Between the Lines: Exploring Science Narratives in Bulgarian Mass Media through Embedded Topic Modeling and Sentiment Analysis

Svetlomir Zdravkov[1], Martin Jordanov Ivanov[1]

[1]Institute of Philosophy and Sociology, Bulgarian Academy of Sciences, Bulgaria

DOI 10.3217/978-3-99161-033-5-002, CC BY 4.0

Abstract. In Bulgaria, as in many other countries in Eastern Europe, the status of science in society changed rapidly after the fall of the Iron Curtain, which created an unfavorable environment for the development of the public image of science. This study explores the coverage of science related news in Bulgarian mass media from 2018 to 2023, using embedded topic modeling and sentiment analysis on over 1.7 million articles from seven major outlets with varying editorial profiles. The findings reveal that science related news is around 0,89% of total news. Also, there is significant variability in science coverage, influenced by each outlet's ideological orientation. Media with liberal and pro-European orientations prioritize scientific topics, while mainstream and government-supporting outlets show the least emphasis. Key topics include "Politics and Science", "Human Health and Biology" and "Astrophysics." Sentiment analysis highlights negative perceptions of climate change and positive views on astrophysics and technological advancements. The paper concludes with speculative explanations of these results. It notes that the results could be explained by a shift towards regulatory science in the Bulgarian context, where some scientific topics are increasingly seen as a tool for managing risks and informing policy.

1. Introduction. Institutional and media context of Bulgarian science

In Bulgaria, as in many other countries in Eastern Europe, the status of science in society changed rapidly after the fall of the Iron Curtain. Its status declined after decades of being praised as one of society's most important pillars (Petryna, 2003). According to sociologist Markku Kivinen (2002), under socialism, science was seen as an instrument of ideological supremacy, symbolizing the progress and modernization achieved under socialist rule. It was heavily promoted and funded by the state, integrated into central planning, and directed towards collective welfare and economic goals. State investment in scientific education and research underscored its importance as a tool for national development and social improvement. As in other socialist countries, science in Bulgaria

played a crucial and valued role in both society and the state (Petkova & Boyadjieva, 1994). The Bulgarian Communist Politburo quickly recognized the growing technological gap with the West and launched a centrally planned scientific and technological revolution aimed at modernizing the country and solving many of its economic and social problems, giving science a cult-like status (Ivanov, 2008).

In stark contrast, contemporary Bulgaria ranks among the least focused on science in the EU when considering its expenditure on science, research, and development as a percentage of its national budget and GDP. The average R&D intensity in the EU, measured as a percentage of GDP is 2.22% in 2022, which is around 262 euro per inhabitant. However, Bulgaria recorded an R&D intensity of just 0.77%, significantly lower than the EU average (around 27 euro per inhabitant), placing Bulgaria among the lowest in the EU for R&D expenditure (Eurostat, 2024). The disparity highlights the significant challenges Bulgaria faces in maintaining and advancing its scientific infrastructure amidst financial constraints, which have profound implications for the country's research output, scientific infrastructure, and retention of talented scientists. This shift in the status of science is starkly illustrated by Bulgaria's response to the COVID-19 pandemic, as the underfunded and neglected institutions was not recognised as an authority on the subject (Tchalakov et al., 2021). With an adult vaccination rate of only 24,9%, Bulgaria has the lowest rate of vaccination in the EU (ECDPC, 2021). The erosion of trust in science and medicine has significant repercussions for public health and the overall well-being of society.

This decline in the status of science in Bulgarian society is also influenced by the condition of the media landscape. In Bulgaria, a country known for high levels of political media parallelism and low levels of press freedom, the media landscape is heavily aligned with the political system, predominantly focusing on political advocacy (Boshnakova & Dankova, 2023). Additionally, the lack of consistent and transparent communication from scientific authorities has exacerbated public mistrust. Instead, people are increasingly turning to alternative sources of information, including social media, religious leaders, and political figures, who may not always promote scientifically accurate information (Mitev, 2021). This raises concerns about how scientific news is reported, as the political influence on media content may skew or overshadow scientific reporting. Despite the importance of this issue, there is a lack of comprehensive studies examining the impact of political media parallelism on the coverage of science-related news in Bulgaria. This gap in research highlights the need for further investigation into how media practices affect public understanding and engagement with science in politically charged environments.

Our study has two main aims. First, we aim to provide an exploratory analysis and measure how science-related news is reported in the Bulgarian media. Second, we aim to demonstrate the usefulness of a bundle of automated analysis techniques that are

underused in the field of public understanding of science (for notable exceptions, see Zorzi et al., 2023; Süerdem, 2018). The article begins with data collection, for which we used the web-scraping technique, using the Python library Selenium to extract data from online sources. This is followed by the application of BERTopic for embedded topic modeling, which allowed us to identify latent topics in the texts. We also used an GPT-3.5 Turbo to perform sentiment analysis, categorizing articles as positive, negative or neutral. We then present descriptive data. Finally, we combine the results of embedded topic modeling and sentiment analysis to map the difference in science-related news reporting in the media landscape. Our intention is to provide descriptive results from our study, so we do not go into the theoretical framework. Nevertheless, when we collected all the data, we found a pattern of science-related news coverage that clustered certain topics, media and sentiments. Therefore, we conclude by hypothesizing that these differences can be explained by the shift from science to regulatory science (Yearley, 1994), which polarizes sentiment between media.

2. Methodology

2.1. Web Scraping

In our study, we used an automated technique called web scraping, a method for systematically extracting large amounts of data from websites. The importance of this technique is highlighted in numerous academic papers that explore different dimensions of social life. For example, Lupton (2014) discusses the commodification of patient opinion in healthcare, Brooker et al. (2018) examine stigma and public discourse in the media, and Milian (2016) explores advertising practices in education. Building on these foundations, discussions by Noortje Marres and Esther Weltevrede (2013) provide deep insights into the methodological implications of scraping in social research. Marres and Weltevrede highlight that web scraping not only collects data, but also introduces pre-ordered categories of analysis embedded in the online platforms themselves.

2.2. Embedded topic modelling

After creating our initial database, we analyzed it using an advanced technique called embedded topic modeling, specifically through a package known as BERTopic (Egger & Yu, 2022; Grootendorst, 2022). Several advantages of BERTopic meet the needs of our research on Bulgarian science related news. First, BERTopic's ability to capture context through its underlying BERT model is crucial for interpreting the nuanced discussions in these forums. This contextual understanding is essential when discussions are rich in technical language and scientific discourse, allowing the model to distinguish between different uses of terms based on their context within conversations.

Secondly, the dynamic nature of BERTopic allows it to adapt to the data, identifying a range of topics from broad themes to more nuanced discussions. This adaptability is particularly beneficial in science-related news, where discussions can range from non-scientific articulations of scientists' actions to specific debates on particular scientific discoveries. BERTopic's flexibility ensures that even the most intricate details within conversations are accurately captured. In addition, BERTopic supports multiple languages, making it highly effective for processing and analyzing discussions in Bulgarian. This feature is essential for capturing the unique linguistic nuances present in forum discussions, ensuring that the analysis is both accurate and culturally relevant. Lastly, a key advantage of BERTopic is its hierarchical topic reduction capability. This feature allows us to categorize topics into a structured hierarchy, facilitating a more organized analysis of the science articles.

To identify science-related topics, we first used BERTopic to create clusters of articles based on their content. After generating these clusters, we examined the top 50 keywords generated by the algorithm and associated with each topic. By manually checking these keywords, we identified which topics were related to science. This method involved checking for the presence of terms and phrases that explicitly mentioned scientific fields (astronomy, mathematics, medicine, etc) or scientists (biologists, sociologists, etc). This careful manual classification provided a solid basis for the subsequent sentiment analysis and qualitative review, allowing us to effectively analyze the representation of science in the Bulgarian mass media.

The preliminary topic names were generated based on the top keywords extracted from each cluster. However, to ensure that these names were accurate and unambiguous, a manual review was carried out. This involved adjusting the names where necessary to ensure that they accurately reflected the content of the clusters. An essential part of this process was the reading of documents representative (these articles being in the center) of each cluster. This step was crucial in understanding the context and nuances of the topics, which allowed for more precise and meaningful topic names. By examining these representative documents, we gained deeper insights into the main themes of each cluster, which helped to refine and confirm the topic names. The final step was to assess the coherence of the topics using coherence scores. Through iteration, we refined the model to improve the quality of the topics, eventually arriving at names that were concise, descriptive and representative of the main themes within each cluster.

This type of classification technique creates categories inductively using an automated algorithm. As a result, the science-related topics do not adhere to any initial definitions or taxonomies. Therefore, this technique should be considered exploratory rather than one that allows for direct comparison. This is also the reason why we do not delve further into the definitions of the themes.

2.3. Sentiment analysis

Sentiment analysis, a crucial subfield of natural language processing (NLP), offers numerous benefits for academic research. Firstly, it enables the categorization of text into positive, negative, or neutral sentiments, providing insights into public opinion and emotional tone behind the text. This capability is invaluable across various domains, including business, politics, and social sciences, allowing researchers to understand public attitudes and responses effectively. One of the key advantages of sentiment analysis is its ability to handle vast amounts of data efficiently. By extracting meaningful information from large datasets, researchers can identify trends, patterns, and shifts in public opinion over time (Tan et al., 2023).

In the business sector, sentiment analysis helps companies gauge customer satisfaction, enhance brand reputation, and make data-driven decisions. In politics, it aids in understanding public opinion towards policies, candidates, and political parties, thus informing campaign strategies and public engagement efforts. Moreover, in finance, sentiment analysis can predict stock market trends and investment opportunities by analyzing news articles and social media posts (Feldman, 2013.). Furthermore, sentiment analysis facilitates the analysis of unstructured data, turning qualitative feedback into quantifiable insights. Overall, the integration of sentiment analysis in academic research enhances the depth and breadth of analysis, enabling a comprehensive understanding of public sentiment and its implications across various fields (Birjali et al., 2021).

We utilized a large language model (LLM) to categorize them by sentiment, adding a new analytical dimension to differentiate the scientific topics. Several sentiment extraction methods were tested, including the spaCy sentiment library. However, these methods showed low accuracy. The highest accuracy was achieved using GPT-4 Turbo, but due to its high cost, we opted for GPT-3.5 Turbo, which provided approximately 90% accuracy. This balance between cost and performance made GPT-3.5 Turbo a suitable choice for our study. Using GPT-3.5 Turbo allowed us to effectively analyze the sentiment of the articles, identifying whether the content was positive, negative, or neutral. This sentiment analysis is crucial as it enables us to understand the tone and emotional context of the science narratives within the Bulgarian media. By combining topic modeling with sentiment analysis, we can offer a comprehensive view of how science is discussed and perceived in different media outlets. We categorize every topic as negative (-1), neutral (0) and positive (+1), and then we calculate the average sentiment for every topic or media.

The integration of these advanced NLP tools not only improves the accuracy of our sentiment analysis but also enhances our ability to provide nuanced insights into media reporting of science related news. This dual-layered approach ensures a robust and

detailed analysis, contributing significantly to the field of media studies and public understanding of science.

3. Data

Our dataset consists of articles collected from seven major Bulgarian media outlets for the period between 2018 and 2023, each with distinct types and political orientations. The diversity of these sources allows for a comprehensive analysis of how different media portray science in Bulgaria. The first media outlet in our dataset is PIK, a tabloid known for its sensationalist content. We collected a total of 144,255 articles (1733 science related) from this source. Blitz, an online news platform with a pro-Russian and conservative stance, contributed a substantial 468,469 articles (6782 science related). This significant volume reflects Blitz's prolific output and strong influence in the Bulgarian media landscape. Trud, another key source, is described as clientelist with centrist and conservative leanings, and its editorial politics are vaccine-skeptic. From Trud, we gathered 284,603 articles (4221 science related). DUMA, the official newspaper of the Bulgarian Socialist Party, provided 125,773 articles (2367 science related). This outlet's pro-Russian and conservative orientation aligns with its political affiliation (Yakimova, 2022). It should be noted that due to the nearly 50-year rule of the Communist Party, conservative ideology in Bulgaria is aligned with statism, traditional values, and often the glorification of the socialist period (Konstantinov, 2024), unlike in the West, where conservatism is typically associated with free market values. Dnevnik, a liberal media outlet with a pro-Western orientation, added 173,783 articles to our dataset (3539 science related). This source is known for its critical stance towards the government and support for European and American perspectives. Nova, a mainstream media outlet that generally supports the government, contributed 265,618 articles (2774 science related). Nova's large audience and government alignment make it a significant player in the media landscape. Finally, Telegraph, a tabloid with a pro-Russian bias, added 308,662 articles (4008 science related). Its content often leans towards sensationalism and aligns with conservative viewpoints (ibid). This diverse collection of articles from varied media types and political orientations provides a rich dataset for analyzing the portrayal of science in Bulgarian mass media. The dataset includes a total of 1,771,163 articles, ensuring a comprehensive and balanced examination of how different media outlets influence public perception of science. After selecting only articles related to science, we ended up with 25424 articles.

After implementing all the procedures, we categorized the articles using several dimensions: probability, sentiment, representativeness, total publications, publication time, and topics. These categories form the basis for our analysis, detailed in the second section of our study. This multi-faceted approach allows for both qualitative and

quantitative analysis. Firstly, calculating the relative weights of each topic enables us to filter topics based on other variables, such as publication time or media source. This step is crucial for understanding the prominence and evolution of specific science topics over time. Secondly, by identifying "probability" and "representative articles," which are central to each topic cluster, we can conduct a qualitative analysis. These representative articles provide a clear picture of the core content and discourse within each topic, serving as a basis for in-depth examination. Through the text we give examples from the titles of the most representative articles according to a category (for example, negative sentiment, Dnevnik, “Animal Behavior & Domestic Care”) This methodology ensures that our analysis is not only data-driven but also contextually rich, offering insights into how different media outlets portray scientific topics. Using these comprehensive categorizations, we present a detailed analysis that highlights trends, sentiments, and the overall representation of science in Bulgarian mass media.

4. Results

4.1. Science related publication between 2018 and 2023 according to the embedded topic modeling.

This analysis examines the coverage of various scientific topics in Bulgarian media from 2018 to 2023, based on the frequency of articles and their percentage of total yearly content (fig. 1). We are not going to list all the categories due to the limitations of the article. As we can see from the fig. 1 the science related articles vary from 1,4 percent to 0,56%. In 2018 the most prominent were the topics "Politics and Science" (31.88%) and "Human Health and Biology" (16.29%). "Astrophysics & Extraterrestrial Exploration" and "Earth Sciences" also received notable attention, while "Nobels" and "Energetics & Energy Solutions" were least covered.

In 2019 the “Politics and Science" (27.07%) and "Human Health and Biology" (10.10%) remained top topics. "Astrophysics and Extraterrestrial Exploration" continued to be significant. "Nobels" and "Energetics & Energy Solutions" were again among the least covered. The focus in 2020 shifted heavily to "Human Health and Biology" (81.64%) due to COVID-19, out of 2176 articles. "Politics and Science" (28.12%) remained significant. "Nobels" and "Energetics and Energy Solutions" had minimal coverage.

In 2021 Politics and Science" led (38.35%), followed by "Human Health and Biology" (54.17%). Other significant topics included "Astrophysics & Extraterrestrial Exploration" and "Education and Science Finance" with "Nobels" and "Energetics & Energy Solutions" remaining low. In 2022 "Politics and Science" (28.18%) and "Human Health and Biology" (12.47%) were still prominent. "Astrophysics" & Extraterrestrial Exploration" and "Education and Science Finance" were notable, while "Nobels" and "Animal Behavior &

Domestic Care" had low coverage. For 2023 "Politics and Science" (31.95%) and "Human Health and Biology" (8.54%) were the most prominent. "Astrophysics & Extraterrestrial Exploration" and "Climate and Weather" were also significant, while "Nobels" and "Animal Behavior & Domestic Care" remained least covered.

In summary, the data indicates that Bulgarian media prioritizes topics related to political impacts on science and health issues, particularly during times of crisis such as the COVID-19 pandemic. This focus underscores the significant role of the media in shaping public discourse around critical scientific issues. Conversely, topics like Nobel achievements, animal behavior, and energy solutions receive relatively little coverage, highlighting potential areas for increased media attention and public engagement.

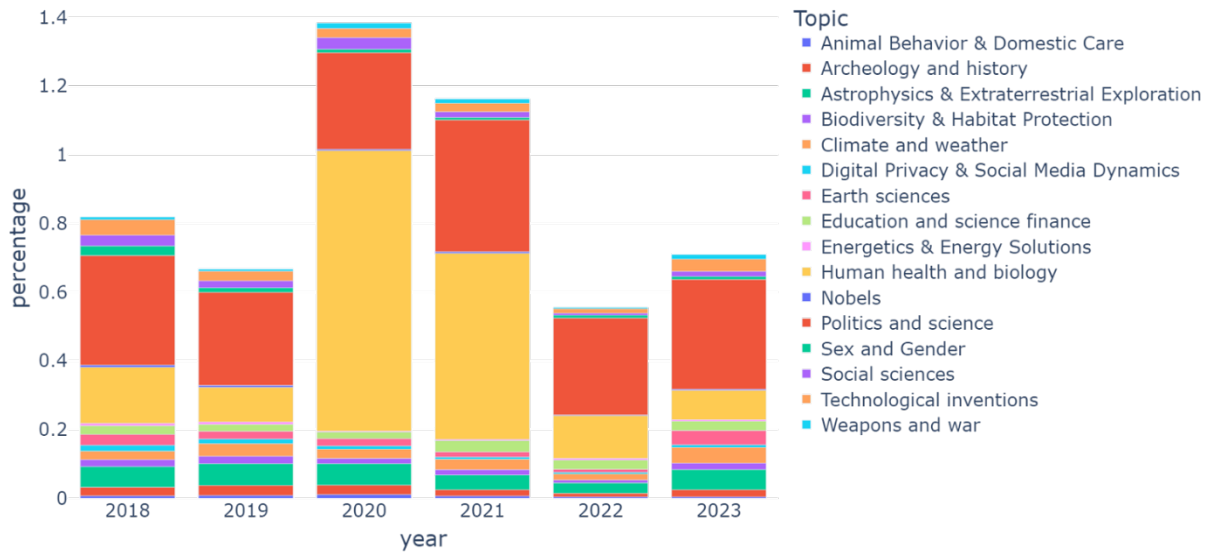

Figure 1. Percentage of science related news by year (2018-2023).

4.2. Media mapping by using BERTopic. Contrasting Approaches to Science Journalism in Bulgarian Media.

We mapped the studied media sources using a scatter plot (fig. 2). The vertical axis of this plot represents the percentage of scientific articles as part of the total articles published by each media outlet, while the horizontal axis represents the total number of articles of the media. From this mapping, it is evident that Dnevnik, a liberal media outlet, features the highest percentage of science-related news, indicating a strong emphasis on scientific reporting relative to its total content. The second highest is Duma, the socialist newspaper, which also shows a significant focus on science despite having a lower overall article count. Interestingly, NOVA, one of the most popular and mainstream media outlets, has the lowest percentage of science-related news. This disparity highlights how different editorial policies and audience targeting strategies can influence

the amount of scientific content published. This scatter plot not only visualizes the distribution of scientific articles across various media sources but also underscores the contrasting approaches to science journalism in Bulgarian mass media.

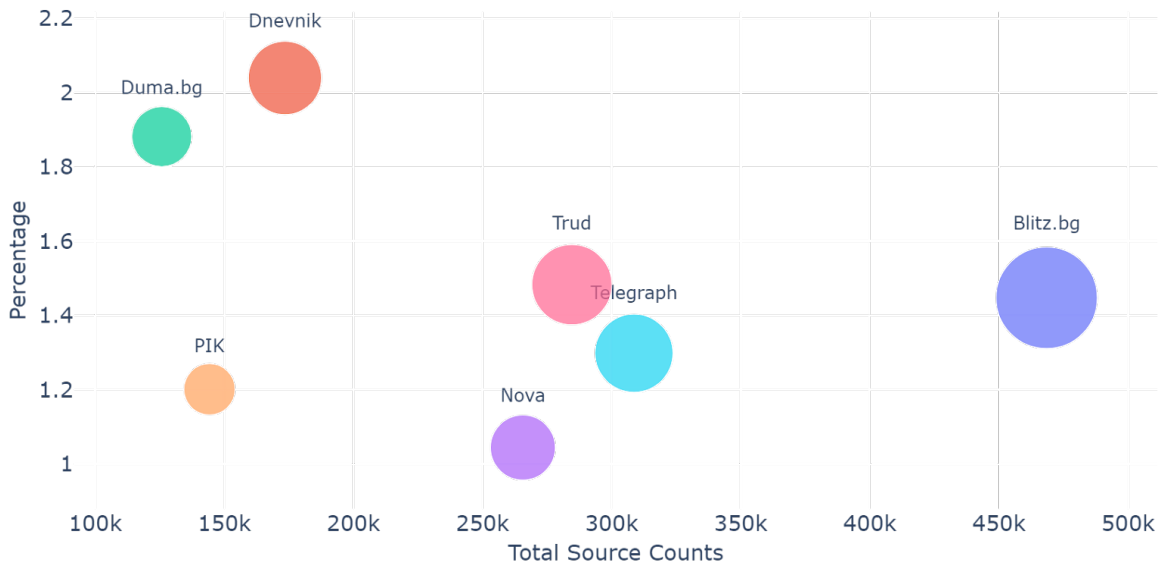

Figure 2. Map of the studied media based on number of articles and total percentage science related articles.

In our analysis, we examined articles with the highest probability levels, which are representative of each media outlet's approach to science reporting. For Blitz, a conservative media outlet, the representative article was titled "How we increased Bulgaria's population from 5.5 to 9 million thanks to scientific management during socialism." This reflects a nostalgic view of the socialist past, aligning with the outlet's statist leanings influenced by Bulgaria's totalitarian history. Dnevnik, a liberal media outlet, featured a study comparing Nazism and Communism, underscoring its anti-statist, pro-market, and pro-personal freedoms stance. This contrasts sharply with Blitz, highlighting the ideological diversity in Bulgarian media. Telegraph, another conservative and pro-Russian tabloid, showcased an article discussing how trust in vaccines was undermined by repression and censorship during the COVID-19 pandemic, indicating an anti-vaccine sentiment prevalent in its coverage.

PIK, known for its sensationalism, is represented by a study about hospital admissions due to acute alcohol intoxication, reflecting its preference for sensational scientific stories. Trud, another conservative media source, had an article on how Americans want to send Bulgarian yogurt to Mars, blending humor with a defense of Bulgarian cultural identity against globalization. Lastly, DUMA, the official newspaper of the Bulgarian Socialist Party, featured an article commemorating the USSR's launch of the first artificial Earth satellite, emphasizing its pro-Russian orientation and historical reverence for

Soviet achievements. These examples highlight the varied approaches to science reporting across Bulgarian media, influenced by each outlet's ideological stance and audience preferences. This diversity underscores the importance of context in media analysis, particularly in understanding how scientific topics are framed and presented to the public.

4.3. Analyzing Sentiment in Bulgarian Media's Science Reporting.

Our sentiment analysis of science-related topics in Bulgarian media reveals a spectrum of perceptions, from highly negative to notably positive. These sentiments provide valuable insights into public and media attitudes towards various scientific fields, reflecting broader societal concerns, interests, and hopes.

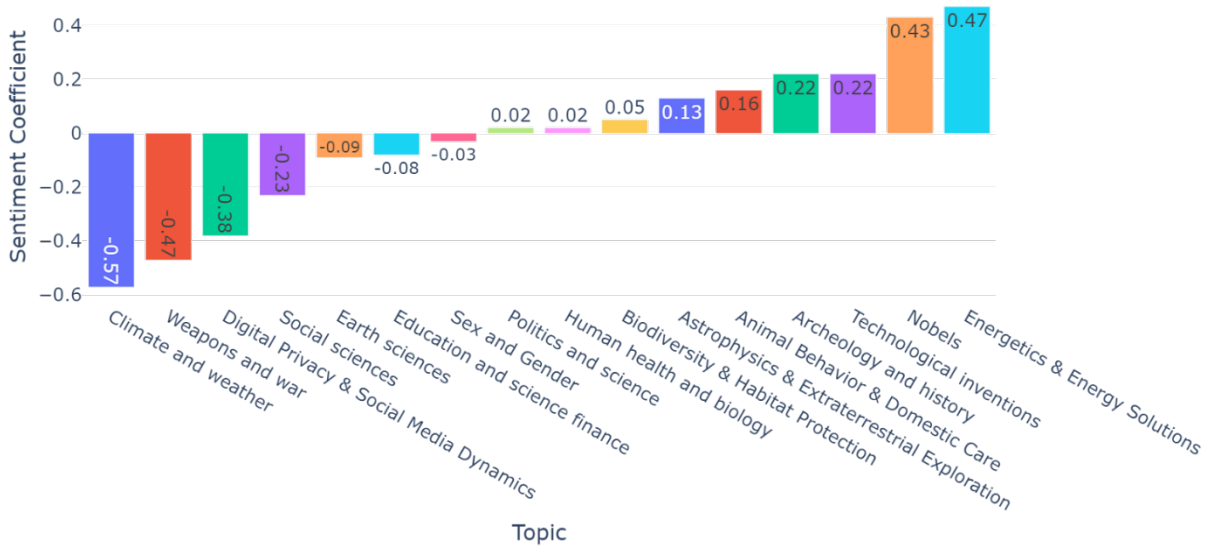

Figure 3. Sentiment analysis of the topics.

Negative Sentiment Scores

"Climate and Weather" stands out with the most negative sentiment score of -0.57. This significant negativity likely stems from widespread discussions about climate change and its dire consequences, such as extreme weather events and environmental degradation. The consistent coverage of these alarming topics underscores a deep-rooted concern for the future of our planet. Following closely is "Weapons and War", with a sentiment score of -0.47. This category naturally evokes negative emotions due to the destructive and tragic nature of warfare and military technologies. Discussions here often focus on conflicts, armament developments, and the human toll of war, contributing to the overall negative sentiment. "Digital Privacy & Social Media Dynamics" has a sentiment score of -0.38, highlighting concerns over privacy issues and the negative impacts of social media

on individual well-being and societal cohesion. The proliferation of data breaches, privacy scandals, and the mental health implications of social media use are likely drivers of this negative sentiment.

Social Sciences (sentiment score: -0.23) and Earth Sciences (-0.09) also lean towards negative perceptions. Social Sciences often involve contentious issues such as inequality, social justice, and political debates, which can provoke strong, often critical, reactions. Earth Sciences, while crucial for understanding environmental processes, are frequently associated with negative news about natural disasters and environmental degradation. Education and Science Finance (-0.08) and "Sex and Gender" (-0.03) are slightly negative, reflecting concerns over educational funding, access, and ongoing gender-related issues.

Closest to Neutral Sentiment Scores

As arbitrary as it is, we consider the scores between -0.05 to 0,05 to be neutral as in the practice score of 0,00 is rarely achieved. In this category we have “Politics and Science" and "Human Health and Biology” both have neutral sentiment scores of 0.02, indicating balanced discussions. "Politics and Science" likely covers both the positive impacts of science-informed policies and the controversies surrounding political interference in scientific matters. Similarly, "Human Health & Biology” combines the breakthroughs and advancements in medicine with ongoing health crises and challenges, resulting in a neutral overall sentiment. “Biodiversity & Habitat Protection” (0.05) edges towards positivity, reflecting the growing awareness and efforts towards conservation and environmental protection. This slight positivity suggests a cautious optimism about the effectiveness of these efforts and the public's support for biodiversity initiatives "Sex and Gender" (-0.03) is also in this category.

Positive Sentiment Scores

With the highest positive sentiment score of 0.47, “Energetics and energy solutions” reflects media enthusiasm for advancements in sustainable energy technologies. Coverage highlights innovations in renewable energy and energy efficiency, emphasizing the importance of these developments in combating "Climate" change and promoting environmental sustainability. News about Nobel Prizes scored 0.43, showcasing the media’s admiration for groundbreaking contributions in various fields. Stories often celebrate the achievements of Nobel laureates, their inspiring journeys, and the significant impact of their work, fostering a sense of pride and inspiration. "Technological Inventions" received a positive sentiment score of 0.22. Media reports focus on cutting-edge developments in areas like AI, robotics, and biotechnology, highlighting the potential benefits for improving everyday life and addressing global challenges. This reflects an optimistic view of technology’s transformative power. Also scoring 0.22,

"Archeology and History" coverage is driven by fascinating discoveries and historical insights. Positive sentiment is fueled by stories about significant archeological finds and efforts to preserve history, emphasizing the importance of cultural heritage and our connection to the past. "Astrophysics & Extraterrestrial Exploration" (0.13) evokes a positive sentiment, driven by the excitement with space exploration and astronomical discoveries, but nevertheless more on the neutral side.

Differences between media in regard to their sentiment

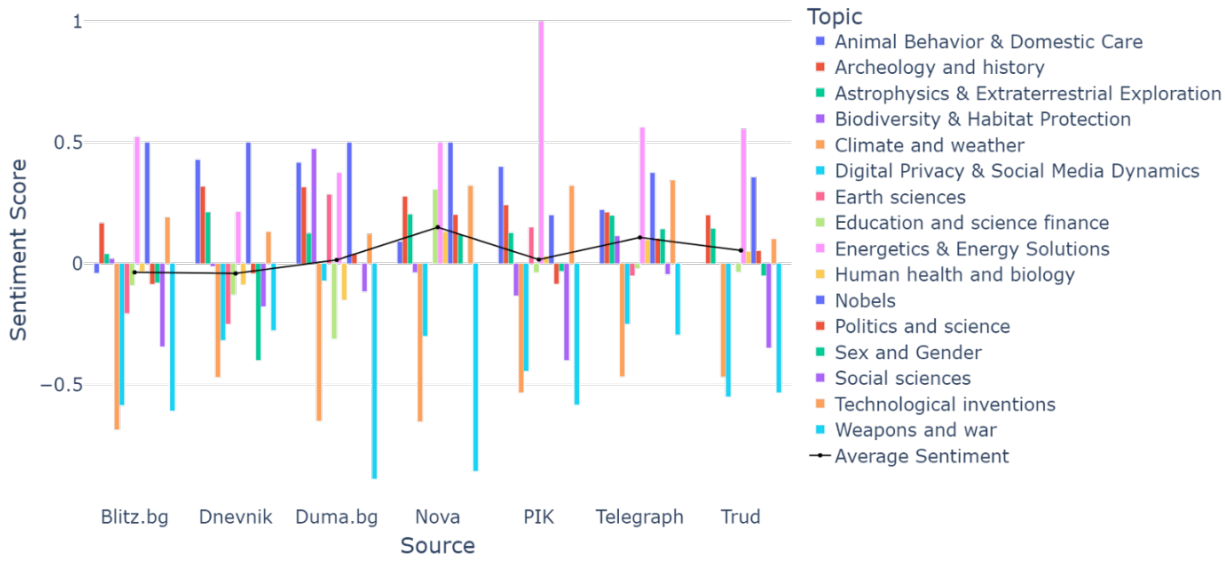

Figure 4. Sentiment Coefficients by Source and Topic.

Our sentiment analysis across various Bulgarian media sources reveals notable differences in how scientific topics are perceived and reported. Each media outlet exhibits unique editorial stances, resulting in diverse sentiments ranging from highly negative to positive. Here, we compare these sentiments, focusing on topics that show significant differentiation between the media.

Sentiments towards “Biodiversity and Habitat Protection” are mixed. Duma.bg stands out with a highly positive sentiment of 0.47, emphasizing its strong support for environmental issues, if we dwell into the articles we see that Duma is represented by an article about finding new animal species. Conversely, PIK (“DISCLOSURE: Greens made Natura 2000 in Bulgaria without scientific data”) and Nova show negative sentiments, with scores of -0.13 and -0.04, respectively, indicating less favorable coverage. Sentiments towards Earth sciences are varied, with Duma.bg showing a positive sentiment of 0.29, highlighting support for environmental sciences. Blitz.bg and Dnevnik are slightly negative, scoring -0.21 and -0.25, respectively, indicating more critical coverage.

Sentiments towards “Education and Science Finance” are mixed, with Nova displaying a positive sentiment of 0.31, emphasizing optimism about educational and scientific funding. Duma.bg shows a negative sentiment of -0.31, reflecting concerns over funding issues. These differences highlight contrasting views on the state of education and science finance across different media. Sentiments towards "Human Health and Biology” are generally neutral, but the media still differentiate. For example Dnevnik (“Two French academies have asked the state to stop funding homeopathy”) reports the topic slightly negative (-0,9) with slight positivity from Nova (“New antibiotic effective against resistant bacteria developed”), scoring 0.13, and slight negativity from Duma.bg, scoring -0.15, reflecting a balanced view of health-related topics across media. Nobel-related articles are consistently positive across all sources, with consistent high scores indicating widespread admiration for Nobel laureates. Blitz.bg, Dnevnik, and Duma.bg all score 0.50, while PIK shows lower positivity with a score of 0.20.

Sentiments towards "Politics and Science" are mixed, with Nova (“Bulgarian students with world recognition in astronomy”) and Telegraph showing positive sentiments, scoring 0.20 and 0.10, respectively, reflecting favorable views on the intersection of "Politics and Science". Blitz.bg and PIK (“BREAKING EXPOSURE ON BSP CASH: Cornelia Ninova blew nearly 6 mln on sociologists and close advisers”) are slightly negative, scoring -0.09 and -0.08, respectively, indicating more critical coverage. These mixed sentiments reflect differing perspectives on how politics influences scientific endeavors. "Sex and Gender" issues are polarized, with Telegraph (“They revealed the secret of female sex appeal”) showing a positive sentiment of 0.14, reflecting support for gender issues. Dnevnik (“In search of balance: do antidepressants stand between sex and health”) is strongly negative, scoring -0.40, indicating critical views on current sex and gender-related discussions.

5. Summary and Discussion

Our study aimed to explore how science-related news is reported in Bulgaria. The lack of funding and all the consequences from this, but also due to lack of media freedom and its preoccupation with political advocacy, create preconditions for a poor media presence of science in the media. We leveraged advanced textual data analysis methods named Embedded Topic Modelling and sentiment analysis to map and analyze the portrayal of science over a five-year period (2018-2023). The dataset consists of articles from seven major Bulgarian media outlets representing a spectrum of editorial stances, from liberal to conservative, and pro-European to pro-Russian.

Our results show that science-related articles account for about 0.89% of total articles. In comparison, the National Science Board reports that science-related news in the US

is around 2% (2014), other studies with a broader definition of science and technology news report 13.25% (comparing tokens in sci-tech news to total word tokens) in English and Italian newspapers (Zorzi et al., 2023). The results suggest the need for a more refined definition and discussion of what exactly should be considered science-related news. One of the limitations of our methodology is that it includes only topics in which the top 50 keywords contain at least one science-related term or reference to a scientist (e.g., biology, astronomy, biologist, astronomer). This rigid approach to selection reduces the number of false positives but, on the other hand, increases the number of false negatives. Furthermore, this effect is exacerbated by the technological aspect of BERTopic, which might rank these science or scientist keywords lower than the 50th keyword, excluding them from our selection. Therefore, readers should take into account that our approach is biased towards lowering the overall results.

A significant portion of these topics focuses on the topic “Politics and Science” (approximately one-third) (“Prof. Ph. Denkov [the prime minister of Bulgaria] is among the top 2% of the world's best scientists.”) which includes articles about the interactions between science and politics. Additionally, a large part of the second largest topic, “Human Health and Biology” is heavily politicized due to the political aspects of the COVID-19 pandemic. Dnevnik, with a liberal and pro-European orientation, and DUMA (official Newspaper of the Bulgarian Socialist Party), aligned with socialist and pro-Russian views, have the highest percentage of science-related news, indicating a stronger editorial emphasis on scientific reporting. In contrast, NOVA, a mainstream outlet that generally supports the government, features the lowest percentage of science-related articles, highlighting how editorial policies and audience targeting strategies influence the amount of scientific content published.

The sentiment analysis reveals that topics related to "Climate and Weather" received the most negative sentiment scores, reflecting widespread concerns about "Climate" change and its impacts. Similarly, topics like "Weapons and War", and "Digital Privacy & Social Media Dynamics" also scored negatively, highlighting public anxiety around these issues. On the other hand, "Astrophysics" & Extraterrestrial Exploration and Energy Solutions received positive sentiment scores, indicating public fascination and optimism towards these areas.

Furthermore, our methodology successfully detected very subtle differences (relative to the size of our initial database) between the studied media. While the majority of topics, such as 'Astrophysics & Extraterrestrial Exploration' (reported with neutral sentiment), 'Technological Inventions' (reported positively), and 'Climate and Weather' (reported negatively), are consistently reported across all the studied media, some topics exhibit different sentiments depending on the media outlet. 'Politics and Science' and 'Human Health and Biology' (mainly due to the controversies around COVID-19) are the most popular and, at the same time, divisive topics. Other divisive topics, though not as

popular, include 'Sex and Gender,' 'Earth Sciences,' and 'Biodiversity,' which are also part of the broader ideological and political struggles within Bulgarian society. These topics are the most politicized, and therefore, the media align their reporting on these topics according to their ideological or other orientations. For example, in the context of COVID-19 science-related news, anti-government media (PIK, Dnevnik, Duma, Blitz) report negatively, while pro-government media report with more neutral or positive sentiments. However, we will propose another explanation that draws on the notion of 'regulatory science.'

How to explain the differences in media reporting? From science-related news to regulatory science-related news.

Why are some science-related topics instrumentalized in ideological, political, and media struggles while others are not? One possible explanation can be found in the notion of regulatory science, a central theme in Science and Technology Studies (STS). According to Yearley (1994:252), “there has begun to be a switch from science being seen as a way of increasing production to a view of it as a means of handling risks and of achieving regulation.” In other words, “When science represents a heuristic for cognitively managing risk and uncertainty, it thrives; alternatively, when it is seen as a tool for bodily regulation, it suffers the stigma permeating all politically charged environments” (Gauchat, 2015:266). Being politically charged, these topics become stakes in the local political struggles.

In our case, we can detect politically loaded science related news thanks to sentiment analysis. Media reports on topics such as “Human Health and Biology”, which is its most obvious example particularly in the context of COVID-19, illustrate this shift. The focus of science shifted from understanding microbiology and biology (represented by articles about the benefits of Vitamin D) as tools to extend and save human life, to being used as political tools to regulate human behavior and the body through public institutions. The same conclusion can be drawn for the theme of “Politics and Science” where science becomes an instrument for legitimizing a wide range of political agendas (legitimacy of the prime minister for being from academia). Media coverage on this topic shows how science is often used to support or oppose political decisions, reflecting Yearley's concept of regulatory science. “Earth sciences and biodiversity” also fall into this category and have been well studied in this regard (Jasanoff, 1990). In our case it is linked to the debate of keeping biodiversity versus the farmer interests. Liberal media tends to support and advocate for the regulatory use of science, emphasizing the importance of scientific input in policy-making. In contrast, conservative media often portrays regulatory science with skepticism, highlighting concerns about government overreach and the implications of regulation on personal freedoms, which is seen as alien to Bulgarian traditions.

In comparison, topics like “Astrophysics” and “Technological Inventions” are reported positively by all media. If we follow our “regulatory science” explanation, these topics are viewed as ways of increasing production and are therefore interpreted outside of the political ideologies framework (Gauchat, 2012). Additionally, these science themes, at least in Bulgaria, are not institutionalized within the state or closely tied to the political domain, which is why they are often republished after being translated from foreign sources.

Acknowledgements

The authors would like to express their gratitude to the Bulgarian National Science Fund for the support with the funding of the project "Images of Science in the Digital World: Between Social Networks and Internet Media", with contract No. КП 06-Н55/6 of November 2021 and the base organization Institute of Philosophy and Sociology at BAS.

References

Birjali, M., Kasri, M., Beni-Hssane, A. 2021. A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl Based Syst. Aug 2021;226:107134. [doi:10.1016/j.knosys.2021.107134]

Boshnakova, D. & Dankova, D. (2023). The Media in Eastern Europe. The Media Systems in Europe, 163.

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

Eurostat. (2024) GBARD by socioeconomic objectives (NABS 2007) https://doi.org/10.2908/GBA_NABSFIN07

(ECDPC) European Centre for Disease Prevention and Control. (2021). COVID-19 Vaccine Tracker. Available here: https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#uptake-tab, accessed on 5 December 2021.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

National Science Board. (2014). Science and Engineering Indicators 2014. Arlington VA: National Science Foundation (NSB 14-01).

Feldman, R. 2013. Techniques and applications for sentiment analysis. Commun ACM. Apr 2013;56(4):82-89. doi:10.1145/2436256.2436274

Gauchat, G. (2015) The Political Context of Science in the United States: Public Acceptance of Evidence-Based Policy and Science Funding, Social Forces, Volume 94, Issue 2, December 2015, Pages 723–746, https://doi.org/10.1093/sf/sov040

Gauchat, G. 2012. Politicization of science in the public sphere: a study of public trust in the United States, 1974 to 2010. Am. Sociol. Rev. 77:167–87

Ivanov, M. (2008).Revolution of the Proletariat vs. Revolution of Technocracy. Sociological problems, 40(3-4), 269-280. [Революция на пролетариата vs. революция на технокрацията. Социологически проблеми, 40(3-4), 269-280.]

Jasanoff, S. 1990. The Fifth Branch: Science Advisors as Policy Makers. Cambridge, MA: Harvard University Press.

Konstantinov, M. 2024. "Right-Wing Leftists, Left-Wing Rightists, and Traditionalist Liberals: Core Political Values and Ideological Inconsistency at the Party-Elite Level in Bulgaria" Social Sciences 13, no. 1: 12. https://doi.org/10.3390/socsci13010012

Kivinen, M. (2002). Progress and chaos: Russia as a challenge for sociological imagination.

Mitev, T. (2021). The Health Wars: On the Crisis of Vaccine Confidence. Sociological problems, 53/2, 582 - 611

Petkova, K., & Boyadjieva, P. (1994). The image of the scientist and its functions. Public Understanding of Science, 3(2), 215.

Petkova, K. & Todorov, V. (2012). The Image of Science in Bulgaria and UK, 1992–2005: Does Generation Matter? in: Bauer, M. et al. (eds) The Culture of Science: How the Public Relates to Science Across the Globe. Routledge (Taylor&Francis). London

Petryna, A. (2003). Science and citizenship under postsocialism. Social Research: An International Quarterly, 70(2), 551-577.

Süerdem, A. (2018). Science news in Turkey: Data mining techniques for science culture mapping. In The Cultural Authority of Science (pp. 137-154). Routledge.

Tan, K.L., Lee, C.P., Lim, K.M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. Applied Sciences, 13(7):4550. https://doi.org/10.3390/app13074550

Tchalakov, I., Mileva, B., Atanasov, D. (2021). Vaccination Hesitancy: Mistrust in Medical Science or Mistrust in the State Institutions? Sociological Problems, Special:5-33. [Въздържането от ваксинация в българския контекст: недоверие в медицинската наука или недоверие в институциите на държавата? Социологически проблеми: 5-33]

Yearley, S. (1994) “Understanding Science from the Perspective of the Sociology of Scientific Knowledge: An Overview.” Public Understanding of Science 3:245–58.

Yakimova, M. (2022). Strah i Propaganda. Iztok-Zapad, Sofia.

Zorzi, V., Neresini, F., & Cammozzo, A. (2023). Public communication of technoscience in the news: A cross-linguistic Multidimensional analysis of English and Italian newspapers. Discourse & Communication, 17(6), 811-835.