2023

NORWEGIAN SCHOOL OF ECONOMICS

BAN432 – APPLIED TEXTUAL DATA ANALYSIS FOR BUSINESS AND FINANCE

EXAM 2023

CANDIDATE NUMBERS: 108, 118, 145, 154

Table of Contents

Introduction

The release of ChatGPT and the development of generative artificial intelligence in general has marked a pivotal shift in the technological landscape. In this report we will present our findings from a thorough analysis of which industries the release of ChatGPT has affected, and what types of risks firms face regarding AI. The analysis will be based on 7000+ earnings call transcripts between October 2021 and August 2023, allowing us to observe industry dialogues before and after the release of ChatGPT across sectors.

Initial Exploration of data

To familiarize ourselves with the dataset, we conducted a range of initial inspection procedures. We inspected the structure of the files and created various visualizations. We started by calculating the total number of documents per sector for the entire time-period, displayed in Table 1. Due to the lack of data regarding industries in the dataset, we define sectors as industries in this analysis to be able to answer the tasks provided.

Sector	**Count**
Health-Care	1404
Information-Technology	1307
Industrials	1229
Financials	1107
Consumer-Discretionary	872
Materials	417
Communication-Services	362
Energy	322
Consumer-Staples	297
Utilities	115
Real-Estate	110
Electric-Utilities	58
Gas-Utilities	16
Diversified-Utilities	14
Water-Utilities	9
Trucking	3
Closed-End-Fund-Debt	1

Table 1 - Number of documents per sector

From Table 1, we observe that the number of documents is quite varied between the sectors. Healthcare stands out as the sector with the most documents, while various Utility sectors, Trucking and Close-End-Fund-Debt have less than 100 documents. Next, we plotted the number of documents for each sector over time, displayed in figure 1.

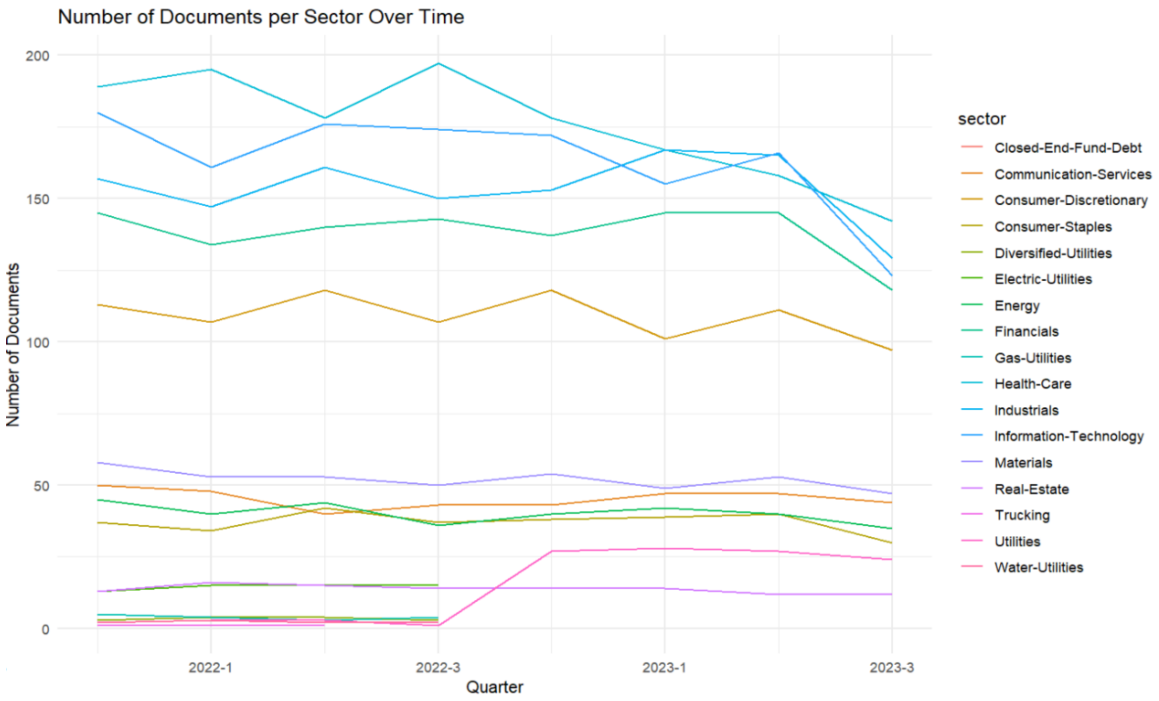


Figure 1 - Number of documents per sector over time

From this graph we observe that most of the sectors have quite a steady number of documents over time. However, there is not an exact equal number of documents per sector in each quarter. This will impact our analysis, as we will have to consider the fact that not all sectors have the same number of documents for each quarter. There can be observed a slight dip in some of the sector's number of documents towards the last quarter, which could be because not all earnings calls for the third quarter of 2023 was published at the time the data was downloaded (August 2023). From the graph we observe that some of the sectors, namely gas-, water-, electrical- and diversified-utilities, only have data available up until the third quarter of 2022. Upon further inspection we observed that the dataset grouped all utilities together after the third quarter of 2022, which can explain the upswing in the number of utilities documents.

Preprocessing

To preprocess the dataset we were provided with, we first converted the text from html to pure text and transferred the text into a data frame. We also appended a column indicating quarter and year in which each earnings call was published.

In addition to these preprocessing procedures, we consolidated all utility-based industry categories, such as gas and electrical utilities, into one broader category named "Utilities". This change was made based on our observation that all utilities were grouped together after the third quarter of 2022. To ensure a consistent approach throughout all the quarters, we adjusted the first part of the dataset to reflect this categorization.

Furthermore, we made the decision to exclude two sectors, namely 'Truckers' and 'Closed-End-Fund-Debt', due to the low number of earnings calls in these sectors. This step maintains the integrity and reliability of our analysis, as a low number of data points in these sectors could lead to unrepresentative results. By focusing on sectors with more data points, our analysis becomes more accurate and gains higher quality.

We also decided to split the transcripts into two parts, a management section, and a Q&A section. This was done to ensure that we conduct analysis on the parts of the transcripts which were relevant for the individual tasks.

Task 1: Which industries are affected by the release of ChatGPT?

Roadmap

In this task, we will perform textual analyses to answer the question: *"Which industries are affected by the release of ChatGPT?"*. We will start by presenting our methodological choices, followed by our method for keyword selection in relation to a Keyword in Context (KWIC) analysis. We then perform a word frequency analysis using our curated list of keywords, where we present a line chart of the average frequency of keywords per document per sector across the quarters in the dataset. To wrap up our analysis, we conduct a Wilcoxon Test to see if there has been a significant increase in the keyword frequency per sector from before the release of ChatGPT compared to after. We also supplement with an overview of the difference in the average keyword frequency across sectors, comparing periods before and after the release of ChatGPT. Based on the findings from these analyses, we provide a conclusion that addresses the research question at hand.

Methodological Choices

For this task we decided to use the management discussion section of the transcripts. Our rationale is to base our findings on the firms' own presentations, ensuring that we only include effects explicitly stated by the companies themselves. A potential limitation of this approach is the exclusion of valuable insights that may present itself during the Q&A sessions of the calls. Nonetheless, we believe that it is crucial to avoid potential biases introduced by the questions asked in these sessions. The decision to exclude Q&A content was made because we believe that the benefits of a more controlled analysis outweigh the risk of overlooking additional information.

Assumptions

We operate under the assumption that a substantial increase in the average frequency of keyword mentions in the management section of the earnings calls within a sector, before the release of ChatGPT compared to after, serves as an indicator that the relevant sector was affected by the release of ChatGPT. This assumption is based on the fact that the firms present elements of importance for the business and things that typically affect the operations in the company in this section of the transcript. Furthermore, if the keywords related to the circumstances around ChatGPT is discussed in this section it would indicate that it affects the industry.

Keywords in Context

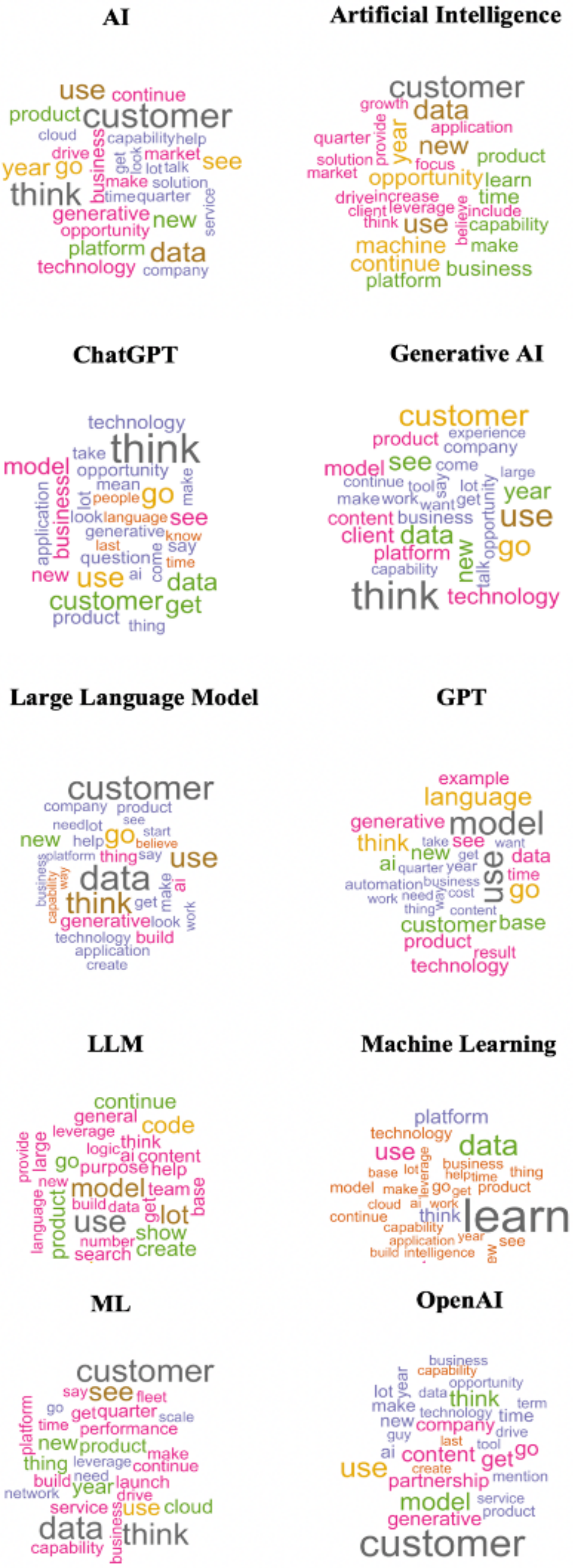
Based on an analysis of which firms are using the term *ChatGPT* in the transcript, we found that Microsoft stood for around 15% of the use of this term, and that the overall frequency was low. Due to this, we found it more convincing to use a broader specter of keywords to determine which sectors were affected by the release of ChatGPT. To identify related terms, we started by conducting a KWIC analysis using "chatgpt" as pattern. Here we used regex to make sure we could catch all forms of formatting of the phrase. We then combined the context into a single paragraph, which was cleaned and lemmatized. We then created bigrams from these results, which is displayed in table 2.

Word 1	Word 2	Count
artificial	intelligence	11
machine	learn	11
language	model	9
azure	openai	8
customer	partner	7
drive	growth	7
generative	ai	7

Table 2 - Related bigrams to ChatGPT

Based on these bigrams we see that artificial intelligence is the most frequently used bigram, followed my machine learning, language model, and azure openai. This tells us that these phrases are commonly used in the same context as ChatGPT. Based on this information we created our initial keyword list which we will use in our KWIC analysis: "machine learning", “gpt”, “large language model”, “ml”, “ai”, “artificial intelligence”, “chatgpt”, “openai”, “generative ai” and “llm”. Here we also used regex to account for different formatting of the phrases.

After creating our initial keyword list, we moved on to further analyze these words to ensure they were used in relevant contexts. We started by conducting a KWIC analysis, and generated word clouds for each keyword based on this, only including nouns, adjectives, and verbs, to assess relevance of said words. These word clouds are displayed in Figure 2. Our word clouds show words like “costumer”, “data”, “capability”, “performance” and “opportunity”. These words indicate that our keywords are mentioned in context where influence on the companies is discussed, and we therefore choose to include all keywords in the further analysis.



AI

Artificial Intelligence

ChatGPT

Generative AI

Large Language Model

GPT

LLM

Machine Learning

ML

OpenAI

Figure 2 - Wordclouds for terms related to ChatGPT

Word frequency analysis

With a relevant keyword list, we moved on to conducting a word frequency analysis.

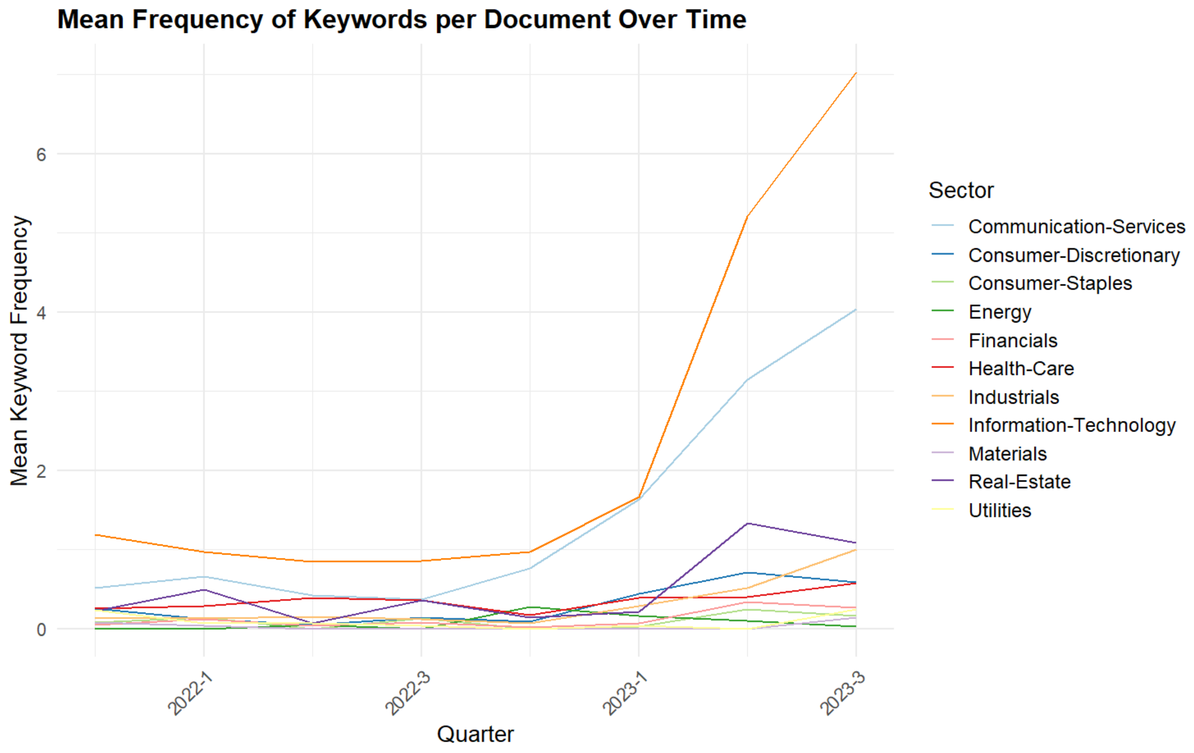


Figure 3 - Mean frequency of keywords per sector over time

Our graph of mean frequency of keywords per document over time, displayed in figure 3, shows a noteworthy upswing in the use of our chosen keywords for several of the sectors after the release of ChatGPT. Before the release, which was in the fourth quarter of 2022, the keywords were mentioned relatively rarely across every sector. Every sector has an average frequency of below 1 across every quarter pre-release. However, Information-Technology seems to be mentioning them the most. After the release, we observe that this sector, along with Communication-Services, show the biggest increase in the use of keywords. Other sectors, such as Real-Estate, Industrials and Consumer-Discretionary also seem to have had a slight increase in the mention of our keywords after the release. Determining whether the increase in keyword mentions is significant across the different sectors requires a statistical analysis, which we will conduct in the next section of this report.

Table 3 below shows the average keyword mention frequency per document both before and after the release for every sector. It also displays the average increase of mentions per sector. Here we can observe that the sectors Information-Technology and Communication-Services have had the largest average increase in use of keywords, matching what we saw in figure 3. All other sectors have had an increase of below 1.

To establish whether these increases can be regarded as statistically significant increases, we have conducted a Wilcoxon Rank Sum Test. Here we had the following null hypothesis: "None of the sectors show a significant increase in the average use of keywords per document". In the test calculation we used the absolute value of keyword mentions per sector both before and after the release of ChatGPT. The p-values for the different sectors are displayed in the right column of the table below, with green numbers for significant values and red for insignificant values. The sectors Information-Technology, Communication-Services, Financials, Industrials and Consumer-Discretionary show significant increases in the use of keywords with a 95% significance level.

Sector	Mean Frequency Before 2022 Q4	Mean Frequency After 2022 Q4	Difference	P-value
Information-Technology	0.9710313	4.4797297	3.5086984	0.0000000
Communication-Services	0.5535714	2.9202899	2.3667184	0.0000193
Financials	0.0658083	0.2254902	0.1596819	0.0047880
Industrials	0.1250000	0.5683297	0.4433297	0.0107010
Consumer-Discretionary	0.1314387	0.5857605	0.4543218	0.0128740
Materials	0.0261194	0.0469799	0.0208605	0.1144544
Real-Estate	0.2638889	0.8421053	0.5782164	0.1617782
Consumer-Staples	0.0691489	0.1467890	0.0776401	0.1979859
Energy	0.0634146	0.1025641	0.0391495	0.2272782
Health-Care	0.2956243	0.4518201	0.1561958	0.5877604
Utilities	0.0827068	0.0886076	0.0059008	0.8434475

Table 3 - Difference in frequency before/after 2022 Q4 and p-value from Wilcoxon Test (per sector)

Results and analysis

Based on this analysis, the Information-Technology and Communication-Services sectors seem to be the most influenced by the release of ChatGPT, followed by Financials, Industrials and Consumer-Discretionary. This conclusion is based on our assumption that a significant increase in the frequency of mentions of our ChatGPT-related keywords indicate that the relevant firms notice an effect on their operational processes and business strategy as a consequence of the release of ChatGPT.

We believe the effect on these industries can be explained by the fact that the introduction of ChatGPT can change the way companies in these sectors work. They might update their digital structure, use AI to make new products, improve customer service with AI help, and/or manage data more effectively.

In contrast, other sectors, whose businesses are less reliant on or connected to ChatGPT, may experience less impact. The core nature of their business operations seems more independent of AI, showing they're less likely to be affected by the release of ChatGPT. It can also be the case that the new technology is not yet suited for use in different industries. For instance, the use of AI in the health-sector might introduce many ethical and legal complications that will require new regulations.

The release of ChatGPT seems to have increased the interest in AI technologies in general. Therefore, the effect on industries from ChatGPT's release most likely extends beyond just its specific functionalities. Other AI-driven technologies may have gained more attention after the release due to the media attention around ChatGPT, which can be a reason for the observed increase in the frequency of AI-related keywords in earning calls.

Limitations of the analysis

Our analysis has certain limitations that should be considered. Firstly, we have not examined whether the observed increase in average keyword mentions is driven by a single dominant company in an industry or is a broader trend across the entire industry. If the former is true, it’s most likely not sufficient to conclude that the entire industry as a whole was affected by the release.

Furthermore, there is a possibility that we might have overlooked some crucial keywords in our KWIC analysis, or included some that do not give our analysis any additional value. This could interfere with the credibility of our analysis, considering that the keyword selection process in the KWIC analysis is crucial in laying the groundwork for the main portion of this task's analysis.

We also operate under the assumption that a significant increase in the average frequency of keyword mentions in the earnings calls within a sector, before the release of ChatGPT compared to after, signifies that the industry has been impacted by the release. If this assumption turns out to be incorrect, it would substantially weaken the validity of our analysis.

As mentioned earlier in the analysis, we focused exclusively on the management discussion section of the transcript for this part of the analysis, prioritizing data that comes directly from the companies. This method may overlook valuable insights from the Q&A sessions, which could serve as a weakness in the analysis.

Task 2: What are the risks firms face regarding AI?

Roadmap

In this task, we will perform textual analyses to answer the question: *"What are the risks firms face regarding AI?"*. We will start by presenting our methodological choices, followed by our method for extracting negative segments, and analyzing co-occurrence of words and bigrams. To wrap up our analysis we will present two network models of co-occurrence and interpret the result from this.

Methodological Choices

In our analysis, we have chosen to use the entirety of the earnings call transcripts, including the Q&A segments. We base this decision on the belief that the Q&A sessions may give critical insights, particularly in relation to risks associated with AI that are not disclosed in the management section. Earnings calls can include a degree of corporate optimism about future development, which may lead to an underrepresentation of potential challenges. The interactive nature of Q&A sessions, however, can prompt a discussion of risks and considerations that might otherwise remain unaddressed.

Keywords in Context

For this task we decided to continue with the sectors identified in task 1. This was done to eliminate "negative noise" from the other sectors. We then used the same set of keywords as found through our analysis in task 1 to extract context around AI from the documents. Following this, we extracted the results of the KWIC analysis to separate paragraphs. These paragraphs were then cleaned using standard preprocessing steps, like removing punctuations, numbers, stopwords and whitespace as well as lemmatizing the words. We also decided to filter out duplicates from sentences that contained multiple keywords to avoid overrepresentation in our dataset, ensuring a balanced and accurate frequency analysis. These preprocessing procedures ensure that the following analysis will operate on the core context of the text, excluding irrelevant or redundant elements.

Sentiment Analysis

The next step was to extract the paragraphs where AI was mentioned in a negative setting. To do this we started by analyzing the frequency of negative and positive words in our data defined by the Loughran-Mcdonald sentiment dictionaries. After an initial look at the word frequencies, we found that two of the most common negative sentiment words used in our data was "question" and "ill". We chose to remove these words from our dictionary as "question" isn't a negative word in the context of a Q&A section, and ill stems from "I'll" after the preprocessing we did previously. From here we calculated the negative and positive sentiment for our paragraphs and calculated a Term Frequency-Inverse Document Frequency (TF-IDF) score to weigh importance of negative words. We then created a list of negative keywords we felt might be used in the context of a risk based on negative words with a high TF-IDF score.

Term	Frequency	Doc.freq	TF-IDF score
critical	119	114	0.00255
challenge	97	94	0.00219
loss	74	55	0.00191
threat	48	38	0.00135
problem	47	46	0.00127
fraud	40	32	0.00117
force	35	34	0.00101
decline	24	21	0.000761
crime	22	18	0.000718
late	22	22	0.000692
prevention	20	19	0.000646

Table 4 - Important negative words

We constructed the following list of words: Risk, problem, concern, threat, challenge, critical, breach, vulnerability, and fraud.

After this initial exploration of the AI contexts, we extracted paragraphs that included one or more of the negative keywords in our list, and where the negative sentiment was dominant.

Word Network

The next step was to analyze the topics of these negative paragraphs. For this we explored words that tend to co-occur when mentioning AI in a negative setting. This was done by calculating pairwise correlation of terms. To extract the most important context we also limited our calculation to adjectives and nouns. We then filtered out words that appear more than 7 times to ensure that the words are somewhat frequently mentioned. The plot below shows word network for all words having higher than a 35% correlation.

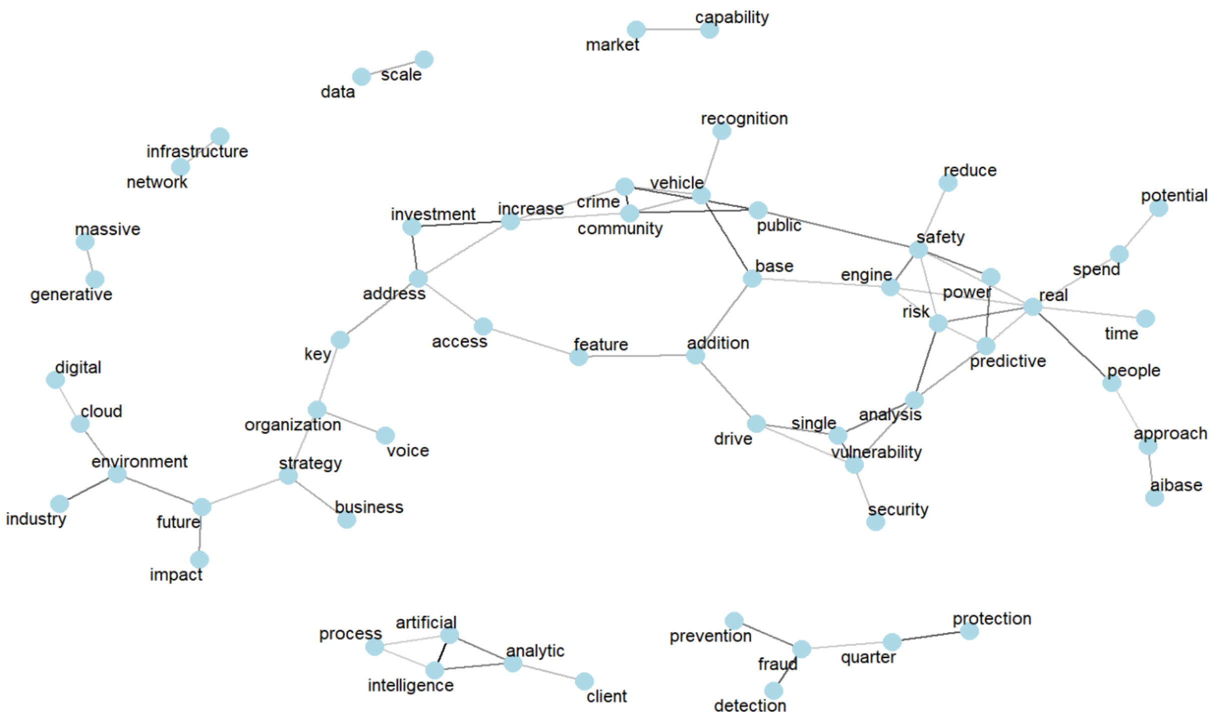


Figure 4 - Word network

From our word network, we can extract valuable insights on how keywords are mentioned in correlation with one another. We can now see that fraud is most likely not mentioned in a negative context, but rather a positive one. On the other hand, security vulnerability, and public safety stands out as possible risks. After this initial look at the co-occurrence of words, we decided that we needed a deeper understanding of the context. To do this we also calculated the co-occurrence of bi-grams in our data. The plot below shows bi-grams mentioned over 4 times with a correlation of 20% or above.

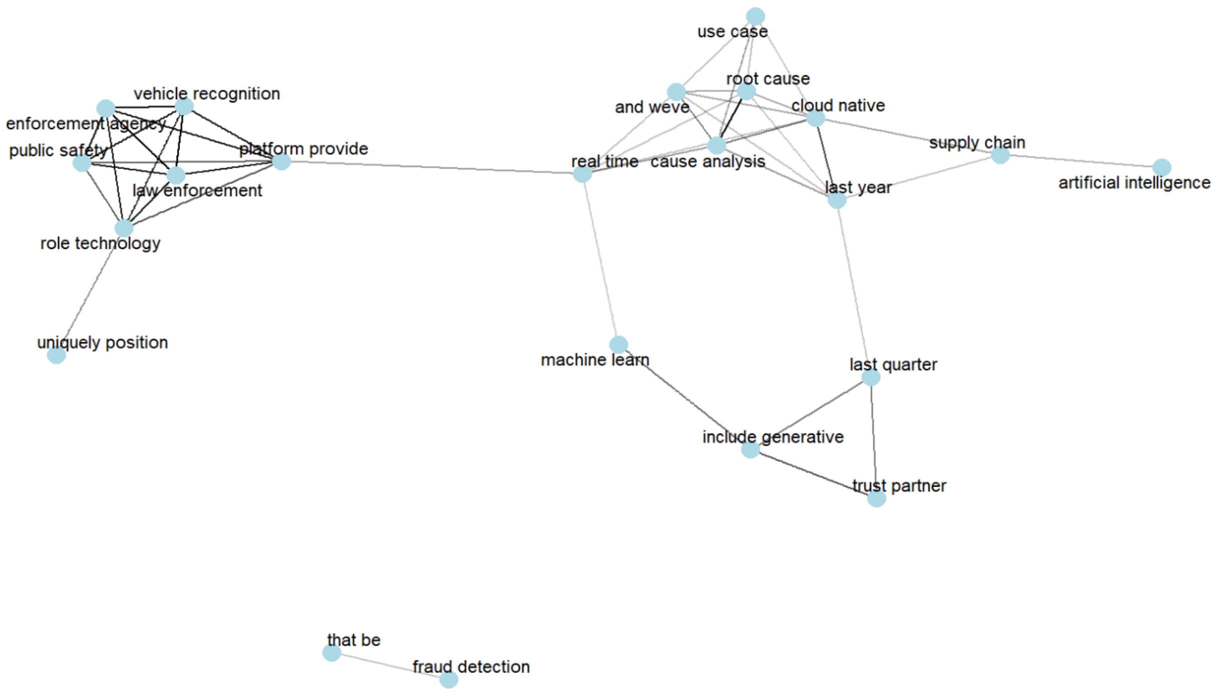


Figure 5 - Bigram network

In the new plot we now are not provided with expanded context of our security vulnerability link from previous analysis. We do however feel that this link is explanatory enough to interpret it as a risk. Firms implementing AI in their digital structure will by the nature of introducing new technology introduce new security vulnerabilities that must be accounted for. We are however provided with deeper context for public safety link in our new plot. We can now see that it relates to the role of AI, and how it will be used by the enforcement agencies. Our interpretation here is that the public is worried about their right to privacy. While this might not be a risk directly connected to firms, it introduces a large reputational risk for companies using AI. If bad PR were to break about a firm crossing a line, the company might lose the trust of the public. There is also a link between law enforcement and role of technology which may represent the risk of regulations from the government. If a firm changes its digital structure to be highly dependent on AI, a new regulation might require companies to use huge amounts of time and money to abide by these regulations. Lastly, we take interest in the link between “uniquely position” and “role of technology”. We interpret this to represent an AI-monopoly approaching in the business-world. This would be a huge liability for other companies, as it would require them to abide by the rules of the monopoly.

Conclusion

By extracting mentions of AI in a negative context, we now have a better overview of which risks firms might face regarding AI. From our initial word co-occurrence analysis, we identified public safety and security vulnerabilities as potential risks. After a deeper dive into the co-occurrence of bigrams we concluded that firms using AI might face risks related to security, reputation, regulation, and the emerging monopolies in the AI-scene.

Limitations of the analysis

Our approach for task 2 has some limitations. Firstly, we have used external lexicons to define our negative and positive sentiments. This might cause positive words to be interpreted negatively and vice versa. Secondly, we only use words that appear somewhat frequently when analyzing co-occurrences. This might cause us to miss important information about risks not mentioned frequently. Another element to consider is that risks may be toned down in earnings calls, because the firms want to present themselves in the best way possible. If they try to use more positive words to front themselves better, the risks may be harder for us to detect through the type of textual analysis we did, and we may have missed some important risks.

Summary

In this report we have investigated which sectors that are affected by the release of ChatGPT and what types of risk firms face regarding AI. We started our analysis with finding related terms to ChatGPT and verify that the chosen terms affect firms. After we had our verified list with keywords, we calculated the frequency of keyword per documents, and aggregated it per sector and quarter. Based on a Wilcoxon Rank Sum test, we could conclude that these five following sectors are affected by the release of ChatGPT: Information-Technology, Communication-Services, Financials, Industrials and Consumer-Discretionary.

Based on these five sectors, we extracted negative segments from the earnings calls in part two of our report. From these negative segments we analyzed co-occurrence of words and bigrams. Our findings are that the risk firms face regarding AI is related to: Public safety, law enforcement, security vulnerabilities and potential monopolies emerging in the AI scene.