Supervisor: Dr. Mark Boukes
Student: Helge Moes, 11348801
Faculty: Social and Behavioural Sciences
Program: Research Masters Communication Science
Course: Research Master's Thesis Communication Science
Date: 28-06-2024

UNIVERSITEIT VAN AMSTERDAM

Navigating the News

Analyzing AI Risks according to Dutch Journalists Preceding the European AI Act of 2024

Word Count:
7499

Graduate School of Communication

0. Abstract

Artificial intelligence (AI) is conquering the world, offering humanity many benefits while simultaneously posing numerous threats for society as well. In response to these challenges, the European Parliament passed the first comprehensive regulation concerning AI, the Artificial Intelligence Act (AI Act), on March 13, 2024. (*The Act Texts | EU Artificial Intelligence Act*, 2024). The AI Act aims to contain the dangers for society and establishes a regulatory framework of AI risks for specific social domains in 4 levels: unacceptable, high, limited and minimal risk.

Currently, little academic research has been executed in the specific role of journalists and news outlets in their coverage on AI in relation with the EU's AI Act. To compare the societal AI risks discussed by journalists from 2014 to 2024 with those structured in the EU's in the AI Act of 2024, newspaper articles on AI were collected from six national Dutch newspapers ranging from tabloids to broadsheets (*de Telegraaf*, *Algemeen Dagblad*, *Trouw*, *de Volkskrant*, *NRC* and *het Financieele Dagblad*). The resulting dataset (*N* = 6,446) was examined by employing an exploratory mixed method approach through automated topic modeling and manually performed content analysis.

Findings indicate a strong overall alignment between the AI risks highlighted by journalists and the regulatory framework of the AI Act and certain topics. In addition, key trends include increased media coverage over time, a shift in focus from solely technology to societal relevant topics such as business, economy, and politics. Notable variations influenced by the type of news outlet were also detected, since tabloid newspapers tend to exaggerate AI risks compared to broadsheets, which align more closely with the AI Act's risk assessments.

However, topic-specific analysis reveals that while general alignment exists, discrepancies arise in areas such as Art, Law, and Politics, where journalists often perceive higher risks than those acknowledged by the AI Act. Conversely, topics such as *Environment* resemble the AI Act's assessment.

In conclusion, this thesis underscores the critical role of journalism in the societal discourse on AI, AI literacy and highlights the importance of regulatory frameworks in addressing the multifaceted risks associated with AI advancements.

Table of Content

1. Introduction

In the last decade, artificial intelligence (AI) has emerged as a critical area of focus in various sectors including industry, governance and academic research (Brennen et al., 2022; Vrabič Dežman, 2024). From facial recognition, medical diagnosis, fraud prevention, controlling industrial processes and selecting our daily news, AI is impacting our lives and society (Hagerty & Rubinov, 2019). So has AI been a recurring topic in popular culture, with sci-fi narratives often exploring their power and risks (Cave & Dihal, 2019; Fast and Horvitz, 2017). Besides the significant impact it has, AI has also become a widespread topic of discussion. Recent technological advancements, especially in machine learning algorithms, have blurred the lines between scientific reality and popular perception, making AI prominent in media discourse (Chuan et al., 2019). However, alongside the enthusiasm, these technologies have raised significant concerns (Araujo et al., 2020).

However, risks of AI have been noticed as well in news outlets, who play a fundamental role in shaping individual opinions on emerging technologies (Nguyen & Hekman, 2022). They cover stories on both the progress offered by AI and the potential harms it has, such as privacy invasions and algorithmic discrimination, influencing the public view on AI risks portrayed in the news and the functioning of democracy. Given the influential role, academia also suggests an urgency to address risks regarding emerging technologies (Helberger & Diakopoulos, 2023; Nguyen, 2023; Kaminski, 2022; Steimers & Schneider, 2022).

This urgency has led to the recognition of risks by the European Parliament (*The Act Texts | EU Artificial Intelligence Act*, 2024). As AI technologies become increasingly integrated into products, services and our daily lives, concerns have been raised about their implications for fundamental human rights (Floridi et al., 2018; Latonero, 2018; Rodrigues, 2020). As a consequence, on the 13th of March 2024 the European Parliament passed the Artificial Intelligence Act, which establishes a common regulatory and legal framework for AI within the European Union (*The Act Texts | EU Artificial Intelligence Act*, 2024).

In order to analyze the alignment between the risk levels described by journalists and the AI Act, and to explore the nuanced portrayal of AI in the European media, this study aims to examine

Dutch news coverage of six prominent national news outlets ranging from tabloids to broadsheets: *de Telegraaf*, *Algemeen Dagblad*, *Trouw*, *de Volkskrant*, *NRC* and *het Financieele Dagblad*. Given the complexity of the concept of AI (Schuett, 2019), factors such as time, news outlets and topics journalists write about may have a significant impact on how AI risks were perceived in the 10 years preceding the AI Act (Sun et al, 2020; Vergeer, 2020). Therefore, this examination aims to answer the following research question:

Research Question:

"To what extent does the regulatory framework established by the AI Act align with the portrayal of AI risks in journalism, and what factors (time, news outlet and topic) influence the coverage of risk?"

To answer the research question, this study operates a mixed method approach of both quantitative and qualitative research techniques (Creswell & Creswell, 2017). It combines topic modeling, with a summative content analysis for the risk analysis of AI (Hsieh & Shannon, 2005). The research shall implement the risk structure of the AI Act as a foundation for a codebook to explore the extent the risk portrayed by journalists aligns with the regulatory framework of the AI Act.

2. Theoretical framework

This section explores the theoretical underpinnings of AI, risk perception, and the EU's AI Act, including historical definitions, the evolution of AI concepts, the role of news media in shaping societal perceptions of AI-related risks, and the regulatory framework provided by the AI Act.

2.1. Artificial Intelligence (AI)

Research on artificial intelligence (AI) began with Alan Turing's 1950 Turing Test, which evaluated a machine's ability to mimic human responses, establishing a groundwork for AI as human-like behavior (Turing, 1950). This concept expanded at the 1956 Dartmouth Conference, where John McCarthy coined ‘Artificial Intelligence’ and described it as creating intelligent machines and programs (Nilsson, 2009; McCarthy, 2007). Over the years, the definition of AI has evolved due to its complexity and ongoing development. Russell et al. (2010) described AI as the study of intelligent agents that perceive their environment and act to optimize performance, reflecting its broad applications and interdisciplinary nature, including machine learning and algorithms. Other contemporary views recognize AI systems as varying in autonomy and adaptiveness, capable of influencing anything based on input (Radley-Gardner & Zimmerman, 2016).

Therefore, AI can be understood as an umbrella term for a range of technologies and systems, such as automated data analysis, machine learning and natural language processing (Deuze & Beckett, 2022). These systems that simulate human behavior and automate processes are implemented in various fields including business, security, and creativity, potentially transforming job markets (Darling, 2015; Williamson, 2017). Due to its versatility, AI remains a challenging concept to understand and determine, leading to misconceptions and anxiety due to its abstract nature and varying definitions across different domains (Wang et al., 2023; Williamson, 2017). For example, policymakers define AI as machine-based systems designed to operate autonomously, capable of generating predictions and decisions that influence environments, highlighting its broad societal impact (Radley-Gardner et al., 2016). While academia, such as Schuett (2019), note the lack of a universally accepted definition, reflecting AI's complexity.

Moreover, the complexity is accompanied by concerns about the societal implications of AI and the recent surge in sensational attention to it, also referred to as an ‘AI hype’ (Vrabič Dežman, 2024). This resembles earlier ‘media hypes’ (Vasterman, 2014) such as the Internet (Rössler, 2001) and Nanotechnology (Cacciatore et al., 2012). It is characterized by creating news by reporting on the excitement (Vasterman, 2014), leading to a cycle of emotional reactions and numerous studies focusing on AI news coverage.

Bunz and Braghieri (2021) analyzed AI in medical contexts over four decades, finding AI often seen as superior to human experts, potentially obscuring human accountability and reflecting creators' biases (Nguyen & Hekman, 2022). Moreover, Vergeer (2020) studied AI reporting in Dutch newspapers from 2000 to 2018, noting increased coverage from 2014 and significant differences among newspapers. Ouchchy et al. (2020) examined media portrayal of AI's ethical issues, finding the coverage realistic yet shallow. Additionally, Cools et al. (2024) identified themes in articles on AI in US newspapers from 1985 to 2020 used LDA topic modeling to identify prominent themes, revealing an optimistic tone but significant attention to ethical dilemmas. Finally, Sun et al. (2020) analyzed articles from major newspapers and identified fourteen major AI-related topics and noted sophisticated framing by journalists, with diverse actors shaping the media discourse.

There have been various procedures used to explore and examine the topics covered in the news each with their own strengths and weaknesses. For instance, studies have utilized frames and topics (Chuan et al., 2019; Cools et al., 2024 Vergeer 2020), while others have opted for a content analysis (Brennen et al., 2022; Bunz & Braghieri, 2021; Ouchchy et al., 2020) or both (Sun et al., 2020; Hase et al., 2020; Nguyen, 2023). Traditionally, identifying topics in media texts has relied on techniques, such as content analysis, an in-depth method employed by communication and social scientists (Krippendorff, 2019; Titscher et al., 2000). However, Recchia (2020) argues for the application of computational methods to study AI narratives, emphasizing their ability to analyze large texts and uncover patterns and nuances often missed without implementing digital tools. Therefore, analyzing these topics through a mixed method approach grants an elaborate understanding of the journalistic discourses triggered by AI, since these discourses give feedback into the development, use, and regulation of the technologies.

Consequently, this study shall combine automated and manual analysis in an attempt to leverage the topics for content analysis purposes, to gain specific context related to the portrayal of artificial intelligence. Besides granting a birds-eye-view of how artificial intelligence is portrayed in the Dutch media, it explores how these topics influence the perspectives of journalists over time. To capture a nuanced understanding of the topics regarding AI, this study addresses the following subquestion:

RQ1:

What topics are predominantly portrayed in the news by journalists with regard to artificial intelligence?

2.2. Risk

Journalists act as critical observers and gatekeepers of information (Shoemaker et al., 2008). Whereas the news provides a platform for discourse on issues shaping our perception of societal contexts (Van Dijk, 1995). Considering AI's versatility and plurality, news reporting is likely to cover its impact across diverse sections such as business, politics and technology (Nguyen & Hekman, 2022). Therefore, news media play a key role in shaping the societal meaning of AI technology, proactively constructing risks (Lupton, 2013). For this reason, Binder et al. (2014) view risk in the news as a combination of probability and subjective interpretation of consequences, known as risk perception. The media significantly influence risk perceptions through factors such as coverage amount, topics, presentation format, and media types (McCarthy et al., 2008).

According to Douglas (1994), risk is the 'danger from future damage', emphasizing negative outcomes. Besides that these negative outcomes are prone to be highlighted in the media, they distort reality by emphasizing violence and negative events, which leads to biases (Hase et al., 2020). News outlets are accused of misrepresenting risk and prioritizing sensational stories, which can lead to irrational public behavior and misplaced fears (Kitzinger, 1999). This tends to occur often in tabloids in comparison to broadsheets (Johansson, 2020).

From social scoring to the use of deep fakes, AI has been portrayed by the media with mixed receptions by journalists and experts (Nguyen & Hekman, 2022; Araujo et al., 2020). Additionally, Radley-Gardner et al. (2016) define risks of AI systems as "the probability of harm and its severity," highlighting the importance of both likelihood and impact, especially given the wide use and integration of AI systems trained on large datasets, which makes them versatile and effective in many tasks. Roslyng & Eskjær (2017) argue that the media often portrays risk technologies, such as AI, as unmanageable rather than controllable.

For example, it has also been stated that artificial intelligence increases risks to both society and individuals' rights and freedoms (Floridi et al., 2018; Latonero, 2018; Rodrigues, 2020). This may stem from a lack of AI literacy, which underlines the understanding, recognition and appreciation of its normative dimensions as much as its impact on society (Cave et al., 2019; Deuze & Beckett, 2022).

In addition, Crépel et al. (2021) show that news media engage critically with AI-related issues, emphasizing the importance of AI literacy and algorithmic awareness in shaping attitudes towards AI (Nguyen, 2017; Gray et al., 2018).

It comes as no surprise that this resulted in tensions between experts and journalists on AI's beneficial and harmful impacts, with recurring subjects of privacy invasion versus data accuracy, and automation benefits versus loss of human influence (Cave & Dihal, 2019). For example, AI enhances efficiency and creativity in various fields, but also introduces risks such as data bias, algorithmic discrimination, surveillance, disinformation, and cybercrime (Nguyen, 2022). These tensions can lead to a clash with policy makers who are responsible for the regulation of AI. Therefore, Kitzinger (2009) advocates that instead of viewing news coverage as 'irrational' or dismissing it as 'sensationalist', it is crucial to understand the factors shaping such coverage, since it involves controllability and preventability.

The traditional approach to studying media reporting of risk has focused on assessing the accuracy of media coverage against expert assessments (Kitzinger, 1999). In the past, researchers have criticized the media for failing to provide sufficient information to make informed risk judgments and for focusing on minor or rare hazards. Such studies have found that media attention to risks is often unrelated to their actual risk or frequency of occurrence, which can be deemed misleading (Singer & Endreny, 1987). Consequently, this study shall examine whether time, topic and news outlet influence the coverage of AI preceding decade before the establishment of the AI Act by addressing the following question:

RQ2:

What factors (time, topics and news outlets) affect how journalists portray risks in their reporting on artificial intelligence?

2.3. Artificial Intelligence Act (AIA)

Initially, the European Commission adopted a non-binding ethical guideline approach in 2019, but in 2021, it shifted towards a legislative approach with the Communication on Fostering a

European approach to Artificial Intelligence (European Commission, 2021). The proposed new regulatory framework aimed to harmonize rules for the development, placement, and use of AI systems, and would complement existing rules set by the European Union. To ensure that AI technologies are aligned with values and principles of the European Union, policy-makers have developed a 'humancentric' approach to AI regulation (European Commission, 2021). This can be described as the 'AI act' or the 'AIA' (*The Act Texts | EU Artificial Intelligence Act*, 2024).

The AI Act is the first regulation produced by the European Parliament on the use of artificial intelligence to stimulate the development of AI in the direction where humans are in control of the innovative technology. Proposed by the European Commission on the 21st of April 2021 and passed on the 13th of March 2024 (European Commission, 2021), the Act encompasses all different types of AI that is implemented in a broad range of sectors into 4 risk levels: minimal, limited, high and unacceptable (*The Act Texts | EU Artificial Intelligence Act*, 2024).

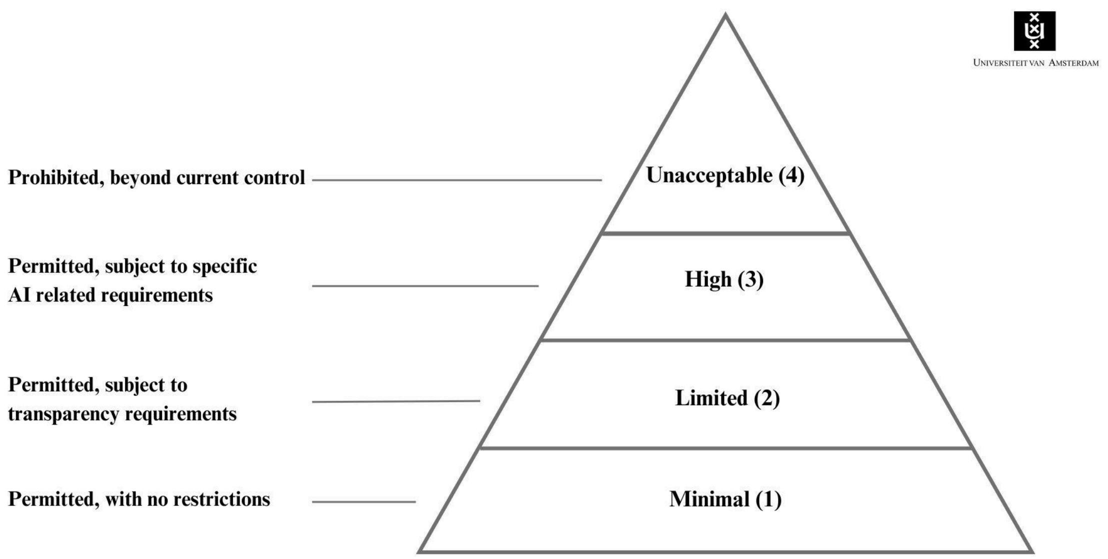

Figure 1: Risk Levels AI Act (2024)

Since the AIA is a regulation and not a law, it does not impose rights on individuals; yet, it regulates the providers of artificial intelligence systems and who uses the technology in a professional context. The regulation has been a topic of scrutiny, since it does not impose any laws. This criticism emerged from the lack of third-party assessments for many high-risk AI systems, raising concerns about safety and the classification of certain technologies, such as deep fakes, for stricter regulation. In this case, it was too late to be prevented (Helberger & Diakopoulos, 2023). Although the AI community has recently endeavored to establish standardized procedures for documenting models, methods, systems, and datasets, there is no methodology specifically tailored that aligns with the risk-based approach outlined in the European AI Act (Hupont Torres et al., 2023).

This research shall implement the risk structure of the AI Act as a codebook to explore the extent the risk portrayed by journalists and the regulatory framework by the European Parliament of the AI Act align. The findings of this study contribute to AI literacy and an improved understanding of how journalists perceive and report on emerging technologies such as artificial intelligence (Deuze & Beckett, 2022; Cave & Dihal, 2019; Araujo et al., 2020). Therefore, following question is formulated:

RQ3:

To what extent does the risk portrayed by journalists and the regulatory framework of the AI Act overlap or differ from one another?

3. Methodology

This section presents the method on how the news data is collected and analyzed incorporating the risk structure set by the European Commission as portrayed in Figure 2.

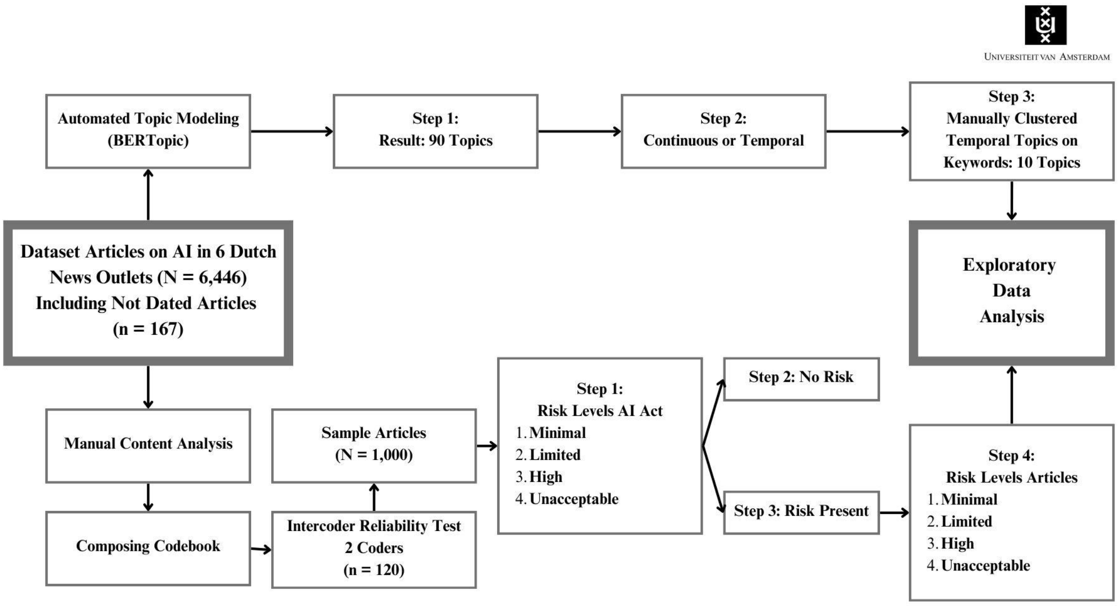

Figure 2: Overview Mixed Methodology Approach

3.1. Data Collection

Web scraping is not inherently illegal, yet scraping copyrighted content as images, videos, or articles without permission is considered to be illegal, emphasizing the importance of respecting copyright restrictions when scraping news articles (Hobby, 2024). Therefore, this study resorted to collecting secondary observational data from the academic data repository of Nexis Uni, which is an online archive for news sources. It grants access to local, national and international newspapers and magazines (Nexis Uni® Home, n.d.). By using the database, this study adheres to the guidelines set by good research practices and ethical standards (Bowman et al., 2022).

This resulted in a list of articles that consisted of links of newspapers and headlines portrayed in Nexis Uni, excluding copyrighted materials such as the original news articles. However, Nexis Uni uses tokenization for accessing the transcripts of the articles. To avoid constraints, geckodriver, a proxy that allows Selenium to interact with the Firefox browser to perform automated tasks, was utilized in order to scrape each article that has been retrieved from Nexis Uni in python (*Selenium*, 2022).

3.2. Dataset

The AI Act is implemented to regulate the use of AI in Europe (The Act Texts | EU Artificial Intelligence Act, 2024). Therefore, a sample of Dutch newspapers was selected based on three criteria: a) national news outlets, b) publication period (time frame), and c) specific keywords used in the selected articles. National outlets are chosen, because they are likely to cover significant AI advancements, regulatory changes, and international AI trends that can influence public opinion and policy (Vergeer, 2020). Consequently, the following newspapers were selected: het *Financieele Dagblad, NRC, de Volkskrant,* and *Trouw* (considered to be broadsheet newspapers) and *Algemeen Dagblad* and *De Telegraaf* (characterized as tabloid newspapers) (Boukes & Vliegenthart, 2017; Alba-Juez, 2017).

Furthermore, the development of AI is well represented and has become a relevant topic since the 10 years preceding the AI Act on the 13^{th} of March 2024 (Dwivedi et al., 2023; The Act Texts | EU Artificial Intelligence Act, 2024). Therefore, the time frame of this sample consisted of longitudinal data from the period March 31 2014 to March 31 2024. Besides the timeframe to determine the sample, the queries 'kunstmatige intelligentie' and 'AI', which - as an abbreviation includes the term 'artificial intelligence' - were used to collect sample articles, similarly executed by Vergeer (2020). This resulted in a dataset of 6,446 articles (*Appendix A*).

3.3. Data Analysis

In this section, the exploratory data analysis (EDA) plan is explained, which consists of topic modeling and manual content analysis by using the AI Act (Páez & Boisjoly, 2022). Firstly, a brief exploration of the data was conducted, which grants an overview of the number of articles on AI published by various news outlets from 2014 to 2024 (*Appendix B*).

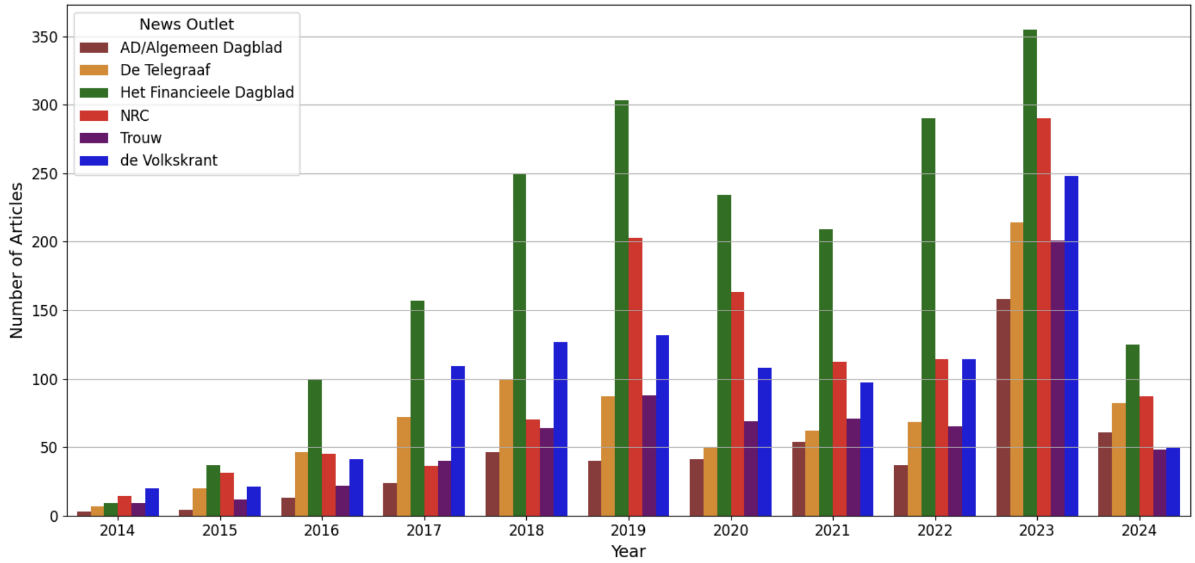

Figure 3: Number of Articles on AI by Year and News Outlet (n = 6,279)

Since some articles contained no dates, these articles were excluded from Figure 3 (*n* = 167) (Appendix C). It portrays a general increase displayed in AI-related articles across all observed media outlets, suggesting a growing interest and relevance of AI topics in news coverage over the years. Broadsheets (*het Financieele Dagblad, NRC, Trouw* and *De Volkskrant*) have written the most articles on AI. Where *Het Financieele Dagblad* has published the most articles (*n* = 2,069) in the past decade. Finally, the representation of articles decreases in the year 2024, since it covers just the first three months instead of a year.

Furthermore, all news outlets seem to decrease in number of articles from 2019 onwards. This is a short-lived decline in 2021. In 2023, the highest number of articles for each news outlet is apparent in the past decade. Broadsheets in general show a more consistent upward trend with a peak in 2023, whereas tabloids (*Algemeen Dagblad* and *De Telegraaf*) displayed moderate, but steady increases over the years. This increase in articles on artificial intelligence suggests a 'media hype' in the 10 years preceding the AI Act (Vasterman, 2014).

3.3.1. Topic Modeling (BERTopic)

The initial 6,446 retrieved news articles were used to structure the data with automated topic modeling to discover topics that categorize content on artificial intelligence. In academic literature several techniques for topic modeling have been mentioned (Egger & Yu, 2022). For instance, Cools et al. (2024) and Nguyen (2022) have implemented a LDA for topic modeling, while Zhang & Rayz, 2022 used Topic2Vec. Unlike traditional methods as LDA and Non-Negative Matrix Factorization (NMF), Bidirectional Encoder Representations from Transformers (BERTopic) offers more coherent topic representations, overcoming the limitations of the conventional bag-of-words approach and making the clustering of topics more accurate (Chen et al., 2023; Grootendorst, 2022; Wang et al., 2023). According to research by Egger and Yu (2022), BERTopic generates well-separated topics and provides novel insights that are not typically surfaced by other models. Hence, BERTopic was considered to be the best option to be implemented for analyzing news articles in the thesis at hand, because it offers enhanced performance in understanding textual nuances and defining topics.

Before the topic modeling was executed, the textual data of the combined body and headline was preprocessed to ensure quality and consistency of the model, since it operates unsupervised. Therefore, special symbols, tokenization, stopword removal and lemmatization was performed beforehand (Egger & Yu, 2022). When the data was preprocessed, the all-MiniLM-L6-v2 embedding model, known for its strong performance with Dutch texts, was used to generate embeddings. Secondly, Uniform Manifold Approximation and Projection (UMAP) was employed to reduce the dimensionality of these embeddings, followed by Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to cluster the reduced embeddings into relevant groups based on semantic similarities, excluding unrelated documents.

Finally, a class-based Term Frequency-Inverse Document Frequency (c-TF-IDF) was applied to simplify and extract topics. BERTopic automatically provided 90 topics for a broad representation of artificial intelligence reporting. As this number of topics was unmanageable for further analysis, these 90 topics were manually clustered by exploring the interface of BERTopic and combining key words presented in the representation from BERTopic that relate to each other, which resulted in 10 topics (Conway, 2006; Egger & Yu, 2022; Rüdiger et al., 2022). For instance, the key words from the articles mentioning: ‘patiënten’, ‘artsen’, ‘ziekenhuizen’ clustered into the topic *Healthcare* (*Appendix E*).

To further analyze the topics (*Appendix F*), the division by Jacobi et al. (2018) was implemented. According to Jacobi et al. (2018), there are two different clusters of topics: continuous topics and temporal topics. Continuous topics show small fluctuations through time, yet do not show a clear trend (Cools et al, 2024). On the other hand, temporal topics fluctuate over time and portray the lifespan of a topic. For this study, continuous topics are noticed, yet the emphasis is on the temporal topics, since these display the changes over time in reporting on artificial intelligence on specific issues by journalists.

The continuous topic in this study was termed *AI in Society*, inspired by Vergeer (2020), since it represents a broader societal context (Cools et al., 2024; Jacobi et al., 2018). Due to the broad scope this topic has, it was considered to not hold noteworthy portrayals for reflecting specific topics regarding artificial intelligence. The articles within this category of news outlets predominantly explore the pervasive role of AI in modern life. Commonly recurring keywords such as 'intelligentie' (intelligence), 'kunstmatige' (artificial), and 'samenleving' (society) indicate a focus on the ethical, social, and transformative implications of AI. These articles frequently discuss how AI is reshaping industries, influencing decision-making processes, and raising ethical concerns. This topic is substantially represented (n = 2,498) making up 38.8% of the analyzed content.

3.3.2. Manual Content Analysis

Before the manual content analysis was executed, a codebook was created and tested for reliability. By combining the AI Act and codebooks from previous academic research (Deng & Matthes, 2023; Kaplan & Garrick, 1981), the codebook (*Appendix G*) was created to organize the data from the AI Act and risk described in articles by journalists. The codebook consists of 4 risk levels for the AI Act: minimal (1), limited (2), high (3) and unacceptable (4). In addition, for the risks described by journalists, the same 4 risk levels are used as in the AI Act with an additional variable 'Risk Presence.' This variable contains a dichotomous value of whether a risk in the article is present (1) or not (0).

Similar to the studies conducted by Chia (2019) and Chuan et al. (2019), the unit of analysis consisted of the articles in their entirety, since it allows for a comprehensive understanding of the context, meaning, and nuances embedded in the material. By examining the full content, the complexity of the article was captured and ensured that no critical information is disregarded, which is essential for accurate and in-depth analysis (Hofman et al., 2017; Riffe et al., 2019).

In the case of the AI Act, the specific technology of artificial intelligence, such as chatbot or deep fake, was categorized with the level of risk that is stated in the AI Act (*The Act Texts | EU Artificial Intelligence Act*, 2024). For the articles of journalists, close reading of the risk levels in the articles was used to identify argumentation patterns based on the similar scale of the AI Act (Chia, 2019; Rössler, 2001). To create a framework for the described risk of journalists, 20 articles of each of the 6 news outlets (n = 120) were randomly sampled to be manually analyzed to resemble the four pillars of risks that are represented in the framework of the AI Act (Hase et al., 2020; Krippendorff, 2004).

To test whether this codebook was accurate in determining risks by the AI Act and journalists by the coders (N = 2), an intercoder reliability test was performed with the same 120 articles. The Krippendorff's alpha test was used (Hayes & Krippendorff, 2007) to estimate the intercoder reliability, since it regards two coders and ordinal data. The results show a relatively high intercoder reliability for all manually coded variables, Q1 to Q9 for Risk and AI Act (*Appendix H*), which indicates agreement between the two coders (α = 0.800). This result was due to intensive coder training and refining the coding scheme (*Appendix I*) in deliberation between the two coders (Landis & Koch, 1977).

After multiple intercoder reliability tests with randomized samples, the manual content analysis was conducted by closely reading 1,000 articles and determining the level of risk. This sample size is determined by previous studies that conducted a similar methodology (Cools et al., 2024; Vergeer, 2020; Hase et al., 2020). To have a proportionate representation of each of the topics from the entire sample (N = 6,446), the 1,000 articles were selected over the 10 topics (n = 100).

The statistical analysis is divided into four parts. Firstly, a bivariate logistic regression was conducted to determine whether or not journalists highlight risks with the dichotomous variable coined 'Risk Presence' as the dependent variable and using time (continuous), topic (dummy), news outlet (dummy), and AI Act (continuous) as independent variables.

Furthermore, an ordinal logistic regression was executed to assess how journalists rate the risk according to the AI Act, with the same independent variables and 'Risk' as coded by the journalists as the dependent variable. For both these analyses the topic *Technology* was considered as the baseline category, since it had a relatively stable and non-significant relationship with the dependent variable and it is the most common or representative category in the data (n = 821) (Krippendorff, 2019; Riffe et al., 2019). Therefore, it can effectively highlight how each specific topic diverges in terms of associated risk levels.

Since 407 articles contained no 'Risk Presence' according to journalists, these articles were included in the minimal risk category (0). To analyze the alignment between the AI Act and journalists, the variables 'AI Act' and 'Risk' were modified and updated to be able to compare the variables by executing a χ^2-test with a Gamma for symmetric variables (n = 1000) (Creswell & Creswell, 2017). This resulted in the ordinal variables that were termed by the author as 'Updated AI Act'—categorized as minimal (0), limited (1), high (2), and unacceptable (3)—and the variable called 'Updated Risk'—also categorized as minimal (0), limited (1), high (2), and unacceptable (3) (*Appendix J*).

Finally, an Ordinary Least Squares was used to investigate the discrepancy between the AI Act and journalists. This is realized by implementing a difference score by subtracting 'Updated Risk' from 'Updated AI Act' and treating it as the dependent variable, ranging from -3 to 3 (Rogosa & Willett, 1983). Time, topic, news outlet and AI Act were used as predictors.

4. Results

This section consists of the results of the mixed method approach including topic modeling and manual content analysis to address RQ1, RQ2 and RQ3.

4.1. Topics in news coverage of artificial intelligence by journalists

This section presents the temporal topics (Appendix D), excluding the continuous topic *AI in Society* (*n* = 2,498) in Figure 4. Additionally, the articles with no publication date (*n* = 167) were excluded from the analysis of topics over time (Appendix C).

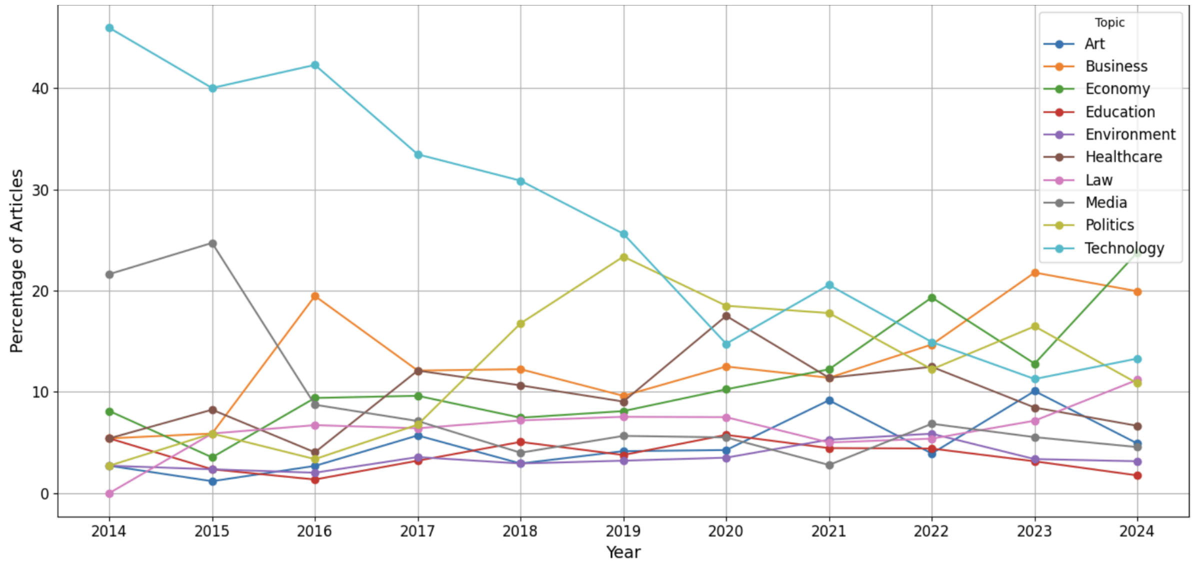

Figure 4: Percentage of Articles on Artificial Intelligence by Year and Temporal Topic (n = 3,837)

In 2014 the topic *Technology* held the highest proportion of AI articles, although its dominance decreases over time and shifts to rising topics as *Business* and *Economy*, which may be due to the representation of AI in *het Financieele Dagblad* that published the most articles. Topic *Media* shows a significant early presence in 2015, but suprisingly declines steadily throughout the years. In addition, topics *Politics* and *Healthcare* increase around 2019-2020 and diminish in representation in 2023-2024. Other topics such as *Art*, *Education*, *Environment* and *Law* remain relatively low and stable below the 10%, suggesting a moderate presence of AI in these topics. To understand these trends, a further analysis of the temporal topics is presented in Appendix F.

To answer RQ1, the current topics predominantly portrayed in the news regarding AI include *Technology*, *Politics*, *Business*, and *Economy*. The *Technology* topic remains substantial, highlighting advancements and applications as robotics and automation, though its dominance has decreased over time. Meanwhile, the *Politics* topic has gained prominence, reflecting AI's influence on international relations and governmental policies, while *Business* and *Economy* focus on corporate activities, market trends, and economic strategies. Additionally, topics like *Healthcare*, *Art*, *Law*, *Media*, *Education* and *Environment* are also discussed, showcasing AI's increasing impact across various sectors in society.

4.2. Content Analysis

The different statistical tests that were executed to answer RQ1 and RQ2 as displayed in Figure 5 are presented in this section.

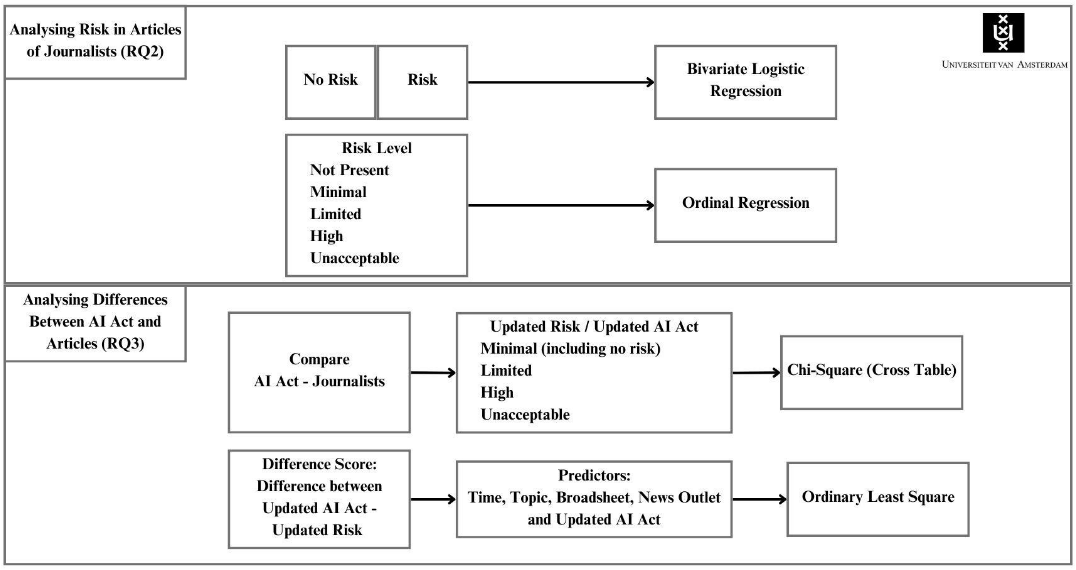

Figure 5: Overview Statistical Tests (N = 1,000)

4.2.1. Influencing factors in journalistic risk portrayals of AI

In response to RQ2, a bivariate logistic regression was performed to evaluate the impact of several factors, including publication year, topic, AI Act and type of news outlet (broadsheet or tabloid), on the likelihood of identifying the presence of risk in articles. The dichotomous variable named ‘Risk Presence’ was used as the dependent variable, consisting of present (1) or not (0). The logistic regression model's goodness-of-fit was confirmed by a significant likelihood ratio test ($\chi^2(13) = 354.039$, $p < .001$), indicating that the model with the predictors fits the data significantly. Moreover, the model's Nagelkerke R^2 value of 0.315 and the Hosmer-Lemeshow (χ^2 (8) = 7.858, $p = .447$) test also supported the model's adequacy.

Several predictors significantly affected the likelihood of risk presence (*Appendix M*). For instance, the predictor of the AI Act influenced the risk presence in articles with a odds ratio (OR) of 3.556 (b^* = 1.269, Wald = 176.456, p < .001, 95% CI: [2.949, 4.288]). Additionally, articles on the topic of *Politics* had an OR of 3.737 (b^* = 1.318, Wald = 13.659, p = <.001, 95% CI: [1.857, 7.518]), *Art* had an OR of 3.027 (b^* = 1.108, Wald = 10.003, p = .002, 95% CI: [1.923, 8.739]), *Law* had an OR of 4.099 (b^* = 1.411, Wald = 13.339, p = <.001, 95% CI: [1.923, 8.739]), *Education* had an OR of 5.114 (b^* = 1.632, Wald = 19.271, p < .001, 95% CI: [2.468, 10.599]) and *Media* had an OR of 2.396 (b^* = 0.874, Wald = 6.529, p = .011, 95% CI: [1.226, 4.685]).

Interpreting the coefficients, the results indicate that significant predictors change the odds of risk presence. For example, the coefficient for the AI Act (b^* = 1.269) corresponds to an OR of 3.556, indicating that articles influenced by the AI Act had 3.556 times higher odds of risk presence compared to those not influenced and the coefficient for *Education* (b^* = 1.632) that corresponds to an OR of 5.114, indicating a 5.114 times higher odds compared to the baseline category topic *Technology*.

To investigate the relationship between the risk associated with the AI Act and various predictors, including publication year, topics, and outlet type (broadsheet or tabloid), an Ordinal Logistic Regression was conducted (*Appendix N*). The dependent variable, 'Risk', was measured on an ordinal scale with five levels, risk 0, 1, 2, 3, and 4, and the predictor variables included the publication year, the AI Act, ten topics, and outlet type. This resulted in a model fit that was statistically significant, $\chi^2(13) = 354.039$, p < .001, suggesting that the model was effective in differentiating between levels of risk. Moreover, the Cox and Snell = .298, Nagelkerke = .315, McFadden = .120, suggest a moderate relationship between the predictors and risk levels.

Furthermore, multiple predictor variables were found to be significant in this model. For instance, the AI Act was a significant positive predictor of risk (b^* = 1.091, SE = 0.070, *Wald* = 244.866, p < .001, OR = 2.977, 95% CI [2.599, 3.414]). In addition, topics such as *Politics* (b^* = -0.814, SE = 0.269, *Wald* = 9.165, p = .002, OR = 0.443, 95% CI [0.261, 0.750]), *Art* (b^* = -0.887, SE = 0.277, Wald = 10.248, p = .001, OR = 0.412, 95% CI [0.239, 0.710]), *Law* (b^* = -0.973, SE = 0.264, *Wald* = 13.594, p < .001, OR = 0.378, 95% CI [0.225, 0.628]), *Media* (b^* = -0.801, SE = 0.274, Wald

= 8.524, p = .004, OR = 0.448, 95% CI [0.262, 0.768]), and *Education* (*b** = -0.531, *SE* = 0.274, *Wald* = 3.751, *p* = .053, OR = 0.588, 95% CI [0.344, 1.069]) were significant negative predictors.

This implies that articles on these topics are associated with lower risk levels compared to the baseline category *Technology*. Moreover, news outlet type also played a significant role, with broadsheet outlets being a marginally significant positive predictor of risk (*b** = 0.322, *SE* = 0.164, *Wald* = 3.842, *p* = .050, OR = 1.380, 95% CI [1.000, 1.905]) and tabloid outlets being a significant positive predictor (*b** = 0.540, *SE* = 0.180, *Wald* = 8.962, *p* = .003, OR = 1.717, 95% CI [1.205, 2.448]). Therefore, tabloid outlets report higher levels of risk, compared to broadsheet outlets.

Although the Pearson Chi-Square test suggested a potential lack of fit, $\chi^2(3947) = 4111.540$, $p = .033$. The overall model was statistically significant, $\chi^2(13) = 354.039$, $p < .001$, indicating that the predictors collectively explain a significant portion of the variance in risk levels. This is also supported by the Deviance test which indicated a good fit of the model to the data, $\chi^2(3947) = 2580.706$, $p = 1.000$. Consequently, the results indicate that the AI Act and several topics, particularly *Politics*, *Economy*, *Art*, *Law*, *Media*, and *Education*, were significant predictors of the risk present and being associated with tabloid outlets also significantly increased the portrayal of risk by journalists.

In conclusion, the logistic regression analysis for RQ2 demonstrates that specific topics, such as *Art*, *Education*, *Law*, and *Politics* significantly influence the portrayal of risks in AI-related reporting. This indicates that during the media coverage from 2014 to 2024, AI was perceived as a risk in these areas. Additionally, the type of news outlet significantly impacted risk portrayal, suggesting that journalists' interpretations or opinions on AI are influenced by whether the outlet is a broadsheet or tabloid. However, the variable of time did not show a significant impact, indicating that the passage of years within the study period does not have a discernible effect on the presence of reported risks.

4.2.2. Comparing journalistic risk narratives and AI Act regulatory framework

The purpose of the χ^2-test was to investigate the alignment between the variable ‘Updated AI Act’ and the variable ‘Updated Risk’. As a result, the Pearson Chi-Square test indicated a significant association between these variables $\chi^2(9, N = 1{,}000) = 238.086, p < .001$). Additionally, the Gamma confirmed a strong positive association between the 'Updated AI Act' and 'Updated Risk' with a value of .550 ($SE = .033, p < .001$). When considering the 'Updated AI Act' as the dependent variable, the value was .417 ($p < .001$), and for 'Updated Risk' as the dependent variable, it was .331 ($p < .001$). These results indicate that as the risk levels identified in the 'Updated AI Act' increase, the risk levels described by journalists (represented by 'Updated Risk') also tend to increase. Showing a strong and significant alignment between the AI Act and risks portrayed by journalists. To explore this association further an Ordinary Least Square (OLS) was executed to explore the discrepancy between 'Updated AI Act' and 'Updated Risk'(Appendix O).

To assess RQ3, various predictors were included, such as AI Act, publication year, topic, and outlet type (broadsheet or tabloid), on the dependent variable 'Difference Score' (Updated AI Act minus Updated Risk). In addition, the linearity, normality, homoscedasticity, and independence of residuals were analyzed to ensure validity of the OLS regression. Consequently, the absolute residuals were explored to detect stabilized errors, provide unbiased estimates and ensure valid standard errors. The regression model proved to be statistically significant ($F(13, 986) = 34.166, p < .001$), indicating that the predictors collectively explain a significant portion of the variance in the difference score and it contained an $R^2 = .311$ and Adjusted $R^2 = .301$.

Furthermore, the regression coefficients indicated several significant predictors of the difference score between 'Updated AI Act' and 'Updated Risk.' For example, topics such as *Politics* ($b^* = -0.300, t = -2.198, p = .028$, 95% CI [-0.568, -0.032]), *Art* ($b^* = -0.367, t = -2.587, p = .010$, 95% CI [-0.645, -0.089]), *Law* ($b^* = -0.471, t = -3.476, p < .001$, 95% CI [-0.737, -0.205]) and *Environment* ($b^* = 0.439, t = 3.206, p = .001$, 95% CI [0.037, 0.395]) were significant predictors of the difference score.

These findings suggest that certain topics significantly influence the difference score in risk levels. Specifically, topics such as *Art*, *Law* and *Politics* are associated with lower difference scores, indicating that the AI Act and the risk described by journalists align for these topics. Conversely, *Environment* and tabloids were associated with higher difference scores, indicating a greater disparity between the AI Act and the risk described by journalists. Additionally, there were no significant results for time, which indicates that this factor did not affect the difference between risk levels by journalists and the AI Act.

5. Discussion

This section includes three parts: interpreting the results, discussing their theoretical and practical implications both supported by literature to address the main research question and finally the limitations of this study:

Research Question:

"To what extent does the regulatory framework established by the AI Act align with the portrayal of AI risks in journalism, and what factors (time, news outlet and topic) influence the coverage of risk?"

5.1. Interpretation of Results

This study reveals a complex relationship between the portrayal of AI risks in journalism and the regulatory framework established by the AI Act. Overall, the AI Act and journalists align with each other strongly, which is portrayed with a positive relationship that both increase and decrease at similar risk levels measured by the significant χ^2-test. This suggests that the European AI Act can be tailored to align risk-based approaches for academic research (Hupont Torres et al., 2023).

However, the OLS yielded negative difference scores for *Art, Law* and *Politics*, emphasizing that journalists describe the risk higher than the AI Act. Surprisingly, the topic *Environment* was the only topic that showed a positive difference score, highlighting that the difference scores resemble risks described by the AI Act. This implies that journalists tend to describe the influence of AI on the environment to be beneficial, since it is used for environmental challenges, such as innovating energy management.

Furthermore, the OLS yielded a positive difference score for the news outlet type *Tabloid*, revealing that this news outlet type resembles risks portrayed by the AI Act more than broadsheets. This suggests that broadsheet newspapers tend to provide more analytical and with the AI Act risk structure aligned risk assessments, while tabloid newspapers often tend to assess the risk higher, or sensationalize AI risks, which might be interpreted as a divergence in journalistic standards and practices (Hase et al., 2020; Johansson, 2020; Roslyng & Eskjær, 2017).

Moreover, the bivariate and ordinal logistic regression portrays that *Art*, *Education, Law, Media* (only for the bivariate regression) and *Politics* yield significant results for risk presence in articles of journalists. Additionally, tabloids magnify the risks present in the articles regarding AI. This implies that while there is some alignment, journalists amplify AI risks more than the regulatory framework. Suggesting that news media spread information about risks and shape our understanding of science, technology and other societal issues, often go beyond technocentric perspectives, emphasizing social political and cultural dimensions of risk (Binder et al., 2014; Roslyng & Eskjær, 2017).

Although time had no significant influence on the alignment between journalistic coverage on AI, when the articles of journalists on AI in the past decade (n = 6,279) are put on a timeline, the results show an overall notable increase in AI-related articles over the past decade, peaking in 2023 (see Figure 3). This trend implies what Vasterman terms a 'media hype' of AI (Vasterman, 2014), probably stimulated by the growing regulatory interest, driven by significant technological advancements and social discourse on the introduction of the AI Act (*The Act Texts | EU Artificial Intelligence Act*, 2024). Additionally, it is apparent that AI has become a news subject that holds more societal driven impact, rather than solely technological.

5.2. Theoretical and Practical Implications

With regard to the theoretical implications, the findings highlight the necessity to understand the role of media in shaping perception of AI risks (Deuze & Beckett, 2022; Roslyng & Eskjær, 2017). The alignment between journalistic risk portrayals and the AI Act underscores the influence media narratives potentially have on regulatory frameworks (Kitzinger, 1999; McCarthy et al., 2008). This suggests that journalists play an important role in guiding how AI risks are communicated to regulatory bodies like the European Commission.

Moreover, the AI Act provides a statistically validated framework for understanding and analyzing the portrayal of AI risks in journalistic content. This makes it a functional measure for academic research, offering insights into how regulatory frameworks align with media narratives and societal perceptions of AI. Additionally, this study portrayed that the AI Act can be implemented for a risk-based approach (Hupont Torres et al., 2023).

Practically, these results emphasize the need for a continuous feedback loop between media portrayals and regulatory updates (Kitzinger, 1999). Such an interaction can ensure that the AI Act remains relevant and comprehensive, effectively mitigating emerging risks associated with AI technologies (Nguyen & Hekman 2022; Helberger & Diakopoulos, 2023). Furthermore, the significant differences in risk portrayal between broadsheet and tabloid newspapers suggest a lack of AI literacy (Cave et al., 2019; Deuze & Beckett, 2022). It is important to note that, over the past decade, the number of AI-related articles in broadsheets was significantly higher than in tabloids. As tabloids tend to highlight or possibly sensationalize these topics more than broadsheets, increasing AI literacy can facilitate critical evaluation of AI-related news, discerning between sensationalized and well-informed reports.

5.3. Limitations

This study has several limitations. Firstly, it only considers articles from the Netherlands, potentially limiting the understanding of AI risk portrayal across different cultural and regulatory contexts. Extending future research to other European countries or globally would provide a broader perspective.

Furthermore, *het Financieele Dagblad* (n = 2,069) contributed significantly more articles than *het Algemeen Dagblad* (n = 481) and Trouw (n = 688), emphasizing news from broadsheets (n = 5,156) over tabloids (n = 1,496). The variable named “Difference Score” suggests linearity and homoscedasticity, but also non-normality. Despite implementing various measures—such as using absolute residuals, robust standard errors, transforming the dependent variable, including additional variables or interaction terms, and weighted least squares (WLS)—to improve the model, future research should consider other methodologies for analyzing news coverage as well. Nevertheless, the results of the OLS should be interpreted with caution.

Future research may also delve deeper into the content within specific topics to understand the underlying reasons for the observed discrepancies and longitudinal studies could provide insights into how journalistic portrayals and regulatory frameworks co-evolve over time. Additionally, exploring the influence of other factors may also contribute to the comprehension of AI risk communication and the explanatory power, such as specific qualitative analysis of content within the topics, the broader regulatory environment or the impact of journalistic narratives on public opinion.

7. Conclusion

This study aimed to analyze Dutch news coverage of six prominent national news outlets ranging from tabloids to broadsheets: *de Telegraaf*, *Algemeen Dagblad*, *Trouw*, *de Volkskrant*, *NRC* and *het Financieele Dagblad* in the 10 years preceding the AI Act. In exploring the alignment of the risk portrayed in news coverage with the regulatory framework created by the EU, a nuanced relationship between these two is revealed. While there is a significant and strong alignment, variations influenced by news outlet type and specific topics highlight the complexity of AI risk communication. Suggesting that further research on AI literacy and a continuous dialogue between the media and policymakers will be crucial in navigating the evolving landscape of AI, regulations and the societal impacts.

8. Bibliography

Alba-Juez, L. (2017). *Evaluation in the headlines of tabloids and broadsheets: A comparative study. In R. Breeze (ed.) Evaluation in media discourse: European perspectives. Berlin: Peter Lang.*

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world. (n.d.). Retrieved May 11, 2024, from https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

Binder, A. R., Cacciatore, M. A., Scheufele, D. A., & Brossard, B. (2014). The role of news media in the social amplification of risk. In R. Kasperson (Ed.), *The Social Amplification of Risk* (pp. 69-82). Routledge. Retrieved from https://books.google.com/books?hl=en&lr=&id=GNtFBQAAQBAJ&oi=fnd&pg=PA69&dq=definition+of+risk+in+news+coverage&ots=s11O2Z7I5N&sig=7r71fIdb4kk9st3VsgCzTh0R6_4

Boukes, M., & Vliegenthart, R. (2020). Hoe populaire en kwaliteitskranten van elkaar verschillen in verslaggeving—De Nieuwe Reporter. https://web.archive.org/web/20201111234730/https://www.denieuwereporter.nl/2017/09/hoe-populaire-en-kwaliteitskranten-van-elkaar-verschillen-in-verslaggeving/

Bowman, N. D., Rinke, E. M., Lee, E.-J., Nabi, R., & De Vreese, C. H. (2022). How communication scholars see open scholarship. *Annals of the International Communication Association*, 46(3), 205–230. https://doi.org/10.1080/23808985.2022.2108880

Brennen, J. S., Howard, P. N., & Nielsen, R. K. (2022). What to expect when you're expecting robots: Futures, expectations, and pseudo-artificial general intelligence in UK news. *Journalism*, *23*(1), 22–38. https://doi.org/10.1177/1464884920947535

Bunz, M., & Braghieri, M. (2022). The AI doctor will see you now: Assessing the framing of AI in news coverage. *AI & SOCIETY*, *37*(1), 9–22. https://doi.org/10.1007/s00146-021-01145-9

Cacciatore, M. A., Anderson, A. A., Choi, D.-H., Brossard, D., Scheufele, D. A., Liang, X., Ladwig, P. J., Xenos, M., & Dudo, A. (2012). Coverage of emerging technologies: A comparison between print and online media. *New Media & Society*, *14*(6), 1039–1059. https://doi.org/10.1177/1461444812439061

Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence*, *1*(2), 74–78. https://doi.org/10.1038/s42256-019-0020-9

Chen, W., Rabhi, F., Liao, W., & Al-Qudah, I. (2023). Leveraging State-of-the-Art Topic Modeling for News Impact Analysis on Financial Markets: A Comparative Study. *Electronics*, *12*(12), Article 12. https://doi.org/10.3390/electronics12122605

Chia, S. C. (2019). Crowd-sourcing justice: Tracking a decade's news coverage of cyber vigilantism throughout the Greater China region. *Information, Communication & Society*, *22*(14), 2045–2062. https://doi.org/10.1080/1369118X.2018.1476573

Chuan, C.-H., Tsai, W.-H. S., & Cho, S. Y. (2019). Framing Artificial Intelligence in American Newspapers. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 339–344. https://doi.org/10.1145/3306618.3314285

Conway, M. (2006). The Subjective Precision of Computers: A Methodological Comparison with Human Coding in Content Analysis. *Journalism & Mass Communication Quarterly*, *83*(1), 186–200. https://doi.org/10.1177/107769900608300112

Cools, H., Van Gorp, B., & Opgenhaffen, M. (2024). Where exactly between utopia and dystopia? A framing analysis of AI and automation in US newspapers. *Journalism*, *25*(1), 3–21. https://doi.org/10.1177/14648849221122647

Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Deuze, M., & Beckett, C. (2022). Imagination, Algorithms and News: Developing AI Literacy for Journalism. *Digital Journalism*, *10*(10), 1913–1918. https://doi.org/10.1080/21670811.2022.2119152

de-Lima-Santos, M.-F., & Ceron, W. (2022). Artificial Intelligence in News Media: Current Perceptions and Future Outlook. *Journalism and Media*, *3*(1), Article 1. https://doi.org/10.3390/journalmedia3010002

Douglas, P. M. (1994). *Risk and Blame: Essays in Cultural Theory*. Routledge. https://doi.org/10.4324/9780203430866

Dwivedi, Y. K., Sharma, A., Rana, N. P., Giannakis, M., Goel, P., & Dutot, V. (2023). Evolution of artificial intelligence research in *Technological Forecasting and Social Change*: Research topics, trends, and future directions. *Technological Forecasting and Social Change*, *192*, 122579. https://doi.org/10.1016/j.techfore.2023.122579

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*, 886498. https://doi.org/10.3389/fsoc.2022.886498

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (AI Act) and amending certain union legislative acts*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC020

Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1), Article 1. https://doi.org/10.1609/aaai.v31i1.10635

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Hagerty, A., & Rubinov, I. (2019). *Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence* (arXiv:1907.07892). arXiv. https://doi.org/10.48550/arXiv.1907.07892

Hase, V., Engelke, K. M., & Kieslich, K. (2020). The Things We Fear. Combining Automated and Manual Content Analysis to Uncover Themes, Topics and Threats in Fear-Related News. *Journalism Studies*, *21*(10), 1384–1402. https://doi.org/10.1080/1461670X.2020.1753092

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. Communication Methods and Measures 1(1), 77-89.

Helberger, N., & Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review*, *12*(1). https://policyreview.info/essay/chatgpt-and-ai-act

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. Science, 355(6324), 486–488. https://doi.org/10.1126/science.aal3856

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hupont Torres, I., Fernández-Llorca, D., Baldassarri, S., & Gómez, E. (2023). *Use case cards: A use case reporting framework inspired by the European AI Act*.

Jacobi, C., Atteveldt, W. van, & Welbers, K. (2018). Quantitative analysis of large amounts of journalistic texts using topic modelling. In *Rethinking Research Methods in an Age of Digital Journalism*. Routledge.

Johansson, S. (2020). Tabloid Journalism and Tabloidization. In *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.877

Kaminski, M. E. (2022). *Regulating the Risks of AI* (SSRN Scholarly Paper 4195066). https://doi.org/10.2139/ssrn.4195066

Kitzinger, J. (1999). Researching risk and the media. *Health, Risk & Society*, *1*(1), 55–69. https://doi.org/10.1080/13698579908407007

Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Inc. https://doi.org/10.4135/9781071878781

Krippendorff, K. (2004). Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human Communication Research*, *30*(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Latonero, M. (2018). *Governing Artificial Intelligence*. Data & Society; Data & Society Research Institute. https://datasociety.net/library/governing-artificial-intelligence/

Lupton, D. (2013). Risk and emotion: Towards an alternative theoretical perspective. *Health, Risk & Society*, *15*(8), 634–647. https://doi.org/10.1080/13698575.2013.848847

McCarthy, J. (2007). *WHAT IS ARTIFICIAL INTELLIGENCE?* Stanford University, CA 94305.

Nguyen, D. (2023). How news media frame data risks in their coverage of big data and AI. *Internet Policy Review*, *12*(2), 1–30. https://doi.org/10.14763/2023.2.1708

Nguyen, D., & Hekman, E. (2022). The news framing of artificial intelligence: A critical exploration of how media discourses make sense of automation. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01511-1

Nilsson, N. J. (2009). *The Quest for Artificial Intelligence* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511819346

Ouchchy, L., Coin, A., & Dubljevic, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & SOCIETY*, *35*. https://doi.org/10.1007/s00146-020-00965-5

Porlezza, C. (2023). Promoting responsible AI: A European perspective on the governance of artificial intelligence in media and journalism. *Communications*, *48*(3), 370–394. https://doi.org/10.1515/commun-2022-0091

Radley-Gardner, O., Beale, H., & Zimmermann, R. (Eds.). (2016). *Fundamental Texts On European Private Law*. Hart Publishing. https://doi.org/10.5040/9781782258674

Recchia, G. (2020). The Fall and Rise of AI: Investigating AI Narratives with Computational Methods. In S. Cave, K. Dihal, & S. Dillon (Eds.), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198846666.003.0017

Riffe, D., Lacy, S., Fico, F., & Watson, B. (2019). *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (4th ed.). Routledge. https://doi.org/10.4324/9780429464287

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the Reliability of the Difference Score in the Measurement of Change. *Journal of Educational Measurement*, *20*(4), 335–343.

Rössler, P. (2001). Between Online Heaven and Cyberhell: The Framing of `The Internet' by Traditional Media Coverage in Germany. *New Media & Society*, *3*(1), 49–66. https://doi.org/10.1177/14614440122225985

Roslyng, M. M., & Eskjær, M. F. (2017). Mediatised risk culture: News coverage of risk technologies. *Health, Risk & Society*, *19*(3–4), 112.

Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence: A modern approach* (3rd ed). Prentice Hall.

Rüdiger, M., Antons, D., Joshi, A. M., & Salge, T.-O. (2022). Topic modeling revisited: New evidence on algorithm performance and quality metrics. *PLoS ONE*, *17*(4), e0266325. https://doi.org/10.1371/journal.pone.0266325

Shoemaker, P. J., Vos, T. P., & Reese, S. D. (2008). Journalists as Gatekeepers. In *The Handbook of Journalism Studies*. Routledge.

Schuett, J. (2019). A Legal Definition of AI. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3453632

Steimers, A., & Schneider, M. (2022). Sources of Risk of AI Systems. *International Journal of Environmental Research and Public Health*, *19*(6), Article 6. https://doi.org/10.3390/ijerph19063641

Sun, S., Zhai, Y., Shen, B., & Chen, Y. (2020). Newspaper coverage of artificial intelligence: A perspective of emerging technologies. *Telematics and Informatics*, *53*, 101433. https://doi.org/10.1016/j.tele.2020.101433

Titscher, S., Meyer, M., Wodak, R., & Vetter, E. (2000). Methods of text and discourse analysis: In search of meaning. Sage Publications.

The Act Texts | EU Artificial Intelligence Act. (2024). Retrieved March 14, 2024, from https://artificialintelligenceact.eu/the-act/

Vergeer, M. (2020). Artificial Intelligence in the Dutch Press: An Analysis of Topics and Trends. *Communication Studies*, *71*(3), 373–392. https://doi.org/10.1080/10510974.2020.1733038

Van Dijk, T. A. (1995). Power and the news media. *Political communication and action*, *6*(1), 9-36.

Vrabič Dežman, D. (2024). Promising the future, encoding the past: AI hype and public media imagery. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00474-x

Wang, W., Downey, J., & Yang, F. (2023). AI anxiety? Comparing the sociotechnical imaginaries of artificial intelligence in UK, Chinese and Indian newspapers. *Global Media and China*, 20594364231196547. https://doi.org/10.1177/20594364231196547

Williamson, B. (2017). *Big Data in Education: The digital future of learning, policy and practice*. 1–256.

Zhang, D., & Rayz, J. (2022). Examining Stereotypes in News Articles. *The International FLAIRS Conference Proceedings*, *35*. https://doi.org/10.32473/flairs.v35i.130642

9. Appendices

Appendix A - Corpus

News Outlet	*Number of Articles*
Algemeen Dagblad	481 (7.46%)
Trouw	689 (10.67%)
De Telegraaf	808 (12.54%)
De Volkskrant	1,067 (16.56%)
NRC	1,332 (20.67%)
Het Financieele Dagblad	2,069 (32.10%)
Total	6,446 (100%)

Appendix B - News Articles per News Outlet 2014-2024

	Year										
News Outlet	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024*
Algemeen Dagblad	3 4.8%	4 3.2%	13 4.6%	24 4.7%	46 6.3%	40 4.7%	41 6.2%	54 8.9%	37 5.4%	158 10.8%	61 13.5%
Trouw	9 14.5%	12 9.6%	22 7.8%	40 7.8%	64 8.7%	88 10.3%	69 10.4%	71 11.7%	65 9.4%	201 13.7%	48 10.6%
De Telegraaf	7 11.3%	20 16%	46 16.3%	72 14%	100 13.7%	87 10.2%	50 7.5%	62 10.2%	68 9.9%	214 14.6%	82 18.1%
De Volkskrant	20 32.2%	21 16.8%	41 14.5%	109 21.2%	127 17.3%	132 15.5%	108 16.2%	97 16%	114 16.6%	248 16.9%	50 11%
NRC	14 22.6%	31 24.8%	60 21.3%	113 21.9%	145 19.8%	203 23.8%	163 24.5%	112 18.5%	114 16.6%	290 19.8%	87 19.2%
Het Financieele Dagblad	9 14.5%	37 29.6%	100 35.5%	157 30.5%	250 34.2%	303 35.5%	234 35.2%	209 34.5%	290 42.2%	355 24.2%	125 27.6%
Total	62	125	282	515	732	853	665	605	688	1,466	453

Note. *N* = 6,446.
2014* = 9 months.
2024* = 3 months.

Appendix C - Total Number of News Articles per Topic 2014-2024 (Continuous & Temporal)

	Year											
Topic	2014*	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024*	Not Dated
AI in Society	25 40.3%	40 32%	118 44.2%	157 35.8%	281 42.8%	322 37.7%	265 39.8%	245 40.5%	279 40.6%	543 37%	167 36.9%	56 33.5%
Technology	17 27.4%	34 27.2%	63 23.6%	94 21.5%	116 17.7%	136 15.9%	59 8.9%	74 12.2%	61 8.9%	104 7.1%	38 8.4%	3 15%
Politics	1 1.6%	5 4%	5 1.9%	19 4.3%	63 9.6%	124 14.5%	74 11.1%	64 10.6%	50 7.3%	152 10.4%	31 6.8%	18 4.2%
Business	2 3.2%	5 4%	29 10.9%	34 7.8%	46 7%	51 6%	50 7.5%	41 6.8%	60 8.7%	201 13.7%	57 12.6%	8 10.8%
Economy	3 4.8%	3 2.4%	14 5.2%	27 6.2%	28 4.3%	43 5%	41 6.2%	44 7.3%	79 11.5%	118 8%	68 15%	6 4.8%
Healthcare	2 3.2%	7 5.6%	6 2.2%	34 7.8%	40 6.1%	48 5.6%	70 10.5%	41 6.8%	51 7.4%	78 5.3%	19 4.2%	3 6%
Art	1 1.6%	1 0.8%	4 1.5%	16 3.7%	11 1.7%	22 2.6%	17 2.6%	33 5.5%	16 2.3%	93 6.3%	14 3.1%	10 1.8%
Law	0 0%	5 4%	10 3.7%	18 4.1%	27 4.1%	40 4.7%	30 4.5%	18 3%	22 3.2%	66 4.5%	32 7.1%	11 6.6%
Media	8 12.9%	21 16.8%	13 4.9%	20 4.6%	15 2.3%	30 3.5%	22 3.3%	10 1.7%	28 4.1%	51 3.5%	13 2.9%	20 12%
Education	2 3.2%	2 1.6%	2 0.7%	9 2.1%	19 2.9%	20 2.3%	23 3.5%	16 2.6%	18 2.6%	29 2%	5 1.1%	7 3.6%
Environment	1 1.6%	2 1.6%	3 1.1%	10 2.3%	11 1.7%	17 2%	14 2.1%	19 3.1%	24 3.5%	31 2.1%	9 2%	25 1.8%
Total	62	125	267	438	657	853	665	605	688	1466	453	167

Note. N = 6,446.
2014* = 9 months.
2024* = 3 months.

Appendix D - News Articles of Temporal Topics 2014-2024

	Year										
Topic	2014*	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024*
Technology	17 45.9%	34 40.0%	63 42.3%	94 33.5%	116 30.9%	136 25.6%	59 14.8%	74 20.6%	61 14.9%	104 11.3%	38 13.3%
Politics	1 2.7%	5 5.9%	5 3.4%	19 6.8%	63 16.8%	124 23.4%	74 18.5%	64 17.8%	50 12.2%	152 16.5%	31 10.8%
Business	2 5.4%	5 5.9%	29 19.5%	34 12.1%	46 12.2%	51 9.6%	50 12.5%	41 11.4%	60 14.7%	201 21.8%	57 19.9%
Economy	3 8.1%	3 3.5%	14 9.4%	27 9.6%	28 7.4%	43 8.1%	41 10.2%	44 12.2%	79 19.3%	118 12.8%	68 23.8%
Healthcare	2 5.4%	7 8.2%	6 4.0%	34 12.1%	40 10.6%	48 9.0%	70 17.5%	41 11.4%	51 12.5%	78 8.5%	19 6.6%
Art	1 2.7%	1 1.2%	4 2.7%	16 5.7%	11 2.9%	22 4.1%	17 4.2%	33 9.2%	16 3.9%	93 10.1%	14 4.9%
Law	0 0.0%	5 5.9%	10 6.7%	18 6.4%	27 7.2%	40 7.5%	30 7.5%	18 5.0%	22 5.4%	66 7.2%	32 11.2%
Media	8 21.6%	21 24.7%	13 8.7%	20 7.1%	15 4.0%	30 5.6%	22 5.5%	10 2.8%	28 6.8%	51 5.5%	13 4.5%
Education	2 5.4%	2 2.4%	2 1.3%	9 3.2%	19 5.1%	20 3.8%	23 5.8%	16 4.4%	18 4.4%	29 3.1%	5 1.7%
Environment	1 2.7%	2 2.4%	3 2.0%	10 3.6%	11 2.9%	17 3.2%	14 3.5%	19 5.3%	24 5.9%	31 3.4%	9 3.1%
Total	37	85	149	281	376	531	400	360	409	923	286

Note. *n* = 3,837.
2014* = 9 months.
2024* = 3 months.

Appendix E - Table Automated Topic Modeling

News Outlet	*Assigned Topics*	*Description*
AI in Society	-1	This topic focuses on the societal impact of AI, including public perception, ethical concerns, and the potential threat or benefits AI poses to daily life.
Technology	0, 2, 9, 27, 29, 53, 57, 65, 84	Articles under this topic generally discuss technological advancements and innovations, particularly in the realm of artificial intelligence (AI).
Politics	5, 10, 11, 21, 23, 28, 40, 41, 48, 49, 66, 69, 70, 74, 79	This topic covers political aspects related to AI, including government policies, Europe, global and international relations, AI governance and policy making.
Business	4, 17, 18, 20, 26, 30, 32, 37, 38, 46, 61, 63, 71, 83	Articles in this topic discuss the impact of mainly Big Tech companies, such as ASML, OpenAI, Google, etc. on the development of AI.
Economy	3, 6, 15, 16, 43, 47, 51, 64, 68	This topic emphasizes the economic aspects of AI, which include the impact it has had on economic growth, job markets, and financial systems.
Healthcare	1, 12, 19, 33, 50, 54, 56, 59	Articles under this topic explore the role of AI in the healthcare industry, including patient care and developments in medical technology as well as ethical and privacy concerns.
Art	13, 22, 25, 36, 45, 81, 87	This topic covers the influence of AI on the arts, including the use in creative processes, digital art, and the impact on traditional artistic practices.
Law	24, 34, 35, 42, 52, 55, 60, 62, 67, 72	This topic includes discussions on legal frameworks, regulations, and ethical considerations surrounding the use and development of AI.
Media	14, 39, 85, 88, 80, 81, 87	This topic explores the role of AI in the media industry, including the effect on journalism, content creation and media consumption.
Education	31, 44, 73, 76, 77, 82, 86, 90	This topic includes articles that examine the role of AI in the educational sector, highlighting the influence it has on teaching and learning.
Environment	8, 58, 75, 78, 89	Articles on this topic address the intersection of AI and environmental issues, such as climate change, sustainability, and ecological impact.

Note. $N = 90$.

Appendix F - Topic Modeling Results

Topics	*Interpretation*	*Count*	*Most representative words*
Continuous topic			
-1	AI in Society	2,498 (38.8%)	Intelligentie (Intelligence), Kunstmatige (Artificial), AI, Toekomst (Future), Veranderingen (Changes), Systemen (Systems), Data, Samenleving (Society), Computer, Ethiek (Ethics), Algoritmen (Algorithm)
Topics with temporal patterns			
0 (base)	Technology	821 (12.7%)	Robots, Interactie (Interaction), Menselijk (Human), Digitale (Digital), Automatisering (Automation), Auto (Car), Integratie (Integration), , activiteiten, zelfrijdende (Self-Driving), Innovatie (Innovation), Drone, Banen (Work)
1	Politics	595 (9.2%)	Europese, Government (Overheid), Unie, Brussel, Democratie (Democracy), China, Amerikaanse, Duitsland (Germany), Macron, Trump, Russische (Russian)
2	Business	594 (9.2%)	Microsoft, Bedrijf (Business), Google, OpenAI, ASML, Tesla, Musk, Chip, Concurrentie (Competition), Bing, Apple, Chip
3	Economy	476 (7.4%)	Economie (Economy), Beleggers (Investers), Rente (Interest), Inflatie (Inflation), AEX, Markt (Market), Crypto, Banen (Work), Banken (Banks)
4	Healthcare	406 (6.3%)	Patiënt, Zorg (Care), Artsen (Doctors), Medische (Medical), Gezondheidszorg (Healthcare), Ziekenhuis (Hospital), Brein (Brain), Sport, Gezichten (Faces), Corona (COVID), Diagnose
5	Art	279 (4.3%)	Science Fiction, Staking (Strike), Acteurs (Actors), Script, Spotify, Film, Boek (Book), Science Fiction, Schrijver (Writer), Verhaal (Story)
6	Law	251 (3.9%)	Rechten (Law), Advocaten (lawyers), Spion (Spy), Foto (Photo), Wapens (Weapons), Deep Fakes, War (oorlog),

			Politie (Police), Drone, Vrijheid (Freedom), Phishing, Privacy, Cyber Criminaliteit (Cybercrime)
7	Media	231 (3.6%)	Game (Spel), Medium, Journalistiek (Journalism), Schrijven (Write), Nieuws (News), Nepaccounts (Fake Accounts), Trends (Trends), Desinformatie (Misinformation), Wereld (World), Interview, Facebook
8	Education	151 (2.3%)	Onderwijs (Education), Studenten (Students), Universiteit (University), Wetenschap (Science), School, Onderzoek (Research), Smartphone, ChatGPT, Taal (Language), Studie (Study)
9	Environment	144 (2.2%)	Dieren (Animals), Aardbevingen (Earthquakes), Natuur (Nature), Planten (Plants), Biodiversiteit (Biodiversity), Weer (Weather), Zee (Sea), Voorspellen (Forecast), Insecten (Insects), Dierenleed (Animal Suffering), Huisdieren (Pets)

Note. *N* = 6,446.

Appendix G - Codebook

Block 1 - AI Act

Classify each article based on close reading each topic on the risk levels from unacceptable, high, limited and minimal risk according to the AI Act.

Q1 - AI Act 4 (Unacceptable)
Can the Risk be considered to be unacceptable according to the AI Act?
Risks that pose a threat to the safety, rights and values to the certain extent that it is not able to be regulated and is therefore prohibited.

- Subliminal Manipulation: AI systems that manipulate human behavior without the person's awareness, such as subliminal techniques.
- Exploitation of Vulnerabilities: Systems that exploit the vulnerabilities of specific groups, such as children or individuals with disabilities.
- Social Scoring: AI systems used by public authorities to evaluate or classify individuals based on their social behavior, which can lead to discrimination and exclusion.
- Real-Time Biometric Identification for Law Enforcement: Use of facial recognition in public spaces for real-time identification by law enforcement, except under specific, narrowly defined conditions.
  - If No (0), directed to the next option.
  - If yes (1), directed to Block 2 - Risk.

Q2 - AI Act 3 (High)
Can the Risk be considered to be high according to the AI Act?
AI systems pose significant risks to health, safety, or fundamental rights, but these can be managed with specific regulatory requirements.

- Critical Infrastructure Management: AI systems managing energy grids or autonomous transportation systems.
- Educational Tools: AI systems that assess students or determine access to education.
- Employment and HR: Systems used in recruitment, employee management, and performance evaluation.
- Law Enforcement: Predictive policing tools and systems used in criminal investigations.
- Healthcare: AI systems used for diagnosing diseases or treatment recommendations.
  - If No (0), directed to the next option.
  - If yes (1), directed to Block 2 - Risk.

Q3 - AI Act 2 (Limited)
Can the Risk be considered to be limited according to the AI Act?
AI systems that present moderate risks and are subject to certain transparency obligations to ensure users are informed about their operations.

- Chatbots and Virtual Assistants: AI systems used in customer service that must disclose their automated nature.
- Recommendation Systems: AI that suggests products or content, where users need to be aware of the automated recommendations.
  - If No (0), directed to the next option.
  - If yes (1), directed to Block 2 - Risk.

Q4 - AI Act 1 (Minimal)
Can the Risk be considered to be minimal according to the AI Act?
Risks where AI systems are those that pose low or negligible risks to users and can be deployed with minimal regulatory oversight.

- Spam Filters: AI used in email services to filter out spam.
- Game AI: AI systems used in video games to control non-player characters (NPCs).
    - If No (0), directed to Block 2 - Risk.
    - If yes (1), directed to Block 2 - Risk.

Block 2 - Risk Presence

classify each article based on close reading of each article on the risk levels from unacceptable, high, limited and minimal risk according to the normative judgment of the journalist.

Q5 - Risk Presence
Is there a clear and explicit indication of a risk mentioned in the article? (Note: it is not enough if the word "risico" is mentioned)

- "Als iemand met kunstmatige intelligentie politiek gevoelige data maakt, komt de politie daar zo achter."
- "Sergej ontkent het bestaan van Covid-19 en meent dat toekomstige vaccinaties een dekmantel zijn van het 'satanische regime' voor het implanteren van microchips in mensen, waardoor de massa's in het vervolg via 'kunstmatige intelligentie zonder mededogen of compassie' aangestuurd zullen worden."
- "Naast al het nepnieuws doemt aan de horizon een nog veel groter gevaar op: synthetische media. Dat zijn afbeeldingen, stukken tekst, audio- of videofragmenten die zijn aangepast of zelfs helemaal gegenereerd door kunstmatige intelligentie (AI)."
- "Straks moeten we bij elk antwoord aangeven dat het door AI is gemaakt. Dan wordt het onwerkbaar een chatbot in te zetten die kan praten. Je ziet al dat Meta met zijn nieuwe app Threads niet naar Europa komt uit angst voor wet- en regelgeving."
    - If No (0), directed to the end.
    - If yes (1), directed to the next option.

Q6 - Risk Level 4 (Unacceptable)
Does the article describe the risk as unacceptable?

Risk level 4: unacceptable (uncontrolled) because the threat is not properly manageable.
Humanity is threatened by the emergence of artificial intelligence and there is no solution for this doom scenario, since it is not solved or regulated to this date. According to the journalist, action should be taken in some cases.

- "Kunstmatige intelligentie die superieur is aan de mens, en die de mensheid potentieel met uitsterven bedreigt - tot voor kort klonk dit als een sciencefictionscenario, maar met de komst van steeds slimmere Artificial Intelligence (AI) is het een serieuze waarschuwing geworden."
- "Er bestaat namelijk geen technische oplossing om de gevaren van AI van bovenmenselijk niveau onder controle te houden."
- "Van phishing via nepwebsites en malafide Marktplaatsverkopers tot aan WhatsAppfraude waarbij een fraudeur zich voordoet als een vriend of familielid: oplichters worden steeds slimmer. Zelfs zo slim dat er sinds kort een scam is waarbij de stem van een geliefde wordt gekloond door middel van kunstmatige intelligentie."

- If No (0), directed to the next option.
- If yes (1), directed to the end.

Q7 - Risk Level 3 (High)
Does the article describe the risk as high?

Risk level 3: high (little controlled) is defined but partly managed.
The article describes the risk as a significant threat to human society. However, it also attempts to argue that the risk has potential to be regulated and the risk can be partially solved.

- "De Nederlandsche Bank vreest dat sommige groepen consumenten in de toekomst geen betaalbare schade- of levensverzekering meer kunnen afsluiten. Door het gebruik van nieuwe technologie, zoals big data en kunstmatige intelligentie wordt het steeds eenvoudiger om de makkelijkst te verzekeren en dus meest lucratieve klanten eruit te pikken [...] Wat hem betreft is het onvermijdelijk dat verzekeraars geavanceerde manieren van analyse gaan gebruiken, waaronder kunstmatige intelligentie. 'Solidariteit en data kunnen heel goed samengaan.'"
- "Europa gaat Big Tech nu aanpakken, maar dat is ontzettend laat. Ze hebben een ontzettende machtspositie, er accumuleert veel kapitaal. Dat vindt zijn weg naar politieke invloed. Daar moet je paal en perk aan stellen. Nieuwe technieken zoals kunstmatige intelligentie zijn zó krachtig. Laten we er alsjeblieft goede dingen mee doen, zoals de grote problemen van onze moderne samenlevingen aanpakken."
- "Benanti vreest dat AI een nieuw voorbeeld wordt van het leegzuigen van het arme Zuiden van de wereld. De paus pleitte in zijn vredeswens voor 2024 voor het opstellen van een verdrag over het ethisch gebruik van AI. Nu krijgen we een vermoeden waar hij dat idee vandaan heeft."
    - If No (0), directed to the next option.
    - If yes (1), directed to the end.

Q8 - Risk Level 2 (Limited)
Does the article describe the risk as limited?

Risk level 2: limited (mostly controlled) is managed but still not without risk level.
The artikel presents the risk as manageable due to regulation. This also highlights the potential for positive aspects of implementation of AI. a comparison with benefits.

- "Big Data, Google and the End of Free Will. Een onwaarschijnlijke kop boven het artikel van Yuval Noah in de FT van 26 augustus. Zijn redenatie is dat kunstmatige intelligentie zo goed wordt dat het alle beslissingen voor ons gaat nemen. Wij waaien slechts mee als vaandels in de wind. Is dit het einde van de vrije wil? ... Hoe moet u nu verder als vrije wil slechts een illusie is? Dat was het al. Dus geen paniek, er is niks veranderd. Sla gewoon de bladzijde om en denk hier even aan als u weer online gaat."
- "In werkelijkheid wordt het geen van beide, maar gewoon weer een technologie als alle andere, met een hoop goede dingen en een hoop nieuwe problemen."
- "Het blijft natuurlijk wel kunstmatige intelligentie die achter de schermen werkt, wat betekent dat je de antwoorden soms met een flinke korrel zout moet nemen."
- "De Zuid-Afrikaanse miljardair Johann Rupert, de oprichter en controlerend aandeelhouder van het Zwitserse luxeconcern Richemont, gelooft dat een nieuwe revolutie aanstaande is: het werk van mensen wordt vervangen door robots. Maar dat is geen reden om te somberen. Omdat mensen meer vrije tijd krijgen, zal in zijn visie reizen steeds belangrijker worden."
    - If No (0), directed to the next option.

   - If yes (1), directed to the end.

Q9 - Risk Level 1 (Minimal)
Does the article describe the risk as minimal?

Risk level 1: minimal (controlled) with little risk left.
The article describes the risk to be nuanced due to benefits with no significant threats or fears. There are risks, but they are not entirely due to the emergence of artificial intelligence. The risk is compared with the benefits to a degree that they are equal or debunked.

- "'Mensen vragen weleens of de coach ooit wordt vervangen door kunstmatige intelligentie', zegt De Boode. 'Maar zo werkt het niet. Het helpt trainers als wij een extra perspectief geven.'"
- "Menselijke beeldende kunstenaars hoeven zich vooralsnog niet veel zorgen te maken. Hooguit krijgen ze er kunstmatige collega's bij die wellicht op een dag echt origineel werk gaan maken. Mits goed gelabeld lijkt er geen reden om dat niet in een museum te hangen. Hoe de kunsthandel er vervolgens mee omgaat, welke prijzen ze er op plakken en aan wie ze die miljoenen dan moeten betalen, dat is hun probleem."
   - If No (0), directed to the next option.
   - If yes (1), directed to the end.

Coding example: [LINK]

Risk, yes (1) or no (0)? (Distance yourself from nuances, one must look at generalizations, thus clearly presenting the risks.)

- What type of risk does the article portray based on the AI Act?
   - Deep fakes (2)
- Risk, yes (1) or no (0)? (Try to read between the lines, focus on whether the journalist describes a risk that is connected to the type of artificial intelligence)
   - Yes (1)
- What type of risk is it according to the normative judgment of the journalist?
   - High (3)

Appendix H - Krippendorff’s Alpha Reliability Estimate

Variable	*Krippendorff’s Alpha*
Q1 - AI Act 4	0.82
Q2 - AI Act 3	0.81
Q3 - AI Act 2	0.63
Q4 - AI Act 1	0.80
Q5 - Risk Presence	0.88
Q6 - Risk Level 4	0.71
Q7 - Risk Level 3	0.67
Q8 - Risk Level 2	0.83
Q9 - Risk Level 1	0.85
Total	**0.80**

Note. N = 120 articles.
Number of Coders = 2.

Appendix I - Framework AI Act

Classification (Risk-based tier)	*Description*	*Case examples*	*Variables*
Unacceptable-risk	Prohibited since it poses unacceptable risk to the fundamental rights of people and is not regulated	This encompasses employing AI for social scoring, which may result in harmful treatment, implementing emotional recognition systems within workplaces, categorizing biometric data to infer sensitive information, and conducting predictive policing on individuals, among other applications. Certain exemptions may be applicable.	• Social scoring • Exploitative AI • Emotion recognition systems in employment contexts • Biometric categorization (for sensitive information inference) • Predictive policing or forecasting criminal activity
High-risk	Permitted, although they pose a significant risk to the rights and health, it can still be regulated by the EU	Includes applications that are critical but manageable through regulatory controls to mitigate risks. These risks are closely monitored and are often already regulated by a government or other EU party	• Recruitment and Employment • (Sustainable) Energy • Transportation • Envrionment • Education • Credit Scoring • Law Enforcement • Forensic analysis • Migration and Border Control • Medical Diagnostics • General-Purpose Models
Limited-risk	Permitted, since it is based on adhering to explicit transparency and disclosure requirements in cases where usage presents a modest level of risk.	This concerns specific AI systems that engage directly with individuals, such as chatbots, and visual or audio content that have been altered by an AI system. It also entails systems that contain a lack of transparency.	• Deep fakes • Lack of Transparency • Chatbots • Generative AI • AI in Content Personalization • Recommendation systems
Minimal-risk	Permitted, without any additional requirements under the AI Act in cases where the utilization presents minimal risk.	This refers to all remaining AI systems that do not fit into the aforementioned categories, since the AI is implemented and is considered by the EU as a risk that does not pose a threat.	• Video Games • Spam Filters • Smartphone AI Assistants

Appendix J - Overview Variables

Variable	*Description*	*Type*	*Digits and Labels*
Article Number	The number of the article in the dataset	Numerical	Number
Headline	The title of the published article	String	Text
Publication	The original data that holds the general information of the article as provided by Nexis Uni	String	Text
URL	The link to the direct published article	String	Link
News Outlet	The newspaper that has published the article	String	Trouw, De Telegraaf, AD/Algemeen Dagblad, de Volkskrant, NRC, Het Financieele Dagblad
Type of News	The column the article has been published within the newspaper	String	Text
Word Count	The number of words used in the article	Numerical	Number
Body	The unprocessed text of the article in its entirety	String	Text
Published Date (Year)	The date of the published article	Metric	DD-MM-YYYY
Combined	Contains a cleaned text combination of the Headline and Body	String	Text
Topic	Each assigned topic number that is allocated through BERTopic modeling	Numerical	-1-10
Probabilities	The probability score of the corresponding dominant topic that is extracted from the BERTopic	Integer	0.0-1.0

Topic Name	The name of the corresponding dominant topic that is extracted from the BERTopic	String	Number
Representation	Specific words that are related to the allocation of the topics	String	Text
Representative_Docs	Contains all the text that corresponds to the topics	String	Topic_Text_Text
Q1 - AI Act 4	Unacceptable risk according to the AI Act	Metric	1-0
Q2 - AI Act 3	High Risk according to the AI Act	Text	1-0
Q3 - AI Act 2	Limited risk according to the AI Act	Text	1-0
Q4 - AI Act 1	Minimal risk according to the AI Act	Binary	1-0
Q5 - Risk Presence	Presence of risk (if 0, no risk present)	Binary	1-0
Q6 - Risk Level 4	Unacceptable risk according to journalism	Binary	1-0
Q7 - Risk Level 3	High risk according to journalism	Binary	1-0
Q8 - Risk Level 2	Limited risk according to journalism	Binary	1-0
Q9 - Risk Level 1	Minimal risk according to journalism	Binary	1-0
AI Act	Manually coded based on the AI Act framework	Ordinal	1-4
Risk Presence	Manually coded based on whether a risk was present in the article	Binary	1-0
Risk	Manually coded based on the risk described in journalism	Ordinal	1-4

Updated AI Act	Manually coded based on the risk described in journalism	Ordinal	0-3
Updated Risk	Manually coded based on the risk described in journalism that includes Q5 - Risk Presence as minimal (0)	Ordinal	0-3
Difference Score (AI Act - Risk)	A difference score by subtracting Risk from the AI Act	Ordinal	-3-3

Appendix K - Chi-Square Test

Statistic	*Value*
Chi-Square	238.086
Degrees of Freedom	9
P-value	3.255e-46 (<.001)

Note. *N* = 1,000.

Appendix L - Crosstabulation Updated AI Act & Updated Risk

		Risk Described by Journalist				*Total*
		0 (Minimal)	1 (Limited)	2 (High)	3 (Unacceptable)	
AI Act	0 (Minimal)	332 33.2%	25 2.5%	20 1.4%	16 1.6%	393 39.3%
	1 (Limited)	128 12.8%	32 3.2%	33 3.3%	22 2.2%	215 21.5%
	2 (High)	145 14.5%	44 4.4%	69 6.9%	37 3.7%	295 29.5%
	3 (Unacceptable)	24 2.4%	7 0.7%	21 2.1%	45 4.5%	97 9.7%
Total		629 62.9%	108 10.8%	143 14.3%	120 12.0%	1,000 100%

Note. *N* = 1,000.

Appendix M - Bivariate Logistic Regression (Risk Presence)

						95% C.I. for Exp (B)	
	B	*S.E.*	*Wald*	*P*	*Exp (B)*	*Lower*	*Upper*
Politics	1.318	.357	13.659	<.001	3.737	2.949	7.518
Business	.623	.345	3.255	.071	1.864	.948	3.666
Economy	.030	.343	.007	.931	1.030	.526	2.018
Healthcare	-.010	.336	.001	.976	.990	.513	1.913
Art	1.108	.350	10.003	.002	3.027	1.524	6.014
Law	1.411	.386	13.339	<.001	4.099	1.923	8.739
Media	.874	.342	6.529	.011	2.396	1.226	4.685
Education	1.406	.372	19.271	<.001	5.114	2.468	10.599
Environment	.226	.343	.435	.510	1.254	.640	2.456
Broadsheet	-.266	.209	1.615	.204	.767	.509	1.155
Tabloid	-.387	.227	2.897	.089	.679	.435	1.060
Year	.000	.000	.604	.437	1.000	1.000	1.000
AI Act	1.269	.095	176.456	<.001	3.556	2.949	4.288
Constant	8.150	13.698	.354	.552	3462.968		

Note. *N* = 1,000.
Baseline category = *Technology*.

Appendix N - Ordinal Regression (Risk)

					95% C.I. for Exp (B)	
	Estimate	*S.E.*	*Wald*	*P*	*Lower*	*Upper*
Risk 0	-8.041	11.587	1.155	.488	-30.751	14.668
Risk 1	-6.854	11.586	.939	.554	-29.562	15.854
Risk 2	-6.208	11.586	.830	.592	-28.915	16.499
Risk 3	-5.022	11.585	.649	.665	-27.729	17.684
Politics	-.814	.269	9.165	.002	-1.341	-.287
Business	-.251	.283	.788	.375	-.806	.304
Economy	.122	.274	.199	.655	-.414	.658
Healthcare	.341	.269	1.607	.205	-.186	.869
Art	-.887	.277	10.248	.001	-1.431	-.344
Law	-.973	.264	13.594	<.001	-1.490	-.456
Media	-.531	.274	3.751	.053	-1.069	.006
Education	-801	.274	8.524	.004	-1.338	-.263
Environment	.444	.275	2.594	.107	-.096	.984
Broadsheet	.322	.164	3.842	.050	<.001	.643
Tabloid	.540	.180	8.962	.003	.187	.894
Year	<.001	<.001	.435	.510	<.001	<.001
AI Act	1.091	.070	244.866	<.001	.955	1.228

Note. N = 1,000.
Baseline category = *Technology*.

Appendix O - Ordinary Least Squares (Difference Score)

	Unstandardized Coefficients					*95% C.I. for B*	
	B	*Std. Error*	*t*	*P*	*Standard Coefficients Beta*	*Lower Bound*	*Upper Bound*
Politics	-.300	.137	-2.198	.028	-.209	-.568	-.032
Business	-.050	.140	-.355	.723	-.214	-.325	.226
Economy	.032	.136	.233	.816	-.061	-.236	.300
Healthcare	.268	.134	1.994	.046	.015	.004	.532
Art	-.367	.142	-2.587	.010	-.212	-.645	-.089
Law	-.471	.135	-3.476	<.001	-.078	-.737	-.205
Media	-.171	.138	-1.239	.216	-.212	-.443	.100
Education	-.152	.140	-1.083	.279	-.166	-.427	.123
Environment	.439	.137	3.206	.001	.067	.170	.708
Broadsheet	.095	.084	1.138	.255	.028	-.069	.260
Tabloid	.216	.091	2.364	.018	.076	.037	.395
Year	<.001	.000	.936	.350	.025	.000	.000
AI Act	.545	.031	17.322	<.001	.483	.607	
Constant	-6.045	5.538	-1.092	.275		-16.912	4.823

Note. N = 1,000. R^2 = .311.
Baseline category = *Technology*.

Appendix P - Preregistration

The Preregistration is added as a separate file for submission.

However, it can also be accessed through the following link to OSF:

LINK TO PREREGISTRATION

Appendix Q - Preregistration Correction

During the research process, there have been slight changes made, such as leveraging BERTopic instead of LDA for topic modeling and more variables were added, such as dummy variables, to analyze the data. These minor details have not altered the main objective of the research procedure of this thesis, which was to examine to what extent AI risks in the AI Act of the EU align Dutch news papers.

Appendix R - Data Sharing Statement

During this thesis study, the primary data that was processed included articles published by multiple Dutch news outlets (*de Telegraaf*, *Algemeen Dagblad*, *Trouw*, *de Volkskrant*, *NRC* and *het Financieele Dagblad*) collected from Nexis Uni. The names of journalists are accessible by registering in Nexis Uni. Consequently, the names of these journalists are not considered publicly available information. Since the names of these journalists were considered to be irrelevant for this study, any names of journalists are excluded in the raw data to ensure confidentiality. In addition, the data does not contain any personal information that could lead to the identification of participants. Moreover, the data was securely stored on the Microsoft 365 OneDrive platform provided by the University of Amsterdam, utilizing two-step verification to protect against potential data breaches. Access to the project folder was restricted to the researcher and the supervisor only. No external organizations were involved in the study, and no personal data was collected, accessed, or transferred by third parties. The research data, excluding copyrighted materials such as the original news articles, shall not be published; however, the code used for content analysis will be made available on the Open Science Framework (OSF) after the thesis submission on the 28th of June 2024.

Appendix S - Open Science Reflection

In social sciences, the notion that there is no 'one-size-fits-all' analytical method is widely accepted (Gašević et al., 2016). Nevertheless, similar to exploratory research, EDA (Exploratory Data Analysis) often receives less attention than it deserves (Bowers & Drake, 2017; Páez & Boisjoly, 2022). For this reason, this thesis utilized an exploratory mixed methods approach for analyzing news coverage of artificial intelligence in this thesis. Furthermore, the EDA for this study involved multiple measures; such as topic modeling, content analysis and multiple statistical analyses, providing diverse analytical alternatives for a deeper understanding of AI risk portrayal in Dutch national newspapers. This enabled an exploration of various topic models and content analysis methods. Additionally, it encouraged ethical data collection practices, enhancing the study's potential for replication and publication. While this approach allows for the flexibility to select the most appropriate methods, it also presents certain challenges.

Conducting an Exploratory Data Analysis (EDA) during the preregistration process supports the selection of suitable statistical tools and techniques, sheds light on critical data characteristics, and helps formulating potential research questions. Additionally, it facilitated informed analytical choices, helping to prevent HARK-ing and p-hacking in later stages of the research process (Bakker et al., 2020). However, this approach can be regarded as a double-edged sword, since it was challenging to anticipate the research direction, making it difficult to draw conclusions, formulate hypotheses or preregister the study. Therefore, adhering to the principles of Open Source Practices (OSPs) was essential for the validity of this thesis.

Furthermore, the research process revealed that outcomes might be less favorable for publication if the p-values are not significant or other inconsistencies. In the case of this thesis, there were results that proved to be insignificant or proved to be less homoscedastic or containing normality than anticipated. Nevertheless, this led to nuanced findings that demonstrate good research practices and provide a more comprehensive picture of how complex the implementation of the AI Act is on analyzing news coverage on AI, leading to greater transparency and more depth for future research to build upon (Hofman et al., 2017). Therefore, one should question the validity of consistently significant *p*-values or leaving out pretests, as this may suggest p-hacking or HARK-ing for favorable outcomes.

Beyond EDA, data transformation plays a crucial role in verifying and ensuring the robustness of a study. Researchers alter and operationalize variables in ways that deviate from their original measurements, which can compromise the validity of their results (Deho et al., 2019). In this thesis, it was necessary to highlight the sensitivity of analytical models to different data structures, which significantly influence the study's conclusions. This underscores the need to present more than a singular dataset and provide raw materials for traceability and transparency, so errors in the data can be resolved.

Given that Communication Science is an interdisciplinary discipline, researchers prefer OSPs, such as presenting statistical notes, as optional rather than mandatory. However, I urge journals, publishers, and researchers to consider making this documentation mandatory for transparency, instead of viewing these practices as impractical. Furthermore, the attitude that only significant findings are publishable should change, as nuanced results reflect the true academic reality. Therefore, I would recommend incorporating EDA as a valid component of the OSPs.

Bibliography of Open Science Reflection

Bakker, B. N., Jaidka, K., Dörr, T., Fasching, N., & Lelkes, Y. (2020). *Questionable and open research practices: Attitudes and perceptions among quantitative communication researchers*. https://doi.org/10.31234/osf.io/7uyn5

Bowers, J., & Drake, K. W. (2017). EDA for HLM: Visualization when Probabilistic Inference Fails. Political Analysis, 13(4), 301–326. https://doi.org/10.1093/pan/mpi031

Deho, O. B., Joksimovic, S., Li, J., Zhan, C., Liu, J., & Liu, L. (2023). Should Learning Analytics Models Include Sensitive Attributes? Explaining the Why. IEEE Transactions on Learning Technologies, 16(4), 560–572. https://doi.org/10.1109/TLT.2022.3226474

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. The Internet and Higher Education, 28, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. Science, 355(6324), 486–488. https://doi.org/10.1126/science.aal3856

Páez, A., & Boisjoly, G. (2022). Exploratory Data Analysis. In A. Páez & G. Boisjoly, *Discrete Choice Analysis with R* (pp. 25–64). Springer International Publishing. https://doi.org/10.1007/978-3-031-20719-8_2