---|
| | **Llama-3** | **LLM** | **Gain (%)** | **Llama-3** | **LLM** | **Gain (%)** |
| Accidents and Safety | 0.94 | 0.98 | 5*** | 0.55 | 0.77 | 40 |
| Advertising and Marketing | 0.95 | 0.97 | 3*** | 0.67 | 0.58 | -13 |
| Airlines and Airplanes | 0.97 | 0.97 | 0 | 0.46 | 0.49 | 7 |
| Appointments and Executive Changes | 0.91 | 0.98 | 8*** | 0.57 | 0.63 | 11 |
| Automobiles | 0.97 | 0.99 | 1*** | 0.73 | 0.78 | 7 |
| Budgets and Budgeting | 0.91 | 0.97 | 6*** | 0.62 | 0.67 | 7 |
| Computers and the Internet | 0.96 | 0.98 | 1*** | 0.72 | 0.74 | 3 |
| Credit | 0.96 | 0.97 | 1*** | 0.53 | 0.63 | 18 |
| Diseases and Conditions | 0.96 | 0.98 | 2*** | 0.72 | 0.78 | 8 |
| Economic Conditions and Trends | 0.78 | 0.96 | 24*** | 0.37 | 0.59 | 59 |
| Elections | 0.97 | 0.98 | 1*** | 0.79 | 0.86 | 9 |
| Finances | 0.84 | 0.93 | 11*** | 0.64 | 0.86 | 35 |
| Immigration and Refugees | 0.98 | 0.98 | 0 | 0.79 | 0.65 | -18 |
| International Relations | 0.91 | 0.95 | 5*** | 0.53 | 0.70 | 31 |
| Labor | 0.94 | 0.96 | 2*** | 0.62 | 0.68 | 10 |
| Medicine and Health | 0.93 | 0.97 | 4*** | 0.72 | 0.81 | 12 |
| Mergers, Acquisitions and Divestitures | 0.98 | 0.98 | 1*** | 0.80 | 0.85 | 6 |
| Politics and Government | 0.75 | 0.92 | 23*** | 0.56 | 0.83 | 48 |
| Reform and Reorganization | 0.85 | 0.97 | 14*** | 0.27 | 0.32 | 17 |
| Taxation | 0.97 | 0.98 | 1** | 0.74 | 0.68 | -8 |

Note. NYT Annotated Corpus includes over 1mln articles with topics discussed hand-labelled by library experts. Meta Llama-3 70b is comparable to other state-of-the-art proprietary LLMs such as ChatGPT on most tasks. In this Table I compare performance in topic detection of my LLM and Meta Llama-3 70b out-of-sample using a random sample of 4000 NYT articles published between 2002 and 2004 as test set (100 articles per topic for 20 topics and 2000 articles on unrelated topics). I choose the top 20 economics-related topics by number of articles featuring each topic. Llama-3 is asked to give a number between 0 and 1 quantifying the extent to which a topic is relevant in an article. The measures compared are Accuracy and Average Precision (area under the Precision-Recall curve). Significance of gains in Accuracy is confirmed using the McNemar's test. Average gain from my LLM model over Meta Llama-3 70b is 14%, 17% and 6% in terms of F1-score, Average Precision and Accuracy respectively.

Table 4: Differences-in-Differences: Geopolitical risk and M&A decisions

| Variable | acquirer 2012-2016 | target 2012-2016 | acquirer 2003-2007 | target 2003-2007 |
|---|---|---|---|---|
| *constant* | 0.20*** | 0.00*** | 0.20*** | 0.00*** |
| | 34.9 | 3.3 | 51.3 | 4.8 |
| $Post \times Treat$ | 0.044*** | -0.004*** | 0.030 | -0.000 |
| | 3.1 | -3.8 | 0.5 | -0.01 |
| $Treat$ | -0.035** | 0.003** | -0.015 | -0.002 |
| | -2.0 | 2.4 | -0.7 | -0.6 |
| Firm Fixed Effects | Yes | Yes | Yes | Yes |
| Time Fixed Effects | Yes | Yes | Yes | Yes |
| N | 2,231 | 2,231 | 1,191 | 1,191 |

Note. On March 6, 2014, President Obama imposed sanctions against Russia in response to annexation of Crimea by the Russian Federation precluding U.S. companies from trading and collaborating with major Russian businesses. Using my Large Language Model I estimate the extent to which each firm was exposed to geopolitical risk in 2014 and assign top 20% of firms in terms of this measure to treatment and bottom 20% to potential control. For each treated firm I further find one matching control firm by matching exactly on year, industry (sic2) and credit rating and minimizing the deviation in terms of Tobin's Q, cashflow, size and cash holdings. I consider 2 years before and after the event (2012-2016) with either being an acquirer or a target as the outcome. Additionally, I repeat the experiment using 2003-2007 window with 2005 as a placebo event. All regressions include firm and year fixed effects. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. t-statistics are derived from Driscoll-Kraay standard errors that are robust to cross-sectional and temporal dependence.

Table 5: Acquisition likelihood and LLM exposures

| | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
| | Coef | ME (%) | Coef | ME (%) | Coef | ME (%) |
| geopolitical risk | 0.045*** | 3.5 | | | | |
| | 2.63 | | | | | |
| financial uncertainty | | | -0.079*** | -6.1 | | |
| | | | -4.71 | | | |
| EPU | | | | | -0.040** | -3.1 |
| | | | | | -2.43 | |
| Firm-level controls | Yes | | Yes | | Yes | |
| Industry-level controls | Yes | | Yes | | Yes | |
| Economy-level controls | Yes | | Yes | | Yes | |
| Firm fixed effects | Yes | | Yes | | Yes | |
| Time fixed effects | Yes | | Yes | | Yes | |
| N | 20,028 | | 20,028 | | 20,028 | |

Note. Using the proposed Large Language Model (LLM) pipeline trained on NYT articles, I produce firm-level measures of geopolitical risk, financial uncertainty and Economic Policy Uncertainty (EPU) using conference calls. This table reports logistic regression coefficients and corresponding marginal effects (ME) modeling probability of being an acquirer next year conditional on controls and a measure of interest. All regressions include firm, time fixed effects and a barrage of firm-, industry- and economy-level controls. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. Standard errors are robust and clustered at the firm-year level. Each regression has 20,028 observations.

Table 6: Target likelihood and LLM exposures

| | (1) | | (2) | |
|---|---|---|---|---|
| | Coef | ME (%) | Coef | ME (%) |
| geopolitical risk | -0.283*** | -58.1 | | |
| | -2.46 | | | |
| economic uncertainty | | | -0.147** | -32.2 |
| | | | -2.25 | |
| Firm-level controls | Yes | | Yes | |
| Industry-level controls | Yes | | Yes | |
| Economy-level controls | Yes | | Yes | |
| Firm fixed effects | Yes | | Yes | |
| Time fixed effects | Yes | | Yes | |
| N | 20,028 | | 20,028 | |

Note. Using the proposed Large Language Model (LLM) pipeline trained on NYT articles, I produce firm-level measures of geopolitical risk and economic uncertainty using conference calls. This table reports logistic regression coefficients and corresponding marginal effects (ME) modeling probability of being a target of an acquisition next year conditional on controls and a measure of interest. All regressions include firm, time fixed effects and a barrage of firm-, industry- and economy-level controls. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. Standard errors are robust and clustered at the firm-year level. Each regression has 20,028 observations.

Table 7: Acquisition likelihood, supply chains and labor shortage

| | (1) | | (2) | |
|---|---|---|---|---|
| | Coef | ME (%) | Coef | ME (%) |
| RUCH_SC_share × GeopRisk | 0.13** | 9.5 | | |
| | 1.98 | | | |
| GeopRisk | 0.04 | -4.4 | | |
| | -1.89 | | | |
| RUCH_SC_share | -0.06* | 2.8 | | |
| | 4.33 | | | |
| LaborShortage × PopAging | | | 0.34** | 27 |
| | | | 1.98 | |
| PopAging | | | -0.16* | -13 |
| | | | -1.89 | |
| LaborShortage | | | 0.21*** | 16 |
| | | | 4.33 | |
| Firm-level controls | Yes | | Yes | |
| Industry-level controls | Yes | | Yes | |
| Economy-level controls | Yes | | Yes | |
| Firm fixed effects | Yes | | Yes | |
| Time fixed effects | Yes | | Yes | |
| N | 20,028 | | 20,028 | |

Note. Using the proposed Large Language Model (LLM) pipeline trained on NYT articles, I produce firm-level measures of geopolitical risk, labor shortage and population aging using conference calls. This table reports logistic regression coefficients and corresponding marginal effects (ME) modeling probability of being an acquirer next year conditional on controls and measures of interest with RUCH_SC_share representing share of Russia and China in supply chain of firm $i$ in year $t$. All regressions include firm, time fixed effects and a barrage of firm-, industry- and economy-level controls. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. Standard errors are robust and clustered at the firm-year level. Each regression has 20,028 observations.

Table 8: Geopolitical risk channels: real option and operational hedge

| | Second-stage Heckman Sample Selection Model | |
|---|---|---|
| | Target Capital Intensity | Vertical M&A |
| geopolitical risk | 0.03*** | 0.07*** |
| | 4.8 | 5.8 |
| Firm-level controls | Yes | Yes |
| Industry-level controls | Yes | Yes |
| Economy-level controls | Yes | Yes |
| Firm fixed effects | Yes | Yes |
| Time fixed effects | Yes | Yes |
| N | 20,028 | 20,028 |

Note. Using the proposed Large Language Model (LLM) pipeline trained on NYT articles, I produce a firm-level measure of geopolitical risk using conference calls. This table reports the second-stage Heckman Sample Selection Model regression coefficients for geopolitical risk in explaining target's degree of investment irreversibility proxied for with the capital intensity ratio and the probability of a merger being vertical (versus not vertical). All regressions include firm, time fixed effects and a barrage of firm-, industry- and economy-level controls. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. Standard errors are robust and clustered at the firm-year level. Each regression has 20,028 observations.

Table 9: Climate change, structural change and M&A outcomes

| | Buyout Premium | CAR -3_3d ff4 |
|---|---|---|
| climate change acquirer | -16.3* | |
| | -1.7 | |
| climate change target | -23.3** | |
| | -2.4 | |
| structural change target | | -0.04** |
| | | -2.4 |
| Firm-level controls | Yes | Yes |
| Industry-level controls | Yes | Yes |
| Deal-level controls | Yes | Yes |
| Industry fixed effects | Yes | Yes |
| Time fixed effects | Yes | Yes |
| N | 76 | 76 |
| $R^2$ within (%) | 65 | 50 |

Note. Using the proposed Large Language Model (LLM) pipeline trained on NYT articles, I produce a firm-level measure of climate change using conference calls. In learning to produce topic probabilities, the model constructs a 24-dimensional representation of each conference call. I use Euclidean distance between consecutive conference call representations for each firm as a measure of structure change. This table reports the coefficients for climate change measures for acquirer and target in explaining buyout premia and the coefficient for target's structural change in relation to market's reaction measured as announcement abnormal returns. All regressions include industry, time fixed effects and a barrage of firm-, industry- and deal-level controls. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. Standard errors are robust and clustered at the firm-year level.

# Figures

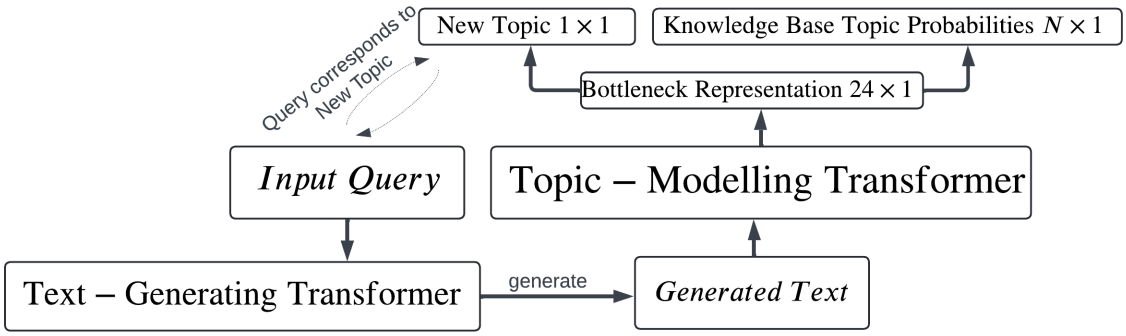



Figure 1: Full Proposed Large Language Model Methodology

Note. The proposed pipeline combines two transformer-based networks – a topic-modelling transformer based on the Longformer architecture, and a text-generating transformer based on Meta Llama 3, the open source ChatGPT alternative. The topic-modelling transformer is first trained on the annotated New York Times corpus producing probabilities each topic is present in an article. Given a new topic of interest (such as "financing constraints" or "cryptocurrencies"), the text generator constructs a corpus of synthetic NYT-style articles on the topic in question. Topic-modelling transformer is further trained on this synthetic corpus and a small sample from the Knowledge Base (the original NYT corpus) learning to distinguish the new topic and produce probabilities that the topic is present. The degree to which the model is certain about the topic being present is the degree of exposure of the text passage or article to the topic of interest. The newly trained topic-modelling transformer can be further applied to any text passage that works best for the application at hand – be it company websites, newspaper articles, conference calls or 10-Ks.

Figure 2: Python Text Generation Prompt Format

```
messages_titles = [
{"role": "system", "content": "You are a New York Times writer."},
{"role":"user","content":"Suggest {} different titles
for an article on {}.".format(str(N),topic)}]

messages=[
{"role": "system", "content": "You are a New York Times writer."},
{"role": "user", "content": 'Write an article on {}
with the following title: {}.'.format(topic,title)}]
```

Note. The two messages sent to the Meta Llama 3 text generator to construct synthetic New York Times articles on the topic of interest in python. The first message asks the LLM to generate a list of N titles for an article on the topic of interest. This guides synthetic article generation allowing for a more diverse generated corpus. The second message is iterated over all titles previously returned by the model and asks that the LLM to generate a synthetic article on the topic of interest with the specific title in mind. The variable "topic" stores the string of interest such as "cybersecurity risks" or "climate change".

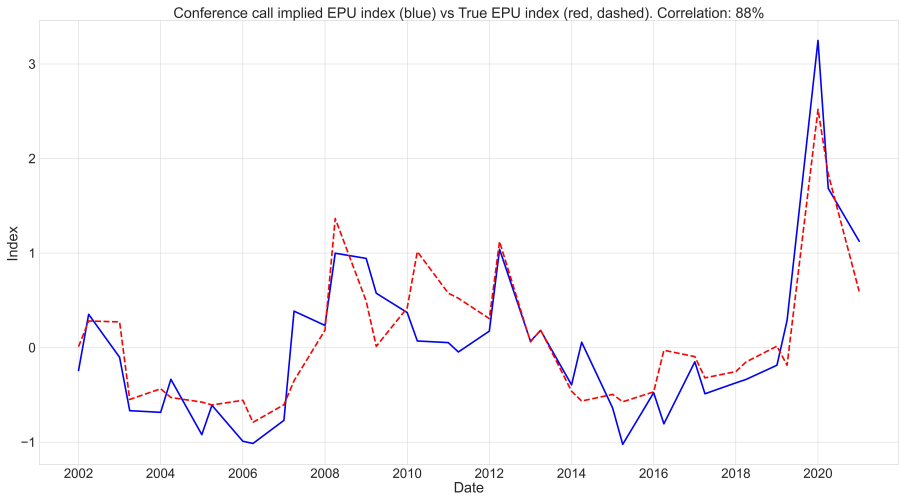



Figure 3: Conference Call implied EPU index against true EPU index

Note. Using the proposed Large Language Model pipeline, I produce economy-wide indices and compare them with true index values. First, I take a quarterly average of the firm-level 24-dimensional representations learned by my model constructing economy-level textual representation. I further regress the index of interest, Economic Policy Uncertainty (Baker et al. (2016)) in this case, onto the constructed economy-level representations. This way I find what combination of textual features reconstructs EPU at the economy-level and can further apply this combination to individual firm-level conference calls to estimate their firm-level exposure to EPU. I term this approach to estimating firm-level exposures as the "top-down" approach. The textual features learned by my model are able to reproduce the EPU index remarkably well.

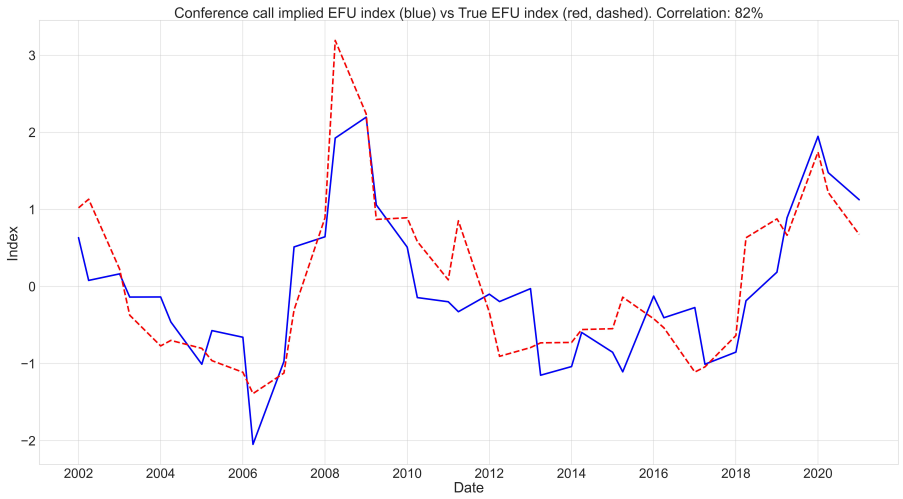



Figure 4: Conference Call implied EFU index against true EFU index

Note. Using the proposed Large Language Model pipeline, I produce economy-wide indices and compare them with true index values. First, I take a quarterly average of the firm-level 24-dimensional representations learned by my model constructing economy-level textual representation. I further regress the index of interest, Economic Financial Uncertainty (Jurado et al. (2015)) in this case, onto the constructed economy-level representations. This way I find what combination of textual features reconstructs EFU at the economy-level and can further apply this combination to individual firm-level conference calls to estimate their firm-level exposure to EFU. I term this approach to estimating firm-level exposures as the "top-down" approach. The textual features learned by my model are able to reproduce the EFU index remarkably well.

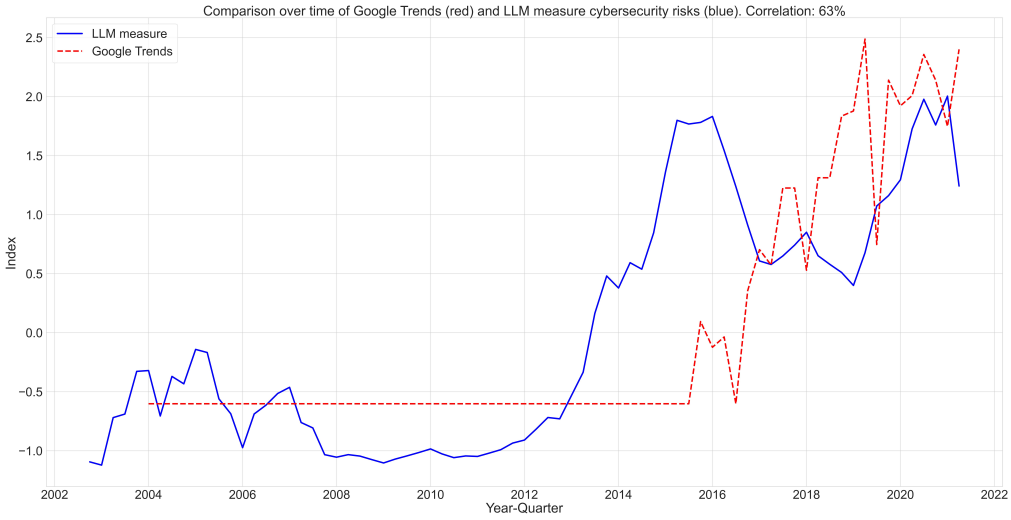



Figure 5: LLM "cybersecurity risks" measure against Google Trends

Note. I apply my LLM model trained on NYT articles to earnings conference calls to produce probabilities of each topic of interest. I term this approach to estimating firm-level exposures as the "bottom-up" approach. I further aggregate the measure corresponding to "cybersecurity risks" to economy-level by taking a 1-year moving average. Comparison is against the Google Trends series for online searches since 2004.

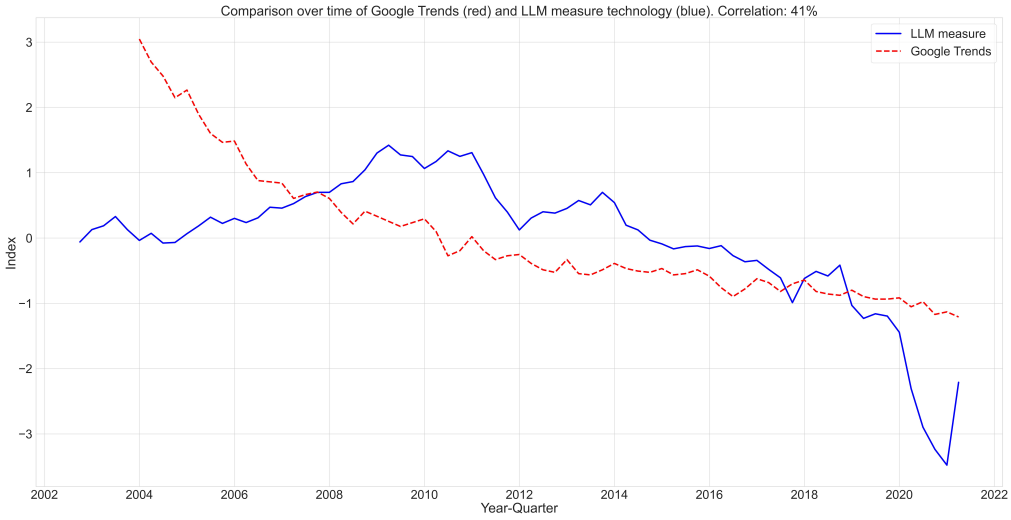



Figure 6: LLM "technology" measure against Google Trends

Note. I apply my LLM model trained on NYT articles to earnings conference calls to produce probabilities of each topic of interest. I term this approach to estimating firm-level exposures as the "bottom-up" approach. I further aggregate the measure corresponding to "technology" to economy-level by taking a 1-year moving average. Comparison is against the Google Trends series for online searches since 2004.

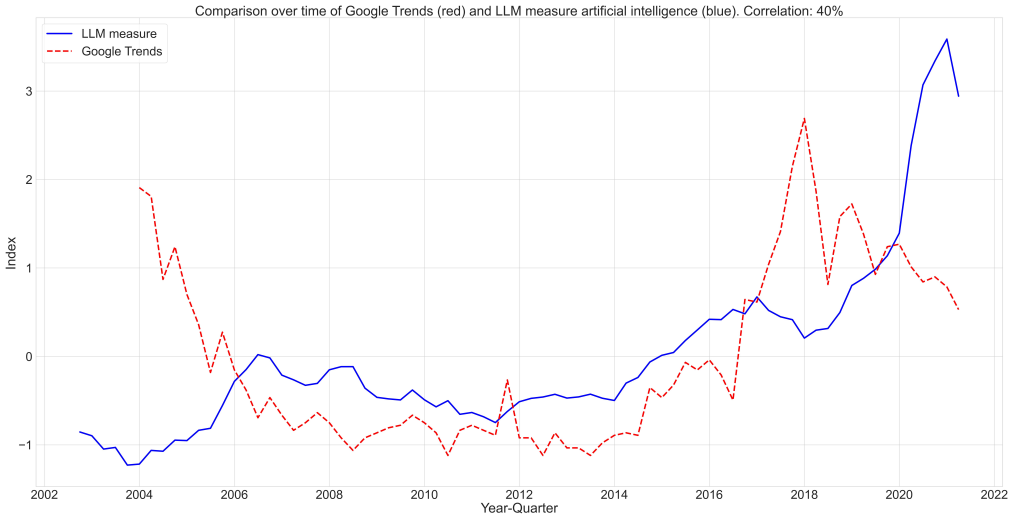



Figure 7: LLM "artificial intelligence" measure against Google Trends

Note. I apply my LLM model trained on NYT articles to earnings conference calls to produce probabilities of each topic of interest. I term this approach to estimating firm-level exposures as the "bottom-up" approach. I further aggregate the measure corresponding to "artificial intelligence" to economy-level by taking a 1-year moving average. Comparison is against the Google Trends series for online searches since 2004.

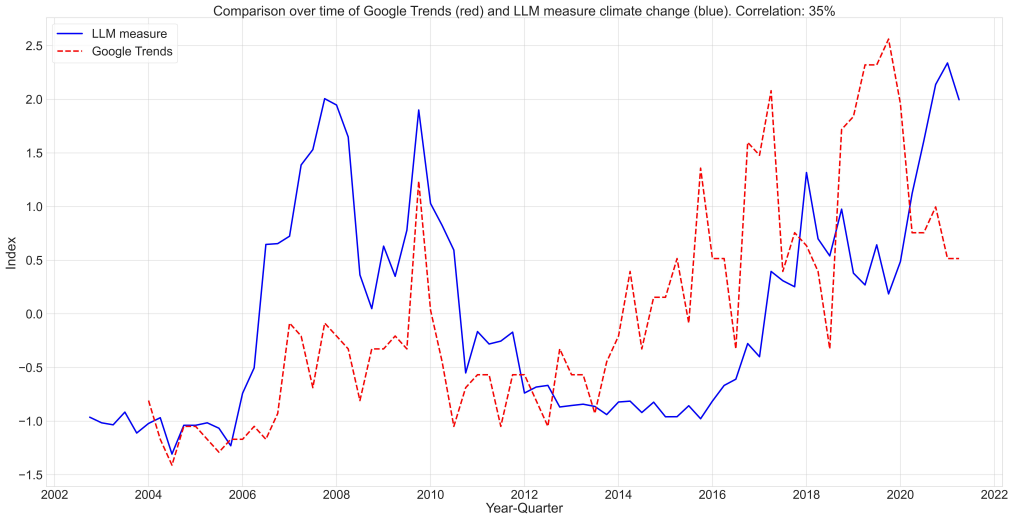
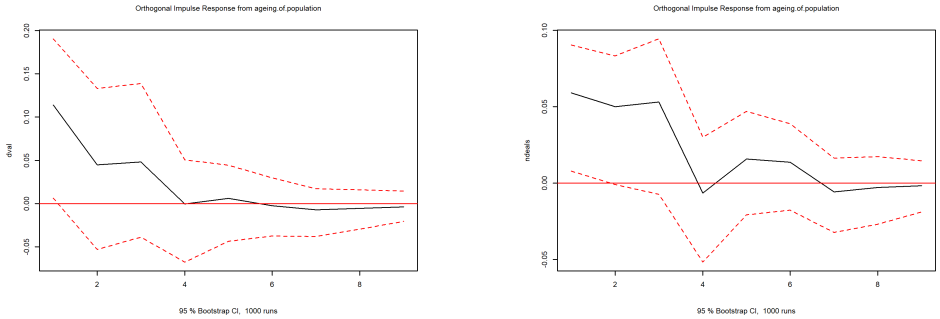



Figure 8: LLM "climate change" measure against Google Trends

Note. I apply my LLM model trained on NYT articles to earnings conference calls to produce probabilities of each topic of interest. I term this approach to estimating firm-level exposures as the "bottom-up" approach. I further aggregate the measure corresponding to "climate change" to economy-level by taking a 1-year moving average of quarterly measures. Comparison is against the Google Trends series for online searches since 2004.

Figure 9: Impulse Response from shock to Population Aging to M&A activity

(a) Aggregate M&A deal value (b) Aggregate number of deals




Note. Left panel depicts response to aggregate deal values, right panel – to number of deals. The red dashed lines mark the 95% confidence band. Using the proposed LLM pipeline, I construct a quarterly firm-level exposure to population aging. The quarterly average of this measure across firms is used as the impulse (the shock variable) in a VARX(2) model.

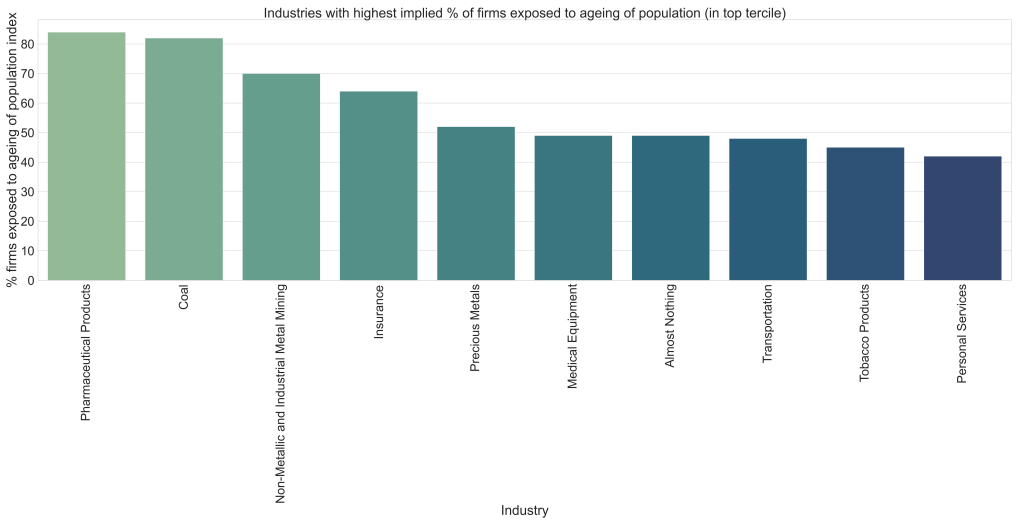



Figure 10: Industries with the highest exposure to LLM Population Aging measure

Note. Using the proposed LLM pipeline, I construct a quarterly firm-level exposure to population aging. Firms in the top tercile of the population aging measure are flagged as highly exposed. Here I plot the percentage of firms highly exposed to population aging for 10 Fama-French 48 industries with the highest percentage.