Sentiment Analysis on the AI Act

Kristoffer Bakke Tvedt

Semester Assignment for INFO371

Department of Information Science and Media Studies
University of Bergen

June 2, 2023

Abstract

This research paper focuses on exploring data generated from the 2021 European Commission proposal for the AI Act, specifically analyzing the sentiment expressed by businesses and NGOs. The AI Act is a legislative framework proposed by the European Parliament to address the challenges and opportunities of artificial intelligence within the European Union. The paper aims to compare the stances of businesses and NGOs on the AI Act and identify the most frequently used positive and negative terms by these groups. The research questions seek to understand the differences in sentiment between the two actor groups and shed light on their perspectives on the proposed legislation.

To conduct the analysis, the paper employs principles of discovery and measurement. Discovery involves uncovering patterns, trends, and structures in the textual data. Measurement, on the other hand, involves quantifying specific aspects of the text, such as word frequencies and sentiment scores, enabling objective analysis and comparison.

The dataset used in this study comprises PDF submissions from various stakeholders, including AI developers, companies, public administrations, academics, and citizens. The data cleaning process involved addressing issues with actor types, removing non-English documents, and filtering the dataset to focus on documents from businesses and NGOs. The analysis of the cleaned dataset revealed distinct patterns in word usage, reflecting the different priorities and concerns of businesses and NGOs.

Furthermore, the paper applies the Wordfish methodology to measure the ideological positions of actors based on their word usage. Wordfish allows for the quantification and comparison of ideological positions by examining patterns of language usage. This approach provides insights into the underlying perspectives of businesses and NGOs regarding the AI Act. The findings from this research contribute to understanding the sentiments and perspectives of businesses and NGOs towards the proposed AI Act. The analysis reveals the distinct characteristics and objectives of these actor groups, highlighting their different priorities and concerns. The research also provides valuable insights into the most frequently used positive and negative terms by businesses and NGOs, shedding light on their perceptions of the proposed legislation.

Overall, this study contributes to the ongoing discourse surrounding the AI Act and its implications for businesses, NGOs, and the European Union as a whole. By analyzing the sentiment expressed by these actor groups, policymakers and stakeholders can gain a deeper understanding of the potential impact and challenges associated with the proposed legislation.

Contents

List of Figures

Chapter 1

Introduction

In this research paper we will explore data generated from the 2021 European Commission proposal for the AI act. We will try to identify interesting patterns, trends and structures through discovery, and look to quantify specific aspects of the textual data generated from the commission. We will mainly look at a sentiment analysis by comparing the papers generated from businesses to the papers written by NGOs.

The European Parliament has proposed the "AI Act" as a legislative framework to address the challenges posed by artificial intelligence (AI) within the European Union (EU). The Act aims to strike a balance between harnessing the benefits of AI and safeguarding against potential risks. Under the AI Act, high-risk AI systems used in critical sectors like healthcare, law enforcement, and infrastructure will be subject to stringent requirements and conformity assessments. Transparency and accountability are emphasized, requiring developers to provide clear information about AI system operations, data used, algorithms employed, and potential limitations or risks.

Data protection and privacy are key considerations, aligning with the General Data Protection Regulation (GDPR). The Act emphasizes privacy safeguards when processing personal data for AI purposes. The AI Act showcases the EU's commitment to responsible AI development, protection of fundamental rights, and fostering public trust. It aims to ensure that AI is developed and deployed in an ethical and accountable manner.

As the proposal progresses, it will undergo further discussions and potential amendments before becoming law. The AI Act will shape the regulatory landscape for AI in the EU, defining rules that govern AI systems and promoting their responsible use for the benefit of EU citizens (Europa.eu, 2021).

The overarching research question we are trying to solve in this paper is: How does businesses stance on the proposed AI act compared to other groups of actors?

As this is a quite broad research question, and far too big to be covered in this research paper, we decided to do a sentiment analysis on the business actor group and the NGO actor group as these are the groups that generated the most papers for the 2021 European Commission proposal for the AI act.

By conducting a sentiment analysis on these groups we aim to answer questions such as: How do NGOs generally feel about the proposed AI act compared to businesses? What are the most frequently used negative and positive terms used by these groups in regard to the AI act?

These research questions set the foundation for conducting a sentiment analysis, enabling the discovery and measurement of the sentiments present in different actor groups on the topic of the proposed AI act.

Chapter 2

Background

2.1 AI act

As the dataset used in this research paper is based on the 2021 European commission proposal for the AI act, a brief introduction to what the AI act will be helpful. The AI Act is a legislative proposal put forth by the European Commission, the executive body of the European Union (EU). Its objective is to establish rules and regulations for the development, deployment, and use of AI within the EU member states. The proposal aims to address both the opportunities and challenges associated with AI while ensuring the protection of fundamental rights and promoting trust in AI systems.

Key aspects and provisions of the AI Act proposal include:

Scope: The AI Act covers a broad range of AI applications and systems, including both public and private sector use cases, AI algorithms, and related data. Prohibited Practices: The proposal identifies certain AI practices as prohibited, such as AI systems that pose a clear threat to the safety, livelihoods, or rights of individuals. It also explicitly bans AI systems that manipulate individuals' behavior in a way that could cause harm or exploit vulnerabilities.

High-Risk AI Systems: The AI Act introduces specific requirements for high-risk AI systems, such as those used in critical infrastructure, education, healthcare, and law enforcement. High-risk AI systems would need to undergo rigorous testing, documentation, and conformity assessments before deployment.

Transparency and Accountability: The proposal emphasizes the importance of transparency and accountability in AI systems. It requires developers and providers of AI systems to provide clear and accessible information about how the systems work, including data used, the algorithms employed, and the potential limitations or risks involved.

Data and Privacy: The AI Act acknowledges the significance of data protection and privacy. It aligns with the General Data Protection Regulation (GDPR) and emphasizes the need to ensure privacy safeguards when processing personal data for AI purposes.

Supervisory and Enforcement Authorities: The proposal suggests establishing competent national authorities within member states to oversee and enforce the AI Act's provisions. These authorities would have the power to issue fines and penalties for non-compliance (Europa.eu, 2021).

On 11th May the European parliament published a press release on their webpage concerning the status of the AI act. The Internal Market Committee and the Civil Liberties Committee adopted a draft negotiating mandate on the first ever rules for AI with 81 votes in favor, 7 against and 12 abstensions. The next step for the AI act will be to get this draft negotiating mandate endorsed by the whole Parliament so that negotiations with the Council on the final form of the law can start. The vote is expected to be held somewhere between the 12-15th of June (www.europarl.europa.eu, 2023).

2.2 Business vs NGO

To get a better understanding of what types of groups we are comparing, it is beneficial to understand the differences between them. The main difference between a business and an NGO lies in their fundamental purposes, legal structure, funding, stakeholders, accountability and approach to impact.

Purpose: Business: The primary purpose of a business is to generate profits and create value for its owners or shareholders. Businesses typically operate with the intention of maximizing financial returns and sustainability. NGO: The primary purpose of an NGO is to address social, environmental, or humanitarian issues. NGOs are driven by a mission to promote social well-being, advocate for causes, and provide support or services to communities in need.

Legal Structure: Business: A business is typically organized as a for-profit entity, such as a corporation, partnership, or sole proprietorship. It operates within legal and regulatory frameworks that govern commercial activities. NGO: An NGO is typically organized as a non-profit entity, often registered as a charitable organization or a non-governmental entity. NGOs operate under specific legal and regulatory frameworks that govern non-profit organizations.

Funding: Business: Businesses primarily rely on revenue generated through the sale of products or services to sustain their operations and generate profits. They may also raise capital through investments or borrowing. NGO: NGOs rely on various sources of funding, including donations, grants, sponsorships, fundraising events, and government or institutional support. Their focus is not on generating profits but on utilizing resources to fulfill their mission and provide services to their target beneficiaries. Stakeholders:

Business: Businesses have stakeholders such as shareholders, employees, customers, and suppliers. Their primary objective is to create value for shareholders while considering the interests of other stakeholders. NGO: NGOs have stakeholders such as beneficiaries, volunteers, donors, partner organizations, and communities they serve. Their primary focus is on addressing the needs and concerns of their beneficiaries and fulfilling their mission.

Accountability and Governance: Business: Businesses are accountable to their shareholders and are governed by boards of directors and executive management. They operate with a profit-oriented mindset and are subject to financial reporting and transparency requirements. NGO: NGOs are accountable to their beneficiaries, donors, and regulatory bodies. They are governed by boards of directors or trustees, and their operations are guided by their mission and organizational values. NGOs often emphasize transparency and accountability in their work.

Approach to Impact: Business: Businesses aim to create economic value through the production and distribution of goods or services. While some businesses may incorporate corporate social responsibility (CSR) initiatives, their primary focus is on generating profits (HAYES, 2020). NGO: NGOs are mission-driven organizations that focus on addressing social or environmental challenges. Their impact is typically measured by the positive change they bring about in the communities they serve or the progress made towards their stated goals (Folger, 2022).

It is important to note that the line between businesses and NGOs can sometimes blur, as some businesses may incorporate social or environmental goals into their operations (e.g., social enterprises), and some NGOs may engage in commercial activities to generate revenue for their mission. Nonetheless, the core differences outlined above help distinguish the general characteristics and objectives of businesses and NGOs.

Chapter 3

Methodology

We will follow the principles of discovery and measurement when exploring and working with the dataset generated from the commission proposal. These principles are fundamental concepts that guide the analysis of textual data, and the principles help us uncover meaningful patterns, relationships and insights from text.

Discovery involves the process of identifying interesting and previously unknown patterns, trends, or structures within a given textual dataset. It aims to extract valuable information or knowledge that may not be immediately evident. Discovery often involves exploratory data analysis techniques, such as text mining, topic modeling, clustering, or network analysis, to uncover hidden patterns or relationships. The objective of discovery is to gain a deeper understanding of the data and generate new hypotheses or insights that can drive further research or decision-making.

Measurement refers to the process of quantifying specific aspects or attributes of textual data. It involves assigning numerical values or metrics to various dimensions of text, allowing for objective analysis and comparison. Measurement in text analysis can encompass a wide range of aspects, such as word frequencies, document lengths, sentiment scores, topic prevalence, or semantic similarity. By quantifying these aspects, we can derive meaningful statistics, perform statistical analysis, or build predictive models to understand patterns and draw conclusions from the data.

Chapter 4

Methods

4.1 Data

The original dataset comprises views from stakeholders with an interest in artificial interest like AI developers, companies, public administrations, academics, citizens etc. The submissions from these stakeholders are sorted as pdfs in a folder as id_nameOfActor_actorType.pdf, for example F2665634_Amnesty_NGO.pdf. The length of each submission varies, as some stakeholders have a lot to say, while others only use one or two pages.

Before cleaning the dataset there are a total of 265 PDF files in the folder. We start by loading the PDF files with quantedas function readtext and save the data in a variable. The raw data is then processed by removing whitespace and separating it into id, actor and type_actor based on the name of the pdf file. Before saving the data as an rds file, we remove the .pdf ending in the document-level variable. This R script was generously provided by Stefan Müller.

When we create a dataframe from this rds file with readRDS and show a table of type_actor within this dataframe we can immediately see that something is not right with the dataset. The table shows that there are 17 type actors, but many of them only have one document associated with them. Some of them are because of spelling errors in the name of the original pdf, for example should Businesss be in Business, Tradu-Union be in Trade-Union while Consumer-Orgnization and Consumer-organization should be in Consumer-Organization. There is also a single document associated with Academic-Research, Stockholm, Commission and Uni.

Figure 4.1: Dataset actor groups before cleaning

```
              Academic     Academic-Research              Business  Business-Association
                    17                     1                    82                    66
             Businesss               Citizen            Commission Consumer-organization
                     1                     6                     1                     1
Consumer-Organization  Consumer-Orgnization                   NGO                 Other
                     2                     1                    50                    17
      Public-Authority             Stockholm           Trade-Union           tradu-Union
                     4                     1                    12                     1
                   Uni
                     1
```

To clean up the issues with the actor_type docvar values we start by changing the pdf names of the documents with the mentioned spelling errors. Academic-

Research was renamed to only include Academic, while Uni was renamed from F2662611_Vrije_Uni_personal_Other to F2662611_Vrije-Uni-personal_Other so that the actor would be Vrije-Uni-personal and type_actor will be Other. Stockholm had a similar naming issue, so we changed an underscore to a dash so that the actor was now City-Stockholm and type_actor was Public-Authority. The single document associated with Commission was from the proposal by the European Union, so I removed the document entirely together with a similar document that was of no interest for the datasat we are making. The second document removed was associated with Act for actor and NA for type_actor as the naming format was different from the submissions.

After cleaning the document names, running the load and clean R script on the new pdf files and loading it as a dataframe we see that there are now only 9 different type actors. For the cleaned dataset there are now for 18 documents under Academic, 83 under Business, 66 under Business-Association, 6 under Citizen, 4 under Consumer-Organization, 50 under NGO, 18 under Other, 5 under Public-Authority and 13 under Trade-Union for a total of 263 documents.

Figure 4.2: Dataset actor groups after cleaning

```
        Academic            Business  Business-Association             Citizen Consumer-Organization                 NGO               Other
              18                  83                    66                   6                     4                  50                  18
Public-Authority         Trade-Union
               5                  13
```

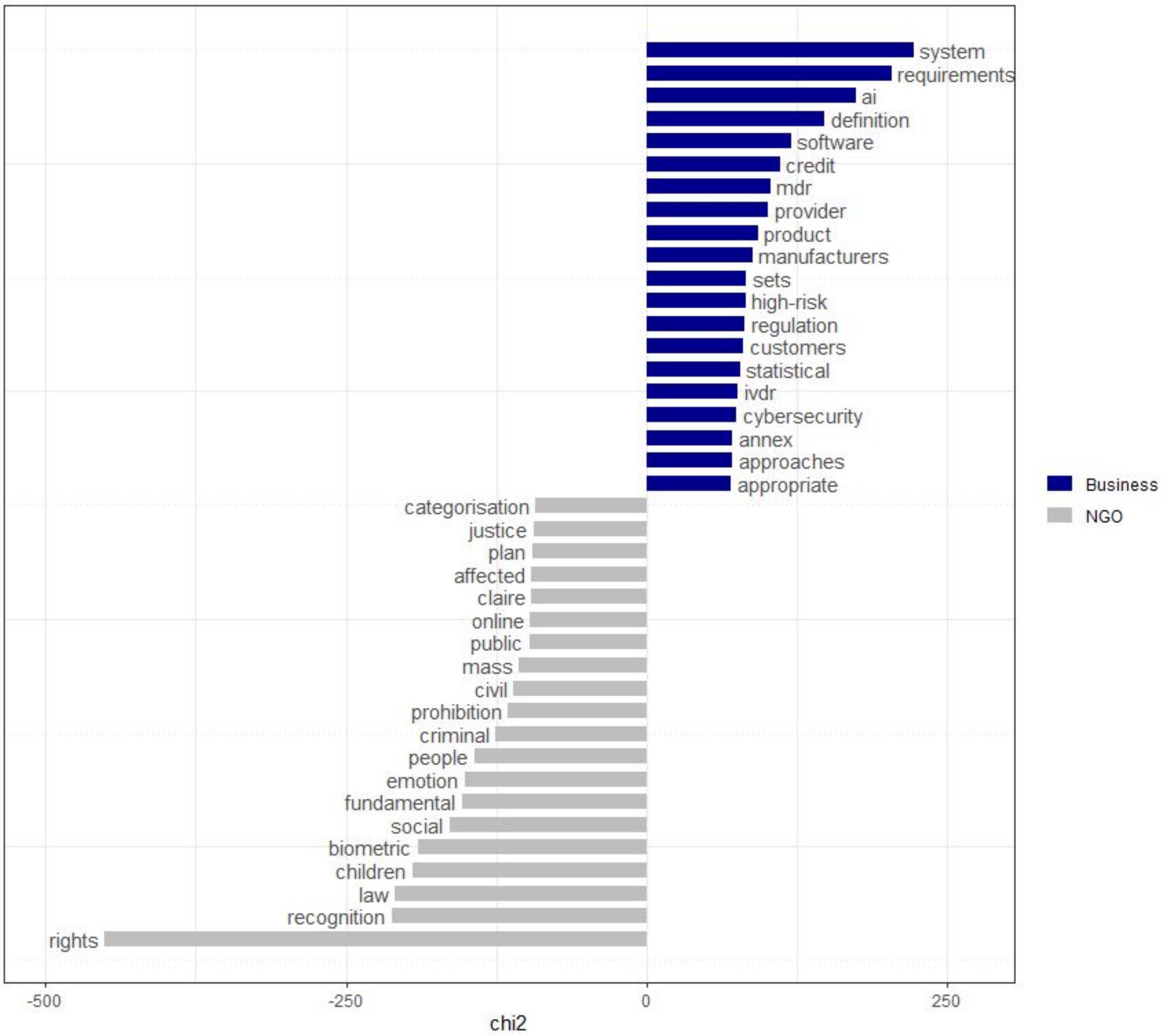
The next preprocessing step for the dataset was to remove any pdf file that was not in English. To do this we manually looked through the documents to see that they were in English and discarded the ones that were in other languages. In the end we removed 12 documents which were in German, 1 Swedish, 1 Dutch, 3 Spanish and 1 French. Four of these documents were from businesses, while none of them were from NGOs.

As we only need the documents associated with Business and NGO for this research paper, the next step was to trim the dataset. This is pretty straightforward with the function filter in R. After filtering the dataset we end up with a dataframe with 129 documents where 79 documents are submitted by businesses and 50 for NGOs.

We noticed that the document names were labeled with text1, text2 etc. which would not be ideal later when plotting graphs, so the next thing we did was to create a variable with the name of every actor, create a corpus from the dataframe we made earlier and change the names of the documents with the function docnames. We also do the same with actor_type and paste that together with the new names to get a good overview with later plots.

To get an idea of the data we are working with, we clean the data by creating a token object from our corpus and remove punctuation characters, numbers and English stop words before creating a document-feature matrix (dfm). We then use quantedas textstat_frequency function to see the 10 most frequent terms per document. To explore the data further we want to conduct a keyness analysis, so we group the dfm by type_actor and use chi-squared as measure. A keyness analysis is a good way to effectively identify frequently used words for a target and reference group. From the analysis we can see that frequent words for business documents include system, requirements, definition, software and credit, while NGOs frequently use words like right, recognition, law, children and social. This is not surprising as the primary purpose of an NGO is to address social, environmental, or humanitarian issues, while for a business its to generate revenue.

Figure 4.3: Keyness analysis



4.2 Implementation

After getting a good look at the dataset we do some measurements. The first model we use on the new dataset is wordfish. Wordfish is a statistical method used to analyze textual data and uncover underlying dimensions or latent variables. It is a technique commonly applied in the field of political science, particularly in the analysis of political speeches, legislative texts, or other forms of written communication. The main goal of Wordfish is to measure the ideological positions of actors based on their usage of words or terms within a given set of documents. It allows us to quantify and compare the ideological positions of different actors or groups by examining their patterns of word usage.

The Wordfish methodology operates on the assumption that word choice reflects underlying ideological or semantic preferences. It assumes that speakers or authors who share similar ideological positions are more likely to use similar words or language to convey their ideas.

To apply Wordfish, the first step is to construct a corpus of texts or documents that represent the actors or groups of interest. This corpus typically consists of speeches, policy statements, or other relevant texts, but in our case we use the documents submitted by stakeholders. The corpus is then preprocessed by removing stopwords (common words like "the," "and," "is," etc.) and transforming the text into a numerical representation suitable for analysis.

Next, the Wordfish algorithm calculates word scores by examining the co-occurrence patterns of words across the corpus. It evaluates the statistical association between words and constructs a dimension or scale that represents the ideological position underlying the observed word usage. The resulting word scores or ideological dimensions can be used to place individual actors or groups along a continuum, representing their positions on the identified dimensions. This allows for comparisons and analysis of the ideological distances between different actors or groups based on their word usage.

Wordfish offers a data-driven approach to uncovering latent ideological dimensions without relying on explicit ideological labels or subjective evaluations. By examining the linguistic choices of actors, it provides insights into their ideological positions and facilitates the quantitative analysis of political discourse.It is however worth noting that while Wordfish is a valuable tool for analyzing ideological positions, it is not without limitations. The methodology assumes that word usage directly reflects ideology, which may not always hold true. It is also sensitive to the specific documents and words included in the analysis, requiring careful selection and preprocessing of the corpus (Anon, 2017).

Finally we conduct a sentiment analysis using quantedas function textstatpolarity and lexicoder sentiment dictionary. With this function we compute sentiment scores using a polarity approach based on assigned categories of positive, negative and neutral sentiment. The Lexicoder Sentiment Dictionary is a widely used resource in sentiment analysis. It consists of a collection of words and their associated sentiment scores or labels. Each word in the dictionary is assigned a positive or negative sentiment score, indicating whether it conveys a positive or negative sentiment.

To perform sentiment analysis using the Lexicoder Sentiment Dictionary, the text data is first preprocessed by removing stopwords, punctuation and numbers. Then, each word in the text is matched against the words in the sentiment dictionary. The Lexicoder Sentiment Dictionary is a widely used resource in sentiment analysis. The dictionary consists of 2858 negative sentiment words and 1709 positive sentiment words.

To perform the analysis, we first create a new object with our token object and apply the Lexicoder Sentiment Dictionary. We then get the sentiment score by using the textstatpolarity function on our new token object and the Lexicoder Sentiment Dictionary.

The sentiment scores or labels associated with the matched words are aggregated to calculate an overall sentiment score for the text. For example, positive sentiment scores may be assigned a value of +1, negative scores a value of -1, and neutral scores a value of 0. We then create a graph to compare different actors, as well as measuring the mean and median of the business and NGO classes. Based on the overall sentiment score, the text can be classified as positive, negative, or neutral. A positive score indicates a predominantly positive sentiment, a negative score indicates a predominantly negative sentiment, and a score close to zero suggests a neutral sentiment (rdrr.io, n.d.).

Chapter 5

Results and Discussion

5.1 Results and Analyses

From the wordfish analysis we can see that the general pattern is that NGOs converge at the top of the graph, while businesses generally stay at the bottom. 50 percentage of NGOs (25 documents) are within the top 31 percentage of all documents (129), while 50.6 percentage of business documents are within the bottom 35 percentage of all documents.

Worth noting with the wordfish analysis is that the majority of the papers have a relatively close estimated theta value between -1 and 1. There are 9 documents that can be considered outliers based on their estimated theta, where 6 of these are from NGOs and 3 from business.

These results could imply that the ideological positions of NGOs and businesses are to some degree different, but not radically so. Its also interesting to see that the majority of NGOs have the same ideological position, which is also the case with the business actors.

Looking at the plot generated with the sentiment scores the first thing we noticed was that there were very few negative scores. There are a total of 4 documents with a sentiment score below 0, and 3 other documents which are very neutral just above 0. Its also worth noting that all 7 of these are by NGOs, and that the 9 lowest sentiment scores are all from NGOs.

Another interesting observation is that most of the documents have a score between 0.25 and 1.25, while only 6 documents have a sentiment score above 1.5. The sentiment scores for NGOs are scattered across the graph, while the business documents for the most part stay around a sentiment score of 1.

If we look at the mean and median of the two type actors we can see that business scores are higher in both with a mean score of 0.89 and a median of 0.87 while NGOs have a mean score of 0.71 and a median score of 0.75.

Finally we looked at what the most frequently used positive and negative words were. For positive, the words rights were used the most, followed by intelligence, ensure, safety and protection. The most frequently used negative words were risk, artificial, risks, oversight and harm.

Figure 5.1: Top 50 most frequently used positive words

rights	intelligence	ensure	safety	protection	innovation	provide
1241	923	640	522	485	442	363
well	relevant	support	right	compliance	quality	learning
358	332	322	297	278	265	261
significant	effective	create	like	ethical	help	principles
236	222	208	200	194	194	190
credit	ensuring	best	trust	better	benefits	accuracy
188	177	176	166	162	160	159
responsible	opportunity	respect	competent	trustworthy	free	benefit
155	155	154	154	149	146	143
validation	coordinated	protect	understand	civil	essential	welcomes
143	139	136	135	134	133	133
welcome	accessible	achieve	providing	improve	knowledge	creating
132	132	131	124	123	123	120
open						
120						

Figure 5.2: Top 50 most frequently used negative words

risk	artificial	risks	oversight	harm
1195	903	638	259	241
bias	prohibited	limited	against	prohibition
226	221	215	194	174
too	avoid	criminal	discrimination	errors
171	167	160	156	141
affected	risk-based	challenges	lack	critical
133	126	122	118	108
harms	uncertainty	difficult	concern	concerned
108	105	102	100	93
unclear	burden	implications	prohibitions	unacceptable
91	91	86	84	83
limitations	harmful	biases	discriminatory	negative
81	76	65	64	64
insufficient	problematic	limit	disproportionate	liability
63	63	60	59	55
unnecessary	limits	challenge	adverse	impossible
55	54	53	52	51
threat	complexity	prohibit	disagree	vague
50	50	49	49	48

5.2 Discussion and future work

The results from the sentiment analysis were somewhat surprising, as before conducting the analysis we believed that NGOs would generally score higher than business actors, as they would have more incentive to be positive towards a legislative framework to address the challenges posed by artificial intelligence. The intention of the AI act is to identify certain AI practices as prohibited, such as AI systems that pose a clear threat to the safety, livelihoods, or rights of individuals. It also explicitly bans AI systems that manipulate individuals' behavior in a way that could cause harm or exploit vulnerabilities, which is in NGOs interest. This could however be because these submissions were generated quite early in the AI act process, and there was a general displeasurement with the definition of AI, with some claiming it to be too broad or unclear and others complaining that the definition were missing aspects of AI.

Another interesting aspect of the analysis was how close together the actors from NGOs and business were on the wordfish analysis. Before conducting these experiments we believed that NGOs and businesses would be very far away from each other ideologically.

Its interesting to see the differences and similarities between the two actor groups, and it would be even more interesting to see how business compares to the other 7 actor groups. For future work with the large overarching research question, we would conduct the same analysis as we have done with NGOs with the other actor groups to compare them with businesses. If we were to continue with this work we would also look more into the most frequently used negative and positive terms, as we believe that there could be some value to removing words like artificial in the negative terms and intelligence in the positive terms. While these are correct on their own, they lose their meaning in the context of artificial intelligence.

Figure 5.3: Wordfish analysis

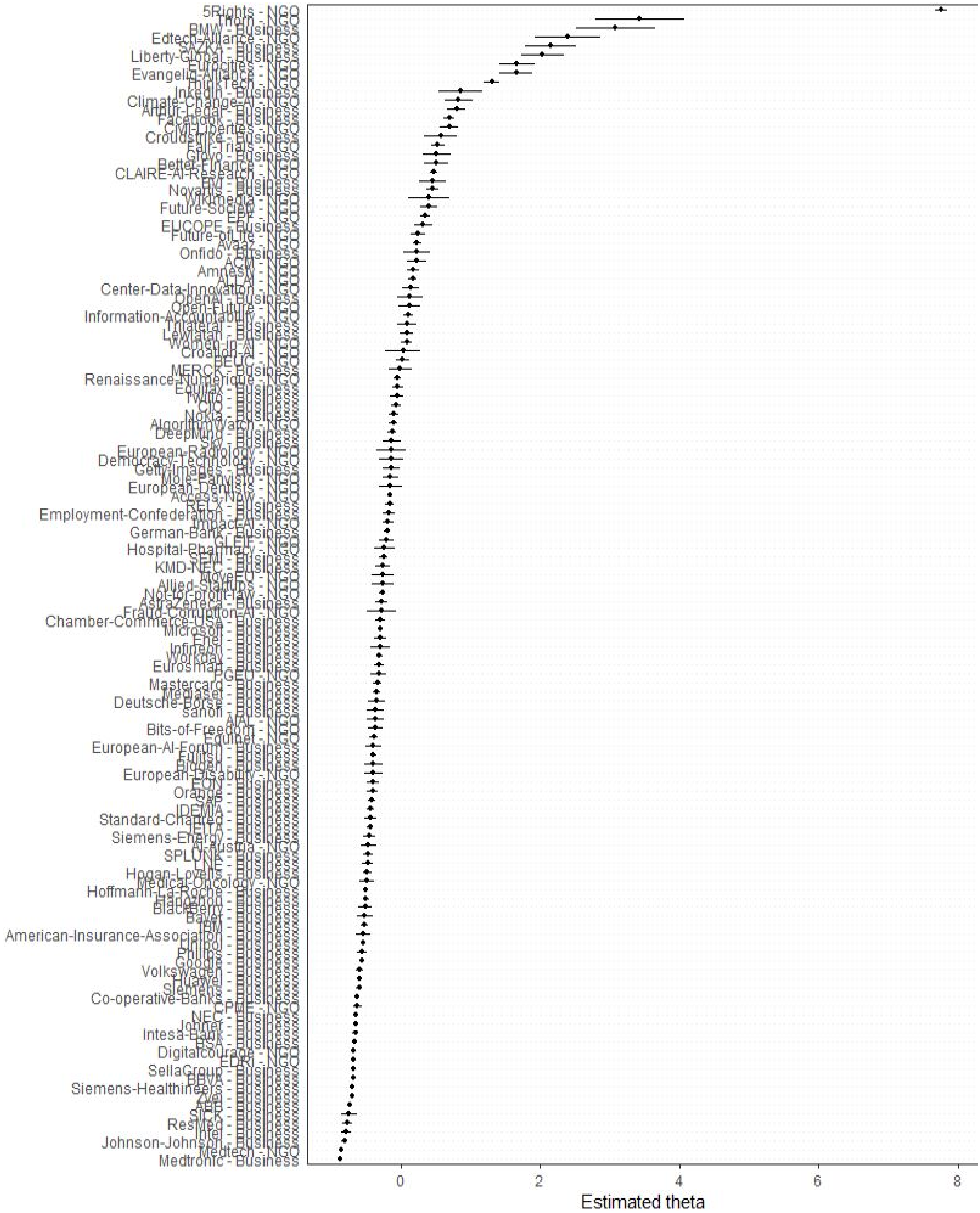
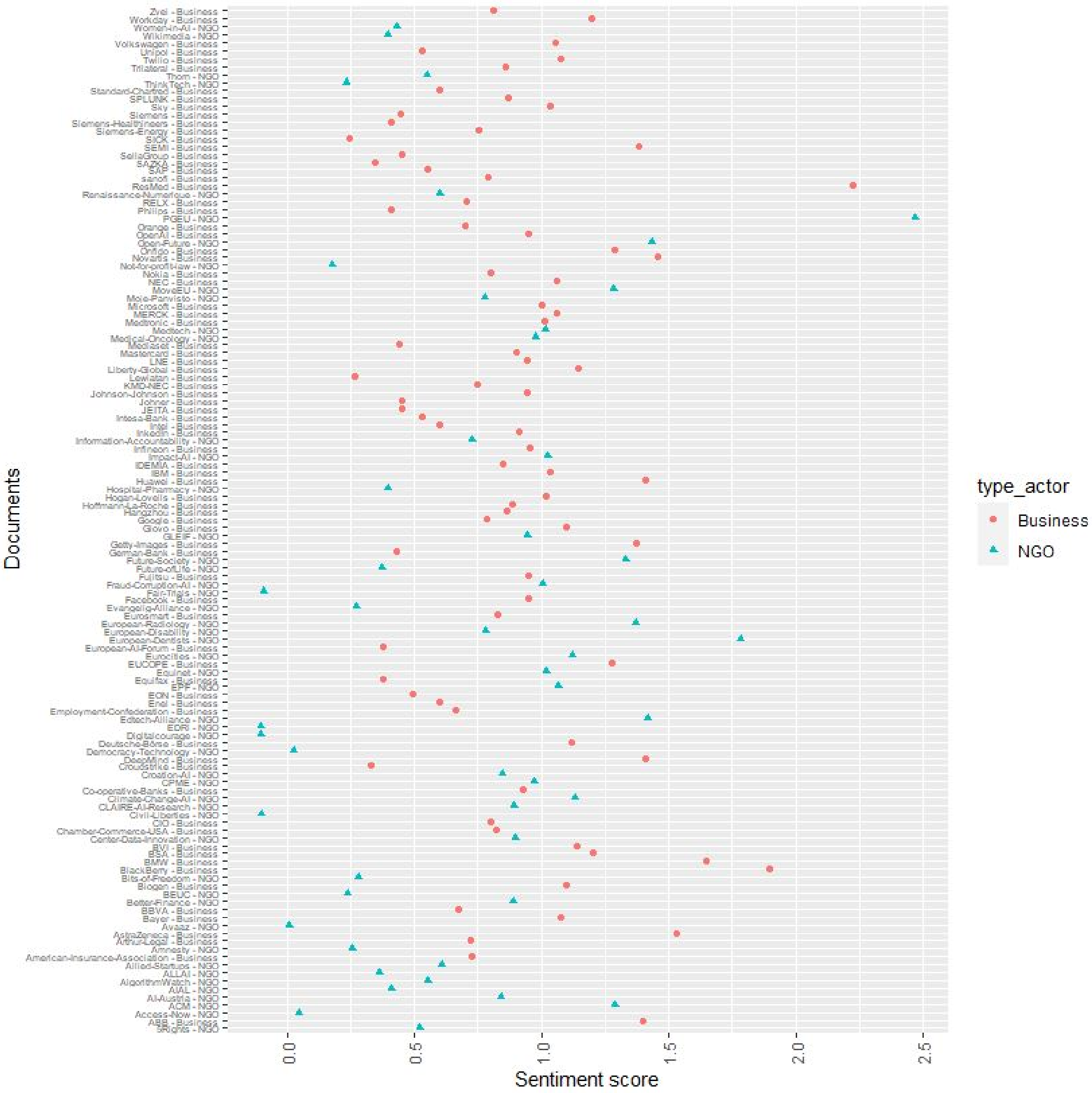


Figure 5.4: Sentiment analysis



Chapter 6

Conclusion

We have answered the two smaller research questions we set out to solve which were: How do NGOs generally feel about the proposed AI act compared to businesses? and What are the most frequently used negative and positive terms used by these groups in regard to the AI act? With a mean sentiment score of 0.71 and a median score of 0.75 we can conclude that the general sentiment from these documents are slightly positive, but they are still below business on both mean and median with business documents scoring 0.89 and 0.87. We did also discover that the most frequent positive words were rights, intelligence, ensure, safety and protection while the most frequent negative words were risk, artificial, risks, oversight and harm.

A part of the bigger overarching research question has been answered, but there is still much more to be done.

Chapter 7

Bibliography

www.europarl.europa.eu. (2023). AI Act: a step closer to the first rules on Artificial Intelligence | News | European Parliament. [online] Available at: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence.

HAYES, A. (2020). What Is a Business? Understanding Different Types and Company Sizes. [online] Investopedia. Available at: https://www.investopedia.com/terms/b/business.asp.

Folger, J. (2022). What is an NGO (Non-Governmental Organization)? [online] Investopedia. Available at: https://www.investopedia.com/ask/answers/13/what-is-non-government-organization.asp.

Europa.eu. (2021). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CE

Anon, (2017). Use WORDFISH for Ideological Scaling: Unsupervised Learning of Textual Data Part I Loretta C. Duckworth Scholars Studio. [online] Available at: https://sites.temple.edu/tudsc/2017/11/09/use-wordfish-for-ideological-scaling/ [Accessed 2 Jun. 2023].

rdrr.io. (n.d.). quanteda/quanteda.sentiment source: R/textstat_polarity.R. [online] Available at: https://rdrr.io/github/quanteda/quanteda.sentiment/src/R/textstat_polarity.R [Accessed 2 Jun. 2023].