This is a repository copy of *Serving the public interest? A computational analysis of the topics of UK national newspaper coverage using Freedom of Information (FOI) requests between 2005 and 2023*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/229124/

Version: Published Version

Article:

Serving the Public Interest? A Computational Analysis of the Topics of UK National Newspaper Coverage Using Freedom of Information (FOI) Requests Between 2005 and 2023

Jingrong Tong

Submit your article to this journal

Article views: 113

View related articles

View Crossmark data

Serving the Public Interest? A Computational Analysis of the Topics of UK National Newspaper Coverage Using Freedom of Information (FOI) Requests Between 2005 and 2023

Jingrong Tong

School of Information, Journalism and Communication, University of Sheffield, Sheffield, UK

ABSTRACT

The Freedom of Information (FOI) Act 2000 was enacted in the United Kingdom (UK) to improve government transparency and accountability. Requesting government information through FOI and using it in coverage support news media's watchdog role and the Act's full potential. However, the public interest served depends on the news story topics journalists pursue. This article employs Latent Dirichlet Allocation topic modelling and *n*-gram analysis to map the topics of 12,132 FOI news stories published by 10 UK national newspapers (and Sunday versions, if any) between 2005 and 2023. Despite fluctuations, FOI coverage increased overall, consistently addressing topics, such as "government and markets," "NHS problems," and "public funds and council spending issues," shaping public agendas, and holding power accountable. Broadsheets and tabloids differed and overlaped in topics, while their political leanings made less difference. Although serving the public interest, some topics were sensational, such as "sexual crimes against children," with some key areas like internal government communications and decision-making underrepresented. This research demonstrates FOI's key role in political communication, showing its potential to reform the political system and shape the relationship between the government, politicians, and news media. However, a balance is needed between fulfilling a watchdog role and avoiding sensationalism.

Introduction

As in many other countries, the Freedom of Information (FOI) Act 2000 was enacted in the United Kingdom (UK) to improve government transparency and accountability. Under this Act, public records that are not available as open data can be requested. As an active user group, journalists request information under the FOI Act themselves, use FOI data obtained by citizens, and report on FOI requests-related news. FOI coverage in the UK has shown its power in high-profile cases such as the MPs' expenses scandal in 2009.

However, a dispute has arisen between politicians and news media, with the former accusing the latter of abusing the Act to hunt for news stories, while the latter justifies their actions by citing the public interest (Mason and Watt 2015; Ministry of Justice 2011). To what extent do FOI news stories serve the public interest? This depends on what topics they cover, i.e., whether these stories are vital to democracy, capable of providing the public with essential information for their well-being and informed decision-making in political events such as elections, and effective in holding power accountable. Scholars (such as Hazell, Worthy, and Glover 2010; Worthy 2017) explored the topics of FOI news, but their studies neither systematically analysed topics by time or media types, nor considered the significant changes since their publication.

This study offers a computational analysis of the topics of the 12,132 news stories using or reporting on FOI requests (referred to as "FOI news" henceforth) between 2005 (when the FOI Act took effect) and 2023. These news stories were covered by ten UK newspapers: *The Times, The Telegraph, the Guardian, The Independent, The Financial Times, The Daily Mirror, The Metro, The Evening Standard, The Express, and The Sun*, as well as their Sunday counterparts if they had one. Latent Dirichlet Allocation (LDA) topic modelling and *n*-gram analysis were used in the analysis.

The remainder of this paper will first discuss the relationship between FOI and journalism, both generally and in the UK. Research methodology will then be discussed before presenting and discussing the findings, summarising the main argument, and laying out the conclusion. This study reveals that FOI coverage increased overall with fluctuations, consistently addressing prominent topics, such as "government and markets", "NHS problems", and "public funds and council spending issues", shaping public agendas, and holding power accountable. While topics like "NHS problems" and "public funds and council spending issues" appeared in both broadsheets and tabloids, "government and markets" was more common in broadsheets with tabloids having more sensational topics, such as "sexual crimes against children". Newspapers' political leanings, however, made little difference.

The Freedom of Information (FOI) Act, its Societal Role and Journalism

With Sweden having the world's oldest FOI Act, countries worldwide passed their FOI law at different times (Galetta 2016; Lemov and Jones 2018; Wald 1984). Despite varying enactment times, FOI laws are designed to enhance government transparency and foster government accountability. In doing so, they play a pivotal role in shaping government operations, making "governments more professional, predictable, and accountable and citizens more proactive in public life" (Michener 2011, 146).

Citizens, the primary target users of FOI laws, however, may not use it frequently. For example, in the United States (US) context, heavy users include commercial companies, law firms, and individuals who seek information about themselves rather than about governments (Kwoka 2018). This means that commercial and individual interests would be the primary force behind seeking information through FOI laws (Wagner 2022). While not heavy users, journalists actively use FOI requests. Their use is crucial for fulfilling FOI laws' goal of improving government transparency and accountability (Wagner 2022), and for helping news media inform voters and hold the government accountable (Dittmer 1973). Local news media's FOI requests can also facilitate public

engagement with local journalism practices (Mihailidis and Gamwell 2022). However, commercial requesters are more successful in receiving the requested information than "public-interest requesters," including journalists and activists, whose requests have greater potential to improve government accountability and society (Wagner and Cuillier 2025).

Despite the FOI Act is closely tied to press freedom (Ricketson and Snell 2002), the extent to which journalists can effectively use it to fulfil their role as a watchdog varies from society to society (Lidberg and Muller 2018). In the US, although the FOI Act is considered a valuable direct legal resource for investigative journalism (Shapira 2018), there are barriers preventing journalists from using it frequently. In social contexts such as the US (Fink 2018; Silver 2019), Nigeria (Ahmad et al. 2024; Antai and Umoren 2023), and India (Relly and Schwalbe 2013), journalists and news media have been found to make insufficient use of the FOI Act to play a watchdog role, primarily due to the limitations posed by contextual factors. A widespread trend has been identified in countries including the US (James 2018; Margaret 2018), Canada (Luscombe and Walby 2022), Ireland (Ryan 2015), Albania, Kosovo, and Montenegro (Camaj 2016), and Spain and Sweden (Appelgren and Salaverría 2018), where the adoption of FOI laws for journalism has declined due to governments' reluctance to release information and news media's financial difficulties in the market.

The Case of the UK

In the UK, the FOI Act was enacted in 2000 and came to effect in 2005, providing public access to information held by public authorities.[1] Journalists can send requests to obtain public records that are not available as open data under the Act. Although not targeted end-users, journalists are a core user group (Worthy 2017), particularly investigative journalists (Žuffová 2023), along with activists, NGOs and professionals (Worthy and Hazell 2017). The law has been seen as an "essential journalistic tool" (Riddell 2013; Worthy 2017, 23). For example, local journalists use it to set "the agenda" for local audiences (Brooke 2006) or to find local angles on national stories (Wyeth 2014), helping enlarge their influence. However, only a limited number of journalists would regularly dedicate time to using FOI, which proves too slow to meet journalistic deadlines (Burgess 2015a, 2015b; Worthy and Hazell 2013, 2017). Most journalists either do not use FOI or only very periodically use it.

Journalistic use of FOI requests gives publicity to information held by public bodies and draws public attention to important public issues, ranging from car parks and restaurant hygiene ratings to government spending and communication (Worthy 2017). Such revelations open political communication about governance. After the Act came into effect, FOI-based news stories soon showed power in holding power accountable by disclosing politicians' misconduct, as exemplified in the revelations of MPs' expenses scandals and the questioning of the UK government's decision to invade Iraq (both in 2009) (Hazell, Worthy, and Glover 2010). The focus of the national newspaper coverage that used FOI was on "the wasting of public money, unethical behaviour or poor performance" (Worthy 2017).

Over time, hostility from politicians and governments toward news media has grown. Politicians questioned the motivations behind journalists' FOI requests and even accused

them of creating a burden on government departments. Tony Blair saw the Act, which was passed under his government, as giving news media "a mallet" to hit them over the head (Blair 2010). News media have been accused of abusing the Act by "fishing" for stories (Ministry of Justice 2011). However, news media argue that FOI is crucial for protecting UK press freedom for enabling the coverage of public interest stories that would otherwise remain untold (Burgess 2015a, 2015b; Mason and Watt 2015).

This situation has recently worsened further. The UK government and politicians have tried to limit journalists' access to information through FOI requests (Tong 2025), particularly during the COVID-19 pandemic (Brooke 2006; Burgess 2015a, 2015b; Knight 2015; Worthy and Hazell 2017; Žuffová 2023). A clearing house was even established to block FOI requests (Evans 2020). Although the clearing house was later disbanded, FOI problems prevail (Greenwood 2021). Journalists have been complaining about their requests being rejected or about delays in FOI responses, resulting in the failure to publish news stories (Žuffová 2023). Their requests may be declined for various reasons, including data provision costs exceeding the allowed limit (Kingston, Elliott, and Thomas 2019) and privacy requirements, imposed by the 2018 General Data Protection Regulation (GDPR) (Tong 2022). The overall environment has become increasingly secretive with escalated tensions between government departments and news media. Police forces, for example, have increased their control over contact with journalists (Ponsford 2016). Journalists were arrested on various occasions involving covering protest actions (Broomfield 2024; Gayle 2022; Morrison 2023).

Apart from information access, using FOI for journalism is also influenced by news media's news values. News values, a set of criteria defining what is worth reporting on, are essential in the news selection process, underpinning news decisions (Harcup and O'Neill 2017). Highly partisan, the UK press is divided between left-wing and right-wing, and between tabloids and broadsheets. Most national newspapers are right-wing, such as *The Times*, *The Telegraph*, and *The Sun*, with left-wing newspapers such as *The Guardian* and *The Mirror* being minority in the press market. Tabloids usually focus on soft and sensational topics, such as sex, sleaze, and scandals (Reinemann et al. 2012), with "some elements of the news values of the serious press," relating to politics and society (Sparks 2000, 15). Tabloids, particularly, *The Sun*, often prioritise political news and are actively engaged in political campaigning and debate, as shown in the Brexit campaign (Javadinejad 2024; Simpson and Startin 2023). They also favour private issues, sensationalism and personalisation. Broadsheets are expected to practise serious or high-quality journalism focused on public issues, politics, economics and society (Sparks 2000). However, the distinction between tabloids and broadsheets has become less clear than it once was due to the spread of tabloidisation to quality newspapers (Umbricht and Esser 2016). Tabloidisation is evident in the adoption of a compact format, as well as in the topics, writing styles, and vocabularies used in stories. This trend is interpreted as broadsheets' response to their difficulties in the market - an attempt to engage as broad an audience as possible – since tabloids enjoy a much larger readership than broadsheets (Lefkowitz 2018). Under such circumstances, therefore, how do UK newspapers use FOI in their coverage?

Few studies have systematically examined FOI news content against the backdrop of increasing tensions between news media and politicians and the blurred distinction between UK broadsheets and tabloids. In current literature on the FOI Act, one clear

stream is about the impact of the Act on UK central and local governments (Birkinshaw 2010; Hazell, Worthy, and Glover 2010; Worthy 2013). Studies have examined the use of FOI in academic research and by various user groups such as MPs (Worthy and Hazell 2013), social scientists (Clifton-Sprigg, James, and Vujić 2020) and medical researchers (Fowler et al. 2013). Studies have also explored universities' FOI responses (Whiting, Campbell, and Awan 2021). A considerable body of literature addresses the Act's development and future (Pearce 2020; Worthy 2008; Worthy and Hazell 2017). Studies (such as Hazell, Worthy, and Glover 2010; Worthy 2017; Worthy and Hazell 2013) that discuss the topics of FOI stories have not systematically analysed news coverage across time or news media. This absence in the literature makes it important to examine FOI news coverage in the past eighteen years since 2005, when the FOI Act took effect.

Key research questions worth examining in this context include: what key topics does these newspapers' FOI news cover? Are there obvious differences between broadsheets and tabloids? Do their stories serve the public interest and hold power accountable? Which areas of transparency can their FOI news improve? Answering these questions reveals the role of FOI requests in political communication and the connection between news media and governance. It can also explain the current interactions and relationship between the UK government, politicians, and news media. The knowledge also aids in understanding the FOI Act and news media in other social contexts, where journalistic use of FOI has declined, as discussed.

Methodology

This study collected the data from the Nexis database by searching any of the keywords including "Freedom of Information Act request", "Freedom of Information request", "FOI request", and "FOIA request" in the full text of news articles published by ten UK national newspapers and their Sunday publications if they had one between 2005 and 2023. These newspapers included *The Daily Mirror*, *The Sun*, *The Times*, *The Independent*, *The Guardian*, *The Financial Times*, *The Daily Telegraph*, *The Express*, *Metro UK*, and *The Evening Standard*.[2] The sample mixed tabloids and broadsheets with left, right, and neutral political leanings (Table 1). The selection aimed to determine whether FOI coverage clearly differed among types of newspapers. Altogether, 12,132 articles were included in the final dataset for

Table 1. The backgrounds of UK newspapers.[a]

Newspapers	Type	Political stance
The Times (*Sunday Times*)	Broadsheet	Right-wing
The Guardian (*Observer*)	Broadsheet	Left-wing
The Daily Telegraph (*Sunday Telegraph*)	Broadsheet	Right-wing
The Independent	Broadsheet	Left-wing
The Financial Times	Broadsheet	Neutral
The Sun (*Sunday Sun*)	Tabloid	Right-wing
Daily Mirror (*Sunday Mirror*)	Tabloid	Left-wing
The Evening Standard	Tabloid	Right-wing
The Metro	Tabloid	Neutral
The Express	Tabloid	Right-wing

[a]The types and political stances of the newspapers are decided according to "UK media bias" published by PressGazette, accessed at (September 1, 2024): https://pressgazette.co.uk/media-audience-and-business-data/media_metrics/uk-media-bias-2024/

analysis after removing duplicates and articles from foreign news, opinion, editorial, and letter sections. Python scripts automatically compared the entries in the dataset and removed duplicates first, and then the ones removed were double-checked manually.

Traditionally, news coverage is analysed by researchers manually. Computational methods based on Natural Language Processing (NLP) techniques have shown effectiveness in detecting, identifying and portraying topics, sentiments, and networks in texts (Jelodar et al. 2020). They enable researchers to analyse large-scale texts that are impractical or resource-heavy to examine manually. Often, programming languages such as Python or R are used to perform the analysis. NLP libraries such as NumPy, NLTK, Gensim, SpaCy, and pyLDAvis can be built in code to handle NLP tasks such as pre-processing texts, calculating the frequencies of words, computing similarity between documents and words, word embeddings, and visualising the results.

While NLP analysis of news requires at least basic computing skills, social science scholars are increasingly adopting it alongside traditional manual methods. Topic modelling, such as Latent Dirichlet Allocation (LDA) and structured topic modelling (STM), are commonly used to identify topics in a text (Field et al. 2018; Jacobi, Atteveldt, and Welbers 2015; Luscombe and Walby 2022; Mellado et al. 2021; Turetsky and Riddle 2018; Züll and Landmann 2004).

The data in this study exceeded the researcher's capability to perform manual analysis. Therefore, this project used Latent Dirichlet Allocation (LDA) topic modelling and *n*-gram analysis to map the topics of these news articles. The existing studies mentioned above served as a valuable foundation for developing the research design of this study.

As "a generative probabilistic model," LDA explores hidden topics in usually large-scale text corpora based on probabilities (Blei, Ng, and Jordan 2003, 993; Blei, Ng, and Jordan 2001). It groups concurrent words into topics. A topic, suggested by high probability words, can be distributed across documents, while a document can be "associated with multiple topics" (Blei, Ng, and Jordan 2003). Since Blei and his colleagues introduced it in 2001 and 2003, LDA has been widely adopted as a type of (unsupervised) topic modelling in mining large-sized text by scholars in different disciplines ranging from software engineering and linguistic science to geography and political science to medical research (Jelodar et al. 2019) and journalism, media, and communication studies (Maier et al. 2021). Other probabilistic topic models such as Correlated Topic Model (CTM) (Blei and Lafferty 2007) or STM (Roberts, Stewart, and Tingley 2019) build on LDA, allowing different functions and addressing the perceived limitations of LDA. While CTM enables the direct modelling of the correlation between topics, STM considers document-level metadata (covariates) such as years and newspapers in the modelling process. However, STM is usually implemented in R. Adapting the STM R package to Python code would compromise some important elements in STM, such as integrating document-level metadata into the modelling process.

Considering the aim of the research and the skills of the researcher, it was decided to use LDA in Python to identify key topics in the whole corpus based on word's probabilities. Metadata was then associated with identified topics in the post-modelling analysis to examine changes in topic distribution over time and across newspapers, as well as topic correlation networks (Lu and Nemet 2020; Roberts, Stewart, and Tingley 2019). Using LDA kept the model simple with more flexibility for analysing how metadata might have influenced topic prevalence or distribution in the post-modelling stage.

However, the findings do not reveal the correlation between topics as explicitly as CTM does, while topic prevalence and content are not adjusted based on metadata in the modelling process, as is the case with STM.

n-gram analysis is another computational method in NLP and text analysis. An *n*-gram is a sequence of *n* consecutive grams such as letters, part-of-speech tags, or words (words in this study) in a given text (Cavnar and Trenkle 1994). *n* can be 1 (unigram), 2 (bigram), 3 (trigram) or other numbers. An *n*-gram analysis chunks the text into grams and calculates their frequencies to identify (or predict) patterns of a text, such as sentiments, semantics, and networks, or to classify documents (Cavnar and Trenkle 1994). Traced back to 1913 (Markov 2006 (originally 1913)), it has been widely used in diverse domains including medicine (Rahimian et al. 2019), the economy (Ojo et al. 2021), law (Mazzi 2018), and education (Rudniy and Elliot 2016).

The two methods have their distinct purposes, strengths, and limitations. LDA is good at automatically discovering hidden topics across a large number of documents, based on the probability of words. The insights provided by the topics go beyond word frequency, as grouping words into topics captures the semantic relationships between words or topics. However, it ignores word orders and is prone to the choices made in tuning the model, such as the number of topics and perimeters. In contrast, an *n*-gram analysis respects the sequence of (adjacent) words and breaks large chunks of data into small, meaningful pieces of information. By capturing the frequencies of grams in a text, it can offer insights into the content and nature of the text. However, it cannot provide knowledge of semantic relationships between words or underlying topics, which LDA can offer. For these reasons, this study adopted both methods, which can complement each other to improve results.

As headlines give us the gist of the content in concise and succinct language, an *n*-gram analysis of all headlines can reveal the topics of the coverage. However, LDA topic modelling of the full text is seen as more accurate than that of the abstracts (Syed and Spruit 2017). Therefore, the LDA topic modelling was used to analyse the full text of all the news articles. The findings from both parts of the analysis were also used to corroborate each other.

The research process of the LDA topic modelling has the following steps: (1) data pre-processing and preparation; (2) training of the model; (3) validation and interpretation of the topics, and (4) further analysis of the topics. In the data pre-processing and preparation step, the data was first pre-processed or normalised, including removing stopwords, tokenisation, removing punctuations, lowercasing words, and lemmatisation. The stopwords included in *nltk.corpus* - a module in the NLTK (Natural Language Toolkit) library (a popular Python library for NLP), a spelling error "hosipital," and words[3] frequently appearing but not adding meaningful meanings in this context of FOI coverage were removed to reduce the noise and increase the accuracy of the modelling. Only nouns, adjectives, verbs and adverbs were kept in the final corpus for analysis. This study used lemmatisation, a procedure to take away the additional suffixes or grammatical variations added to a word, returning it to its base or dictionary form (Balakrishnan and Lloyd-Yemoh 2014; Maier et al. 2021). In the process, bigrams were formed to improve the accuracy of the topic model as bigrams capture the context and meaning of words and are more meaningful than individual words.

After the corpus pre-processing and preparation step, dictionary was created, providing unique ID for each word in the corpus, and corpus, representing documents with

word IDs and their frequencies, for the LDA modelling using the Bag of Words (BoW) model in Gensim, a Python library for topic modelling and NLP. Mean coherence scores (García Avilés et al. 2006; Liu et al. 2020; Nastiti, Hidayatullah, and Pratama 2021; Wang, Feng, and Dai 2018) were calculated to determine the optimal number of topics, which is 32 with a mean coherence sore of 0.429. After having identified the mean coherence score, the dictionary and the corpus were used to build the LDA model in Gensim. This study used both coherence scores and human judgements to validate and interpret the topics generated by the LDA algorithm, an approach suggested by other scholars (such as DiMaggio, Nag, and Blei 2013; Levy and Franklin 2014; Maier et al. 2021; Quinn et al. 2010). Visualisation tools such as WordCloud and pyLDAvis were used to present the topics for validation and interpretation. Later, the topic distribution for each document and topic contribution percentages were calculated and the most representative texts were identified for further analysis. Understanding gained from reading the most representative texts, exploring the corpus, the findings from the *n*-gram analysis, and the researcher's domain knowledge were also used for the validation purpose. The topics were later labelled and analysed statistically to understand their distribution and correlations (Lu and Nemet 2020; Roberts, Stewart, and Tingley 2019). That means, after key topics had been identified, their temporal trends or trends for different newspapers were examined by calculating the distribution of individual topic probabilities across time or newspapers. The correlation between topics on the document-topic distribution level were calculated to map the networks of the topics. This means, the concurrence of two topics in documents was seen increasing their correlation (Chae and Park 2018).
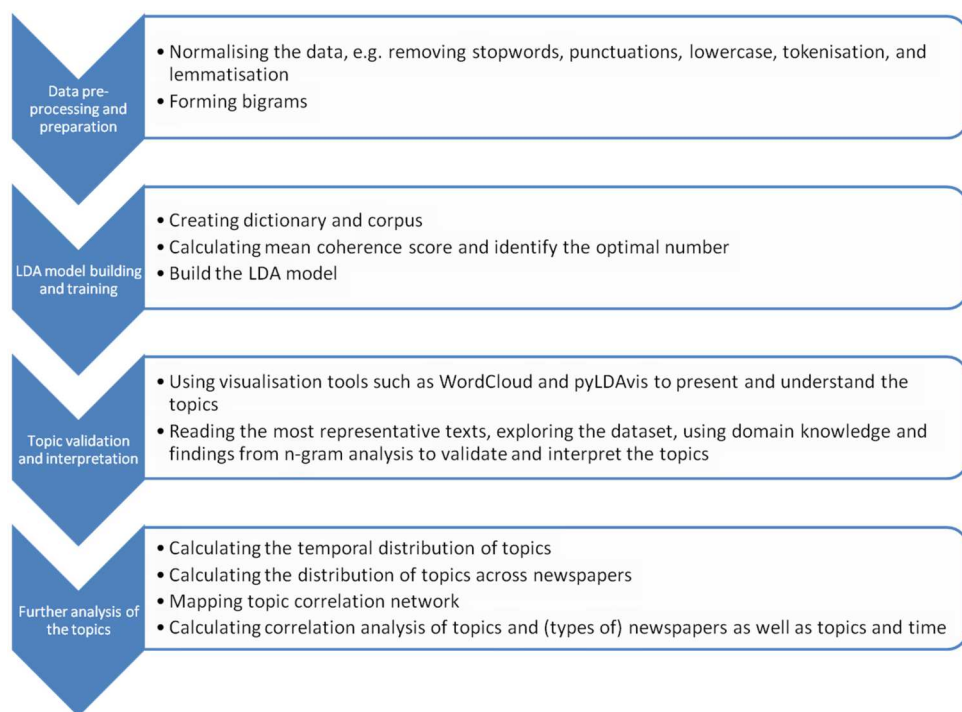
Figure 1. The LDA research process.[8]

Topics and their proportions were shown as nodes while the edges between nodes showed the correlations between topics and correlation coefficients. Only those edges with correlation coefficients being 0.09 and above were included. The correlations between topics and newspapers were also calculated. The LDA research process can be seen in Figure 1.

The *n*-gram analysis is simpler compared to the LDA topic modelling. It follows three steps: (1) data pre-processing and preparation; (2) data analysis; (3) presentation and visualisation of the gram lists. Similarly, the data pre-processing and preparation stage involved the removal of noises as discussed above, removing punctuations and tokenising the data. But for the *n*-gram analysis, only nouns were retained for the analysis, as they indicate the subjects the articles focus on and are dominant in language, conveying meanings that offer "unexpected insights into the ways people conceptualise, categorise, and order the world around them" (Ye 2017, 2017: 2). The data analysis focused on unigrams and bigrams. Their frequencies and distributions over time and per newspapers were also analysed statistically.

In this study, the tasks involved in the process of data collection, pre-processing, preparation, analysis, and presentation were implemented by Python scripts (Python 3.10.2 was used) with Python libraries for NLP tasks such as NLTK, Gensim, NumPy, and SpaCy.

Findings

Overview

The total number of FOI news articles gradually increased between 2005 and 2023, experiencing some fluctuations along the way (see Figure 2). After the Brexit vote in 2016, the number of stories grew rapidly, peaking in 2018 - the year when the GDPR came into force - followed by a sharp drop leading to 2020, when the Brexit deal was reached
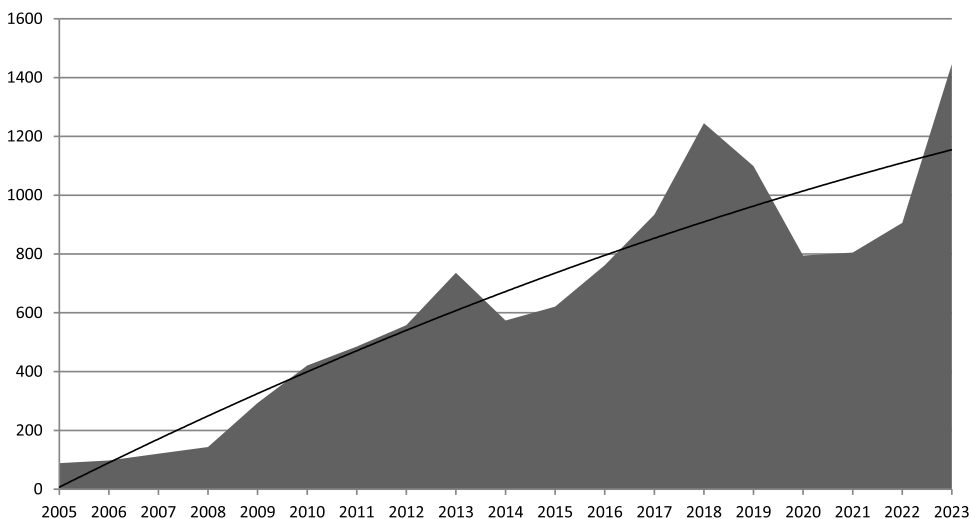
Figure 2. Temporal trend of news articles mentioning FOI published by the selected UK national newspapers between 2005 and 2023.

and the COVID pandemic struck. It climbed steeply again after 2022 and reached a new peak in 2023. This sharp increase was potentially boosted by the clearing of the COVID backlog of FOI requests. This trend reflects the impact of the pandemic, echoing recent reports about the challenges journalists faced in receiving FOI responses during that time, as discussed above.

In the dataset, *The Times* (3279 articles) had the biggest number of FOI stories, followed by *The Sun* (1741 articles) and *The Telegraph* (1623 articles), with *The Metro* (304 articles) having the smallest number (see Figure 3). Notably, *The Times* had the largest volume of FOI coverage throughout all these years (see Figure 4). *The Sun*'s FOI coverage was much
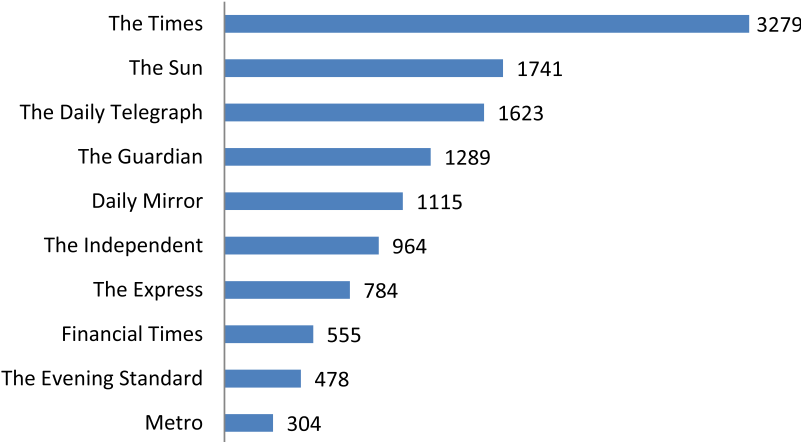
Figure 3. Coverage per newspaper.
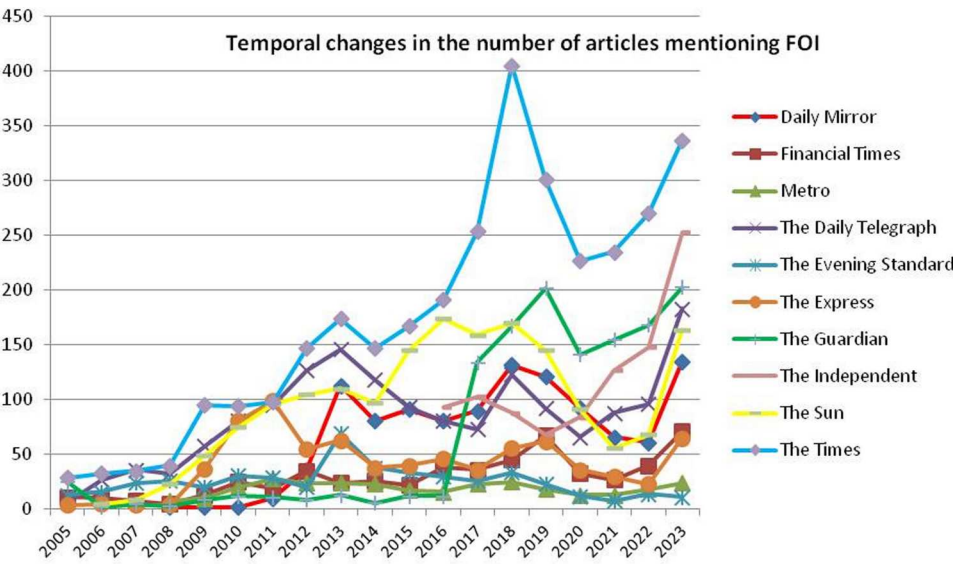
Figure 4. Temporal trend of news articles mentioning FOI published by each UK national newspaper between 2005 and 2023.

smaller after 2018, though it increased in 2023. With coverage growing after 2016, *The Guardian* became the only newspaper gaining a rapid increase in FOI coverage between 2018 and 2019. From 2020, *The Independent*'s coverage increased so quickly that the newspaper became number two in 2023. After 2020, *The Times*, *The Guardian*, and *The Independent* (replacing *The Sun*) published the largest volumes of FOI news.

Topics of FOI News Suggested by the N-*gram Analysis*

The *n*-gram analysis suggests that the FOI coverage focused on government spending, politician misconduct, and crimes (Figures 5–8). Unigrams including "bill," "tax," "government," "cost," "pay," and "cash" that suggest public funds and government spending were among the top 20 unigrams (Figures 5–6). The word "bill" could mean legislation, but a careful look into the texts suggests that "bill" mostly meant expenses, as shown in "A £7 m taxpayer bill for Speaker's House refurbishment; Asbestos and damp add to A £7 m Speaker's House bill" (*The Daily Telegraph*, October 28, 2023) or "'Cover up' over bill for stopping illegal immigrants" (*The Express*, February 8, 2019). In addition, the top 20 bigrams ranked by frequency and those with high frequency (>10) ranked by Mutual Information (MI) scores[4] in the headlines of these news articles also suggest the focus of the topics was on public funds, public spending, crimes in particular sex abuse, police, immigration, social inequalities, the BBC, COVID-19 and (data) security (Figure 5; Table 2).

MPs' expenses, government spending and expenses and the BBC were in the spotlight from the very beginning (Figures 6–8). In 2005, the FOI news focused on the expenses of politicians and those of the BBC. David McLetchie, the then Scottish Conservative Party

Figure 5. Heatmap of the use of top 20 unigram changing across newspapers.

Top 20 Unigram Frequencies Changing Over Time

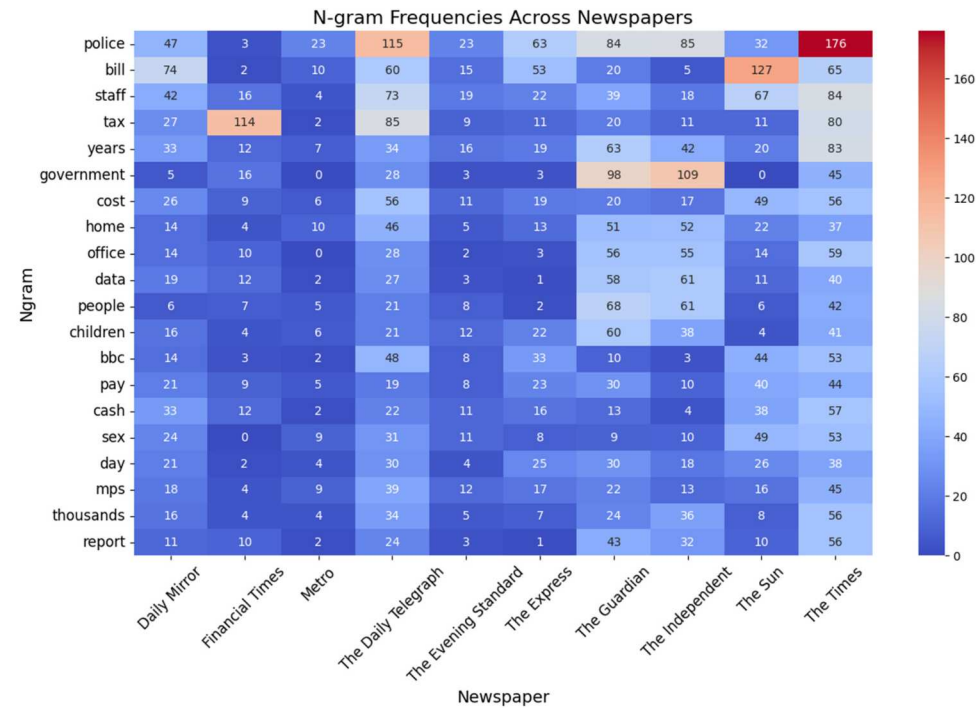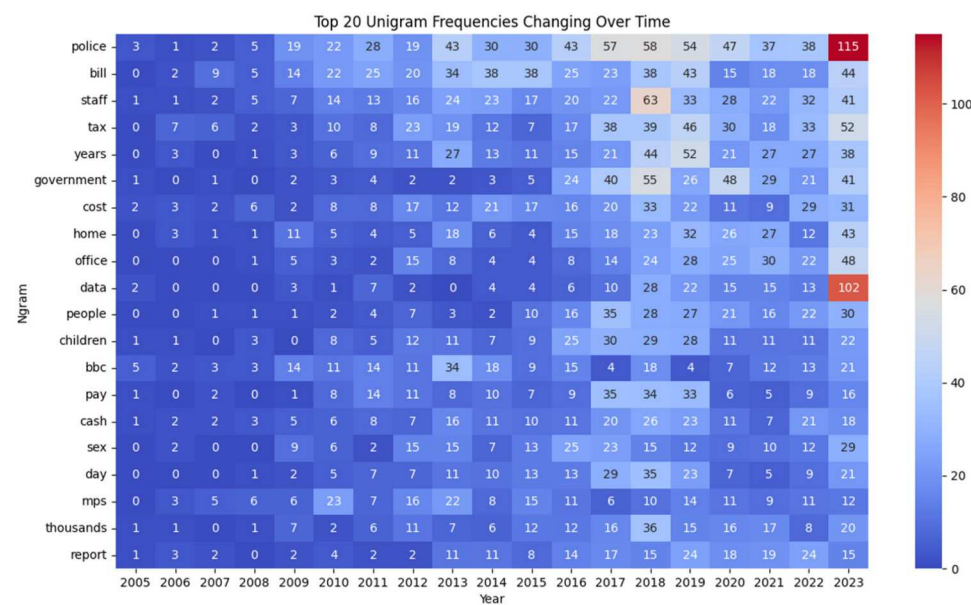| Ngram	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
police	3	1	2	5	19	22	28	19	43	30	30	43	57	58	54	47	37	38	115
bill	0	2	9	5	14	22	25	20	34	38	38	25	23	38	43	15	18	18	44
staff	1	1	2	5	7	14	13	16	24	23	17	20	22	63	33	28	22	32	41
tax	0	7	6	2	3	10	8	23	19	12	7	17	38	39	46	30	18	33	52
years	0	3	0	1	3	6	9	11	27	13	11	15	21	44	52	21	27	27	38
government	1	0	1	0	2	3	4	2	2	3	5	24	40	55	26	48	29	21	41
cost	2	3	2	6	2	8	8	17	12	21	17	16	20	33	22	11	9	29	31
home	0	3	1	1	11	5	4	5	18	6	4	15	18	23	32	26	27	12	43
office	0	0	0	1	5	3	2	15	8	4	4	8	14	24	28	25	30	22	48
data	2	0	0	0	3	1	7	2	0	4	4	6	10	28	22	15	15	13	102
people	0	0	1	1	1	2	4	7	3	2	10	16	35	28	27	21	16	22	30
children	1	1	0	3	0	8	5	12	11	7	9	25	30	29	28	11	11	11	22
bbc	5	2	3	3	14	11	14	11	34	18	9	15	4	18	4	7	12	13	21
pay	1	0	2	0	1	8	14	11	8	10	7	9	35	34	33	6	5	9	16
cash	1	2	2	3	5	6	8	7	16	11	10	11	20	26	23	11	7	21	18
sex	0	2	0	0	9	6	2	15	15	7	13	25	23	15	12	9	10	12	29
day	0	0	0	1	2	5	7	7	11	10	13	13	29	35	23	7	5	9	21
mps	0	3	5	6	6	23	7	16	22	8	15	11	6	10	14	11	9	11	12
thousands	1	1	0	1	7	2	6	11	7	6	12	12	16	36	15	16	17	8	20
report	1	3	2	0	2	4	2	2	11	11	8	14	17	15	24	18	19	24	15

Figure 6. Heatmap of the use of top 20 unigram changing over time.

Top 20 Bigram Frequencies Changing Over Time

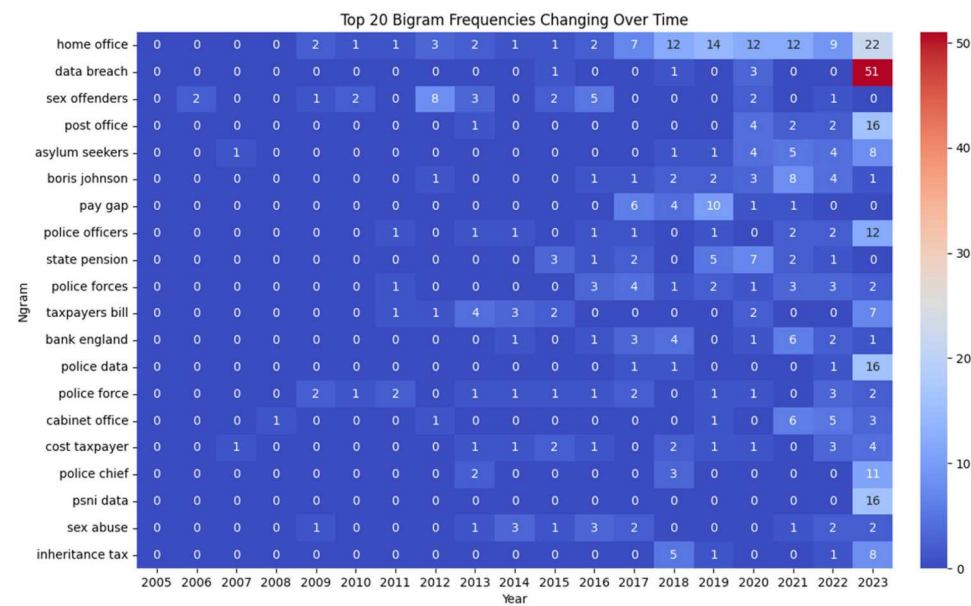| Ngram	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
home office	0	0	0	0	2	1	1	3	2	1	1	2	7	12	14	12	12	9	22
data breach	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	3	0	0	51
sex offenders	0	2	0	0	1	2	0	8	3	0	2	5	0	0	0	2	0	1	0
post office	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	4	2	2	16
asylum seekers	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	4	5	4	8
boris johnson	0	0	0	0	0	0	0	1	0	0	0	1	1	2	2	3	8	4	1
pay gap	0	0	0	0	0	0	0	0	0	0	0	0	6	4	10	1	1	0	0
police officers	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	0	2	2	12
state pension	0	0	0	0	0	0	0	0	0	0	3	1	2	0	5	7	2	1	0
police forces	0	0	0	0	0	0	1	0	0	0	0	3	4	1	2	1	3	3	2
taxpayers bill	0	0	0	0	0	0	1	1	4	3	2	0	0	0	0	2	0	0	7
bank england	0	0	0	0	0	0	0	0	0	1	0	1	3	4	0	1	6	2	1
police data	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	16
police force	0	0	0	0	2	1	2	0	1	1	1	1	2	0	1	1	0	3	2
cabinet office	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	6	5	3
cost taxpayer	0	0	1	0	0	0	0	0	1	1	2	1	0	2	1	1	0	3	4
police chief	0	0	0	0	0	0	0	0	2	0	0	0	0	3	0	0	0	0	11
psni data	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
sex abuse	0	0	0	0	1	0	0	0	1	3	1	3	2	0	0	0	1	2	2
inheritance tax	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0	1	8

Figure 7. Heatmap of the use of top 20 bigram changing over time.

Leader, resigned following the revelation of his substantial spending of taxpayers' money on taxi fares. In 2006, expenses and taxpayers' money were still in the focus, but new areas emerged in relation to Tony Blair and war, as well as the Olympics. The coverage about Blair was still about his expenses, while the UK government was also criticised for drafting

Figure 8. Word clouds of the top 10 unigrams in the headlines for each year.[9]
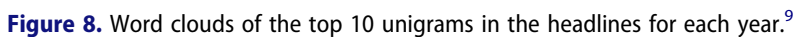
Table 2. Top 20 bigrams ranked by frequency and top 20 bigrams with high frequency (>10) and high-MI sores ranked by MI scores.

Top 20 bigrams ranked by frequency			Top 20 bigrams with high frequency (>10) and high-MI scores ranked by MI scores		
Bigram	Freq.	MI Score	Bigram	Freq.	MI score
home office	101	6.62	town halls	11	10.60
data breach	56	7.58	asylum seekers	24	10.23
sex offenders	26	7.29	city diary	13	9.77
post office	25	7.09	boris johnson	23	9.69
asylum seekers	24	10.23	speed cameras	12	9.67
boris johnson	23	9.69	coronavirus coverage	12	9.08
police officers	22	3.87	england wales	31	8.70
pay gap	22	8.06	sex pests	11	8.20
state pension	21	6.31	tens thousands	13	8.16
taxpayers bill	20	4.57	pay gap	22	8.09
police forces	20	4.45	tv licence	12	7.97
bank england	19	6.28	data breach	55	7.58
police data	19	3.05	sex offenders	26	7.29
police force	18	4.88	post office	25	7.09
cost taxpayer	17	4.73	gender pay	11	7.01
cabinet office	17	6.69	inheritance tax	15	6.74
psni data	16	6.38	cabinet office	19	6.69
sex abuse	16	5.29	home office	101	6.62
police chief	16	4.00	psni data	16	6.38
inheritance tax	15	6.74	state pension	21	6.31

Note: Freq. is short for frequency.

the Iraq weapons dossier by a spin doctor other than a weapons expert. FOI requests were also used to question the plan for the 2012 Olympics in London. From the Speaker's wife in 2007, through the MPs' expenses scandal in 2009, to the government's Rwanda plan bill

and Prince Harry's court costs and security legal bill in 2023, expenses and public spending remained a consistent focus throughout these years.

Newspapers displayed varied interests in these topics, without clear differences between right-wing and left-wing, or between broadsheets and tabloids. *The Sun* gave the greatest attention to "bill" (127 occurrences), followed by *The Times* (65 occurrences), *The Daily Telegraph* (60 occurrences) and *The Express* (53 occurrences). Given *The Express* only had 784 articles in the sample, its attention to "bill" was rather intensive. It is not a surprise to see that *The Financial Times* gave "tax" an extraordinary attention (114 occurrences). *The Daily Telegraph* (85 occurrences) and *The Times* (80 occurrences) also devoted considerable attention to "tax". Among all the newspapers, *The Independent* (109 occurrences) and *The Guardian* (98 occurrences) devoted the most significant attention to "government," while *The Sun* and *The Metro* even did not mention "government" once in their headlines. *The Daily Telegraph*, *The Times* and *The Sun* gave "bill," "cost," "bbc," and "sex" the most extensive attention, among all other newspapers. In fact, throughout these years, a major focus of the FOI news was on the BBC over its spending, pay, sexual harassment, inequality, diversity and (mis)management.

"Police" was a prominent, persistent focus for most of the newspapers from the start. FOI news revealed the lies, misconduct, and problems in the work and spending of the police, as well as their incompetence in dealing with crimes and gangs (Figures 5–8). Particularly, after 2009, "police" was the top 1 unigram in the headlines for most of the time with an extraordinary frequency in 2023 (Figure 6). This may have contributed to the ill relationship between the press and police in the UK (Colbran 2022). "Police" was the top one unigram in the headlines of *The Times*, *The Daily Telegraph*, *The Express*, *Metro*, and *The Evening Standard* (Figure 5). *The Metro* had the smallest coverage, but with 23 out of its 304 articles having "police" in their headlines, it still showed a strong interest in this topic.

The lack of distinction between types of newspapers also showed in other topics. Crimes, particularly those related to sex, started to become prominent from 2009. "Sex" started to consistently appear among the top 10 unigrams from 2009 (Figures 6 and 8). In 2009, FOI news including "sex" in their headlines was about child sex abuse and sex offenders such as teachers and paedophiles, published by *The Times*, *The Daily Telegraph*, *The Sun*, *The Evening Standard*, and *The Express* (Figure 5). However, the FOI requests led to the revelations might not be made by these newspapers' own journalists but by other organisations, such as the NSPCC (The National Society for the Prevention of Cruelty to Children, a UK children's charity), and other news media, such as the BBC and More4News. The top 20 bigrams associated with "child/ren" focused on big numbers of children, childcare, and child crime or abuse - related to sex. Among bigrams containing "child" or "children" with high frequency (>10) and high MI scores are "child sex" (MI = 5.58, freq = 11) and "children care" (MI = 4.46, freq = 11). Two typical examples for a headline containing "children" are "'Creaking' mental health care keeps children waiting years" (*The Times*, March 30, 2015) and "'He was covered in bruises': the vulnerable children being harmed in special schools" (*The Guardian*, July 6, 2019).

Naturally, COVID was prominent between 2020 and 2022 and the PSNI data breach was prominent in 2023. From 2021, immigration, indicated by bigrams such as "home office"[5] and "asylum seekers" were given much focus.

Topics of the FOI Coverage Revealed by LDA Topic Modelling

The findings of the LDA topic modelling match those from the *n*-gram analysis. They reveal a wide range of more detailed topics of the public interest from the impact of Brexit on governance and the market to crimes and public health issues. Overall, those topics show a clear focus on government actions and (mis)conduct, political transparency and misconduct, council spending, MPs' expenses, tax and social benefits, crimes, NHS struggles, and other social issues such as education, diversity, energy (costs of living) and public transportation (see Figure 9). "Government and markets" is the most dominant topic, followed by "NHS problems" and "public funds and council spending issues". Social issues, such as road safety, urban safety, children and young people addiction and mental health issues, public transportation issues and social housing issues are also dominant topics. The topic "data security and leak" was boosted by the PSNI data leak in 2023.

"Government and markets" exhibited high probabilities throughout the entire period, particularly between 2006 and 2009, and again after 2016, reaching its peak after 2020 (Figure 10). As the global financial crisis began to unfold in 2007, it makes sense that newspapers gave much attention to the government and the market between 2006 and 2009 (Figure 10). The topic naturally became prominent after the Brexit vote occurred in 2016, as Brexit has had impacts on the government and markets. This suggests that the use of FOI requests for journalism aligns with social dynamics and societal changes.

"NHS problems" was the second most dominant topic over the years, particularly between 2013 and 2018 – during the lead-up to the Brexit vote and the Brexit deal - and across these newspapers (Figures 10 and 11). The stories reported on NHS waiting times in A&E, ambulance services, hospital beds, and backlogs, highlighting the systemic problems within the NHS.

"Public funds and council spending issues" was the third most dominant topic, particularly among tabloids (Figure 11). Its probabilities slightly declined in recent years since 2020, perhaps due to the difficulties in getting useful responses from councils to journalists' FOI requests (Figure 10). The fourth most dominant topic was "police and crimes,"
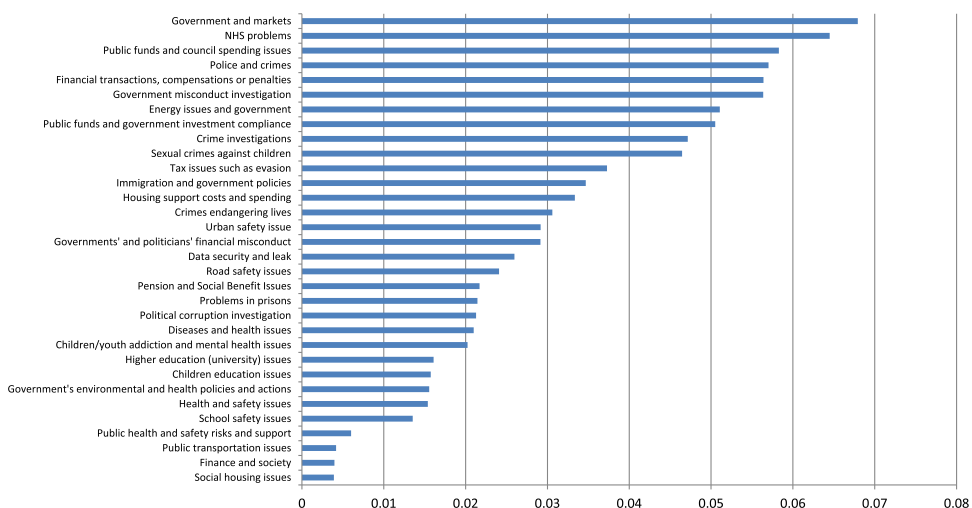
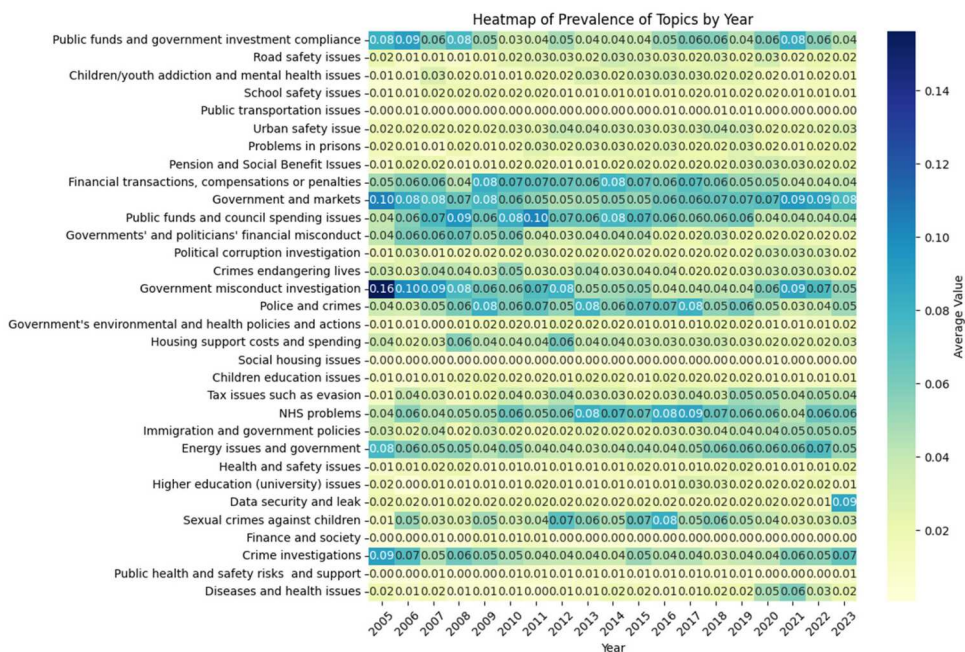**Figure 9.** Topic prevalence based on per-document topic distribution probabilities.

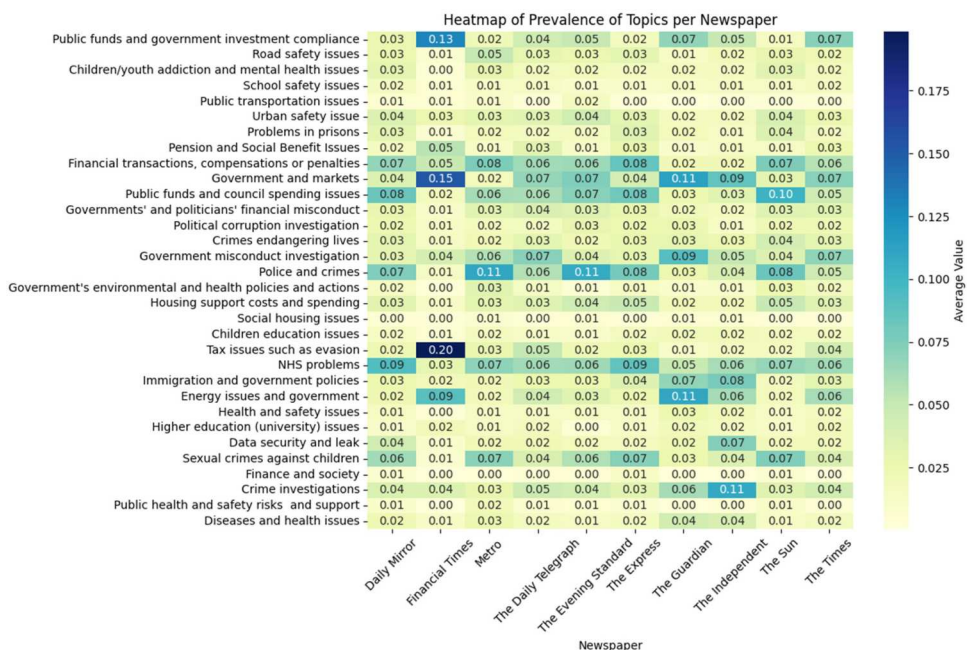**Figure 10.** Heatmap of topic prevalence over time.

**Figure 11.** Heatmap of topic prevalence per newspaper.

which was prominent all the years, with a slight decline in recent years since 2020, for which one possible reason is the difficult relationship between the police and journalism (Figure 10).

"Governments' and politicians' financial misconduct" had a higher probability in *The Daily Telegraph*, a newspaper that played a crucial role in exposing the MPs' Expenses scandal (Figure 11). It was particularly prominent between 2006 and 2010 (Figure 10). In 2005, when the FOI Act came into force, requests were filed to obtain details of MPs' expenses by campaigner and journalists[6] and there have been tugs of war since. Some data was ordered to be published in 2008. In 2008, *The Daily Telegraph* was the first news media that published stories based on leaked data about claims submitted by MPs. Later on, other newspapers such as *The Guardian* also extensively reported this story.

However, other aspects of UK society were largely absent in the coverage. For example, with a focus on the NHS and how it works, i.e., system flaws and problems, the FOI news had insufficient coverage on diseases and health issues. "Disease and health issues" was present as one of the dominant topics with much less prevalence than "government and markets" and "NHS problems" (Figure 9). Government actions and decisions, economics, social justice, public infrastructure, as well as civil rights were also missing in the coverage. While exposing the misconduct and scandals of governments and politicians, crimes, and revealing the problems with public services are important for our democracy, so is scrutinising internal government communications, decision-making, and legislation. These issues that are central to the government system's healthy operation were not represented in the FOI news analysed. Other absent areas included citizenship, military and defence, and environmental issues. The absence of environmental issues might be due to the Environmental Information Regulations (EIR), under which journalists request environmental information. Related news stories were potentially excluded in the sample.

A closer analysis of topic distribution over time (Figure 10) revealed that a significant area of using FOI requests across all these years was government conduct, transparency and accountability, as well as issues regarding crimes, the police and the NHS. Despite the constant focus, "public funds and council spending issues" ($r^7 = -0.50$, $p < 0.05$), "governments' and politicians' financial misconduct" ($r = -0.86$; $p < 0.001$), "government misconduct investigation" ($r = -0.56$; $p < 0.05$) had a statistically significant negative association with year. This means, when it reached 2023, there were fewer FOI stories on these topics. However, "pension and social benefit issues" ($r = 0.72$; $p < 0.001$), "tax issues such as evasion" ($r = 0.68$; $p = 0.001$), "immigration and government policies" ($r = 0.63$; $p < 0.05$), "diseases and health issues" ($r = 0.62$; $p < 0.05$), and "social housing issues" ($r = 0.52$; $p < 0.05$), increased coverage over the years. These suggested that the focus of FOI news gradually shifted from political and government misconduct to wider societal issues.

Different to the *n*-gram analysis, correlations were found between topics and newspapers. The topics were more closely related to the types of newspapers—broadsheets or tabloids—than to their political leanings. "Public funds and council spending issues" ($r = 0.83$; $p < 0.05$), "police and crimes" ($r = 0.88$; $p < 0.05$), "sexual crimes against children" ($r = 0.90$, $p < 0.001$), "financial transactions, compensations or penalties" ($r = 0.76$, $p < 0.05$), "housing support costs and spending" ($r = 0.73$, $p < 0.05$) and "urban safety issue" ($r = 0.75$, $p < 0.05$) were statistically significantly, positively associated with tabloids. "Public funds and government investment compliance" ($r = -0.71$, $p < 0.05$), "government and markets" ($r = -0.74$, $p < 0.05$), "energy issues and government" ($r = -0.82$, $p < 0.05$), and "higher education (university) issues" ($r = -0.86$, $p < 0.05$) were more likely to be published by broadsheets.
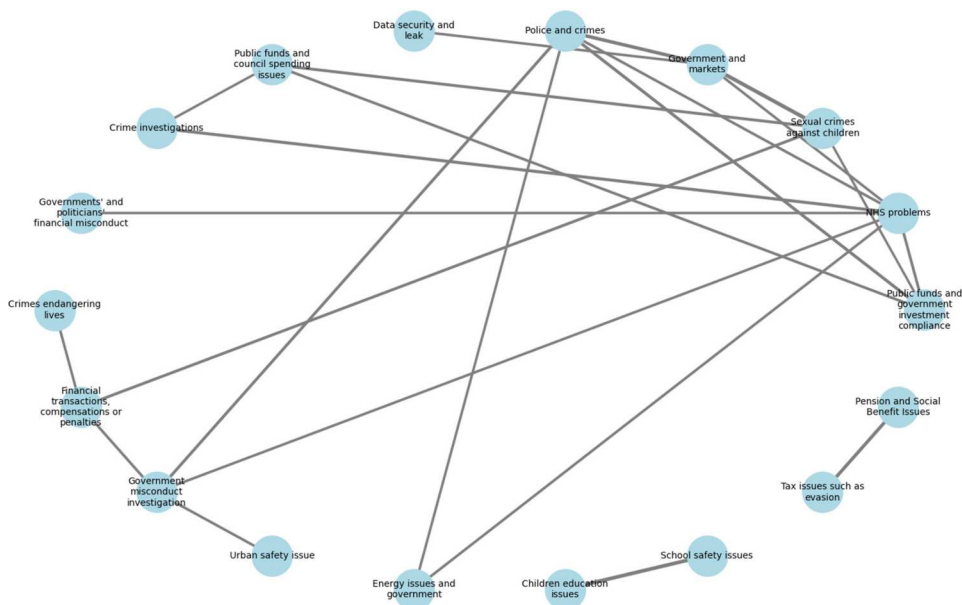
Figure 12. Topic correlation network (threshold = 0.09).

The topics of the broadsheets and the tabloids had both differences and overlaps. Topics: "police and crimes" and "sexual crimes against children" were particularly prominent across the tabloids, such as *Daily Mirror*, *Metro*, *The Evening Standard, The Express*, and *The Sun*. "Financial transactions, compensations or penalties," "public funds and council spending issues" and "NHS problems" were prominent in the tabloids, *The Daily Telegraph*, and *The Times*. "Public funds and government investment compliance," "tax issues," and "energy issues and government" were mostly prominent in the broadsheets. It makes sense that topics related to finance and the market including "public funds and government investment compliance," "government and markets," "tax issues such as evasion" and "energy issues and government" were particularly prominent in *Financial Times*. "Government and markets" was mostly prominent in the broadsheets, but was also featured in one tabloid: *The Evening Standards*.

The topic correlation network (for positive correlations, threshold r = 0.09, see Figure 12) indicates that public funds, government investment, spending and policies were often co-current with topics like the problems in the NHS, crimes, and safety issues. This connection indicates that the focus of FOI news was on the resources, effectiveness, and problems of public services, as well as crimes and safety including children safety. "Tax issues" often appeared together with "pension and social benefit issues", suggesting a theme centred around public welfare and taxation.

Discussion and Conclusion

Echoing previous studies (such as Worthy 2017), the above discussions suggest that the birth of the FOI Act has opened up political communication in the UK. The overall increase in the number of FOI news reflects the UK news media's embrace of the FOI Act, despite the

process for FOI requests has become increasingly restrictive (Žuffová 2023). However, the fluctuations reflect the changes in the social environment, such as the COVID-19 pandemic.

FOI news since 2005 has given persistent publicity to issues in the public interest such as government and markets, public funds and spending, politicians' (mis)conduct, police and crimes, public services' performance, and safety and security, holding power accountable. They are also addressing issues that the public cares about, such as pensions and taxes, crimes, and public services. Despite some similarities, such as a prominent focus on the police, these topics are, however, different from those of FOI coverage in other countries, such as Canada (Luscombe and Walby 2022). Exposing these issues in news coverage informs the public, sets agendas, keeps them at the centre of public attention, facilitates public debates and discussions, and improve government transparency and accountability. These FOI stories demonstrate their ability to compel political elites to address the issues they reveal, generate predictable responses, and hold those in power – governments, politicians, and other elites- accountable. In some cases, this accountability has even led to the removal of politicians, including MPs (Worthy 2017). Through FOI news coverage, UK newspapers has acted as watchdogs, served as a check on power, and helped prevent political corruption and unethical activities. In this way, news media has at least indirectly helped reform the political system, as shown in the establishment of the Independent Parliamentary Standards Authority (IPSA) in 2009 following the revelations of the MPs' expenses scandal. Such coverage may also be able to deter politicians and public officials from getting involved in corrupt and dishonest practices. By revealing social issues such as those problems facing the NHS, UK newspapers can raise public and government awareness of these issues, potentially driving changes in society. Their exposure of these problems may also put the UK government under pressure to address them effectively. In addition, promoting transparency and accountability, FOI news coverage can help build public trust in news media, stimulate public scrutiny of government actions and policies, encourage public engagement and awareness about key issues, prompt public debates, and impact public opinion on public issues, demanding greater accountability. This, in longer term, is beneficial to democracy.

However, because of this pressure on the government and politicians, FOI news may have contributed to their hostility towards news media, as discussed above. In addition, media exposure of political scandals and social problems in FOI news can also influence public opinion about politicians and government departments. This means that FOI news, which has been increasing in number, may undermine political legitimacy, resulting in a loss of public support that could impact local and general elections. Therefore, politicians and government departments, such as the police, may seek to avoid the constant publication of FOI news due to its potential implications for their public images. This potential mentality partially explains the increasing difficulties in UK journalists' access to data through FOI requests as discussed earlier in this paper.

The topics of FOI news are influenced by various factors. One is the access to information under the FOI Act. Different government departments and institutions appear to have different attitudes and responses to journalists' FOI requests. For some areas, journalists can smoothly obtain data through their FOI requests; but for others, their requests may be rejected for various reasons including cost limits, confidentiality and privacy (Tong 2025). This may be one reason to explain why some key areas have been insufficiently covered in FOI news. Another reason is the news values of news

media, which can explain the varying and sometimes overlapping foci of these newspapers. The types of topics in FOI news indicate not only what they are able to report using FOI requests but also what they are interested. The lack of clear distinction in FOI coverage between different types of newspapers and the prominence of sensational topics, such as sex, sexual assaults, especially sexual crimes related to children, particularly among tabloids and some broadsheets such as *The Daily Telegraph* and *The Times*, reflect sensationalism and tabloidation in FOI news, and a blurred boundary between broadsheets and tabloids. A balance needs to be struck between serving the public interest and being sensational and tabloidised.

This study contributes to knowledge of the influence of the FOI Act on political communication by revealing the temporal changes of topics and the distributions of topics across broadsheets and tabloids, and across left-wing and right-wing newspapers since the implementation of the Act. It offers a new lens through which we can gain insights into the relationship between news media and governance. However, it is limited by its focus on UK national newspapers. For future research, it would be interesting to study local newspapers or broadcasting media in the UK to explore the ways in which local journalists and broadcast journalists use the FOI Act and the focus of their news stories. Additionally, the computational approaches used in this study can also be applied to analysing FOI coverage in other democracies such as the United States to understand how their coverage is similar to or different from that in the UK.

Notes

1. https://ico.org.uk/for-organisations/foi/what-is-the-foi-act-and-are-we-covered/
2. *The Sunday Mirror* and *The Independent on Sunday* were not included in the Nexis database at the time of collection.
3. These words include 'freedom', 'information', 'request', 'requests', 'foi', 'foia','act', "exclusive", "timeupdated", and "timeupdate".
4. Mutual information (MI) scores indicate the dependency between two variables.
5. "Home office" may refer to an office located at home. But a closer look into the corpus confirms this term is mostly related to immigration.
6. https://www.telegraph.co.uk/news/newstopics/mps-expenses/6499657/MPs-expenses-scandal-a-timeline.html
7. The *r* (correlation coefficient) and *p* (a statistical measure used to determine whether the result is statistically significant) values represent the results of the correlation analysis between topics and years or between topics and types of newspapers.
8. The actual research process is iterative, not linear.
9. For 2005 and 2006, the word "times" ranked 10th was replaced with the word ranked 11th.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Data availability statement

Key Python scripts, numerical/graph results, and document feature matrices generated from the analysis are shared on the author's GitHub account.

References

Ahmad, J., C. V. Gever, A. Savaşan, and N. A. Tukur. 2024. "Knowledge and Utilization of Freedom of Information law among Journalists in Nigeria." *Information Development* 40 (2): 247–265.

Antai, S. J., and P. E. Umoren. 2023. "Freedom of Information Act (FoIA) and the Challenges of Investigative Journalism in Nigeria." *AKSU Journal of Management Sciences* 8: 117–128.

Appelgren, E., and R. Salaverría. 2018. "The Promise of the Transparency Culture: A Comparative Study of Access to Public Data in Spanish and Swedish Newsrooms." *Journalism Practice* 12 (8): 986–996.

Balakrishnan, V., and E. Lloyd-Yemoh. 2014. "Stemming and Lemmatization: A Comparison of Retrieval Performances." *Lecture Notes on Software Engineering* 2: 262–267. https://web.archive.org/web/20170809060210id_/http://www.lnse.org/papers/134-I3007.pdf.

Birkinshaw, P. 2010. "Freedom of Information and Its Impact in the United Kingdom." *Government Information Quarterly* 27 (4): 312–321.

Blair, T. 2010. *A Journey*. London: Hutchinton.

Blei, D. M., and J. D. Lafferty. 2007. "A Correlated Topic Model of Science." *The Annals of Applied Statistics* 1 (1): 17–35.

Blei, D., A. Ng, and M. Jordan. 2001. "Latent Dirichlet Allocation." Paper Presented at the Advances in Neural Information Processing Systems.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=https://githubhelp.com.

Brooke, H. 2006. "Freedom of Information and Local Journalism." In *Local Journalism and Local Media*, edited by B. Franklin, 290–300. London: Routledge.

Broomfield, M. 2024. "UK Journalists Fall Victim to New Police Tactics." *Index on Censorship*. https://www.indexoncensorship.org/2024/08/uk-journalists-fall-victim-to-new-police-tactics/.

Burgess, M. 2015a. "Freedom of Information Act: 103 Stories That Prove Chris Grayling Wrong." *The Guardian*. https://www.theguardian.com/media/2015/oct/30/freedom-of-information-act-chris-grayling-misuse-foi.

Burgess, M. 2015b. *Freedom of Information: A Practical Guide for UK Journalists*. London: Routledge.

Camaj, L. 2016. "Governments' Uses and Misuses of Freedom of Information Laws in Emerging European Democracies: Foi laws' Impact on News Agenda-Building in Albania, Kosovo, and Montenegro." *Journalism & Mass Communication Quarterly* 93 (4): 923–945.

Cavnar, W. B., and J. M. Trenkle. 1994. "N-gram-based Text Categorization." Paper Presented at the Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval.

Chae, B., and E. Park. 2018. "Corporate Social Responsibility (CSR): A Survey of Topics and Trends Using Twitter Data and Topic Modeling." *Sustainability* 10 (7): 2231.

Clifton-Sprigg, J., J. James, and S. Vujić. 2020. "Freedom of Information (FOI) as a Data Collection Tool for Social Scientists." *PLoS One* 15 (2): e0228392.

Colbran, M. P. 2022. *The Relationship between Police and Media: Is the MET Less Accountable than Ever Before?* https://eprints.lse.ac.uk/116150/1/politicsandpolicy_police_media.pdf.

DiMaggio, P., M. Nag, and D. M. Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U. S. Government Arts Funding." *Poetics* 41 (6): 570–606. https://doi.org/10.1016/j.poetic.2013.08.004.

Dittmer, L. 1973. "The Structural Evolution of 'Criticism and Self-Criticism'." *The China Quarterly* 56: 708–729.

Evans, R. 2020, November 24. "'Orwellian' Government Unit Obstructs Freedom of Information, Says Report." *The Guardian*. https://www.theguardian.com/politics/2020/nov/24/orwellian-government-unit-obstructs-freedom-of-information-says-report.

Field, A., D. Kliger, S. Wintner, J. Pan, D. Jurafsky, and Y. Tsvetkov. 2018. "Framing and Agenda-Setting in Russian News: A Computational Analysis of Intricate Political Strategies." arXiv: 1808.09386. https://arxiv.org/pdf/1808.09386.

Fink, K. 2018. "State FOI Laws: More Journalist-Friendly, or Less?" In *Troubling Transparency: The History and Future of Freedom of Information*, edited by D. E. Pozen, and M. Schudson, 91–115. Columbia: Columbia University Press.

Fowler, A. J., R. A. Agha, C. F. Camm, and P. Littlejohns. 2013. "The UK Freedom of Information Act (2000) in Healthcare Research: A Systematic Review." *BMJ Open* 3 (11): e002967.

Galetta, D.-U. 2016. "The Italian Freedom of Information Act 2016 (Why Transparency-on-Request Is a Better Solution)." *Italian Journal of Public Law* 8 (2): 268–290. https://ssrn.com/abstract = 3014629.

García Avilés, J. A., B. León, K. Sanders, and J. Harrison. 2006. "Journalists at Digital Television Newsrooms in Britain and Spain: Workflow and Multiskilling in a Competitive Environment." *Jouranlism Studies* 5 (1): 87–100.

Gayle, D. 2022. "Senior Officers Ordered 'Unlawful' Arrests of Journalists at Just Stop Oil Protests." *The Guardian*. https://www.theguardian.com/media/2022/nov/23/senior-officers-ordered-unlawful-arrests-of-journalists-at-just-stop-oil-protests.

Greenwood, G. 2021. Written Evidence from George Greenwood (FOI 06). https://committees.parliament.uk/writtenevidence/38160/pdf/.

Harcup, T., and D. O'Neill. 2017. "What Is News?" *Journalism Studies* 18 (12): 1470–1488. https://doi.org/10.1080/1461670X.2016.1150193.

Hazell, R., B. Worthy, and M. a. Glover. 2010. *The Impact of the Freedom of Information Act on Central Government in the UK*. London: Palgrave.

Jacobi, C., W. v. Atteveldt, and K. Welbers. 2015. "Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling." *Digital Journalism*. https://doi.org/10.1080/21670811.2015.1093271.

James, H. 2018. "FOIA and Investigative Reporting: Who's Asking What, Where, and When and Why It Matters." In *Troubling Transparency: The History and Future of Freedom of Information*, edited by P. David and S. Michael, 116–134. New York: Columbia University Press.

Javadinejad, A. 2024. "A Longstanding Duality: Discursive Construction of the EU vs the UK in the British Broadsheets' News Discourse of the Brexit Referendum." *International Journal of English Studies* 24 (1): 49–75.

Jelodar, H., Y. Wang, R. Orji, and S. Huang. 2020. "Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: Nlp using LSTM Recurrent Neural Network Approach." *IEEE Journal of Biomedical and Health Informatics* 24(10): 2733-2742.

Jelodar, H., Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao. 2019. "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey." *Multimedia Tools and Applications* 78: 15169–15211.

Kingston, S., A. Elliott, and T. Thomas. 2019. "'Cost'calculations as a Barrier to Gaining Information under the Freedom of Information Act 2000 from the Police in England and Wales." *Policing and Society: An International Journal of Research and Polic* 29 (7): 834–847. https://doi.org/10.1080/10439463.2018.1424156.

Knight, M. 2015. "Data Journalism in the UK: A Preliminary Analysis of Form and Content." *Journal of Media Practice* 16 (1): 55–72.

Kwoka, M. B. 2018. "First-Person FOIA." *Yale Law Journal* 127 (8): 2204–2269.

Lefkowitz, J. 2018. "'Tabloidization' or Dual-Convergence: Quoted Speech in Tabloid and 'Quality' British Newspapers 1970–2010." *Journalism Studies* 19 (3): 353–375.

Lemov, M. R., and N. Jones. 2018. "John Moss and the Roots of the Freedom of Information Act: Worldwide Implications." *Southwestern Journal of International* 24 (1): 1–40.

Levy, K. E. C., and M. Franklin. 2014. "Driving Regulation: Using Topic Models to Examine Political Contention in the U.S. Trucking Industry." *Social Science Computer Review* 32 (2): 182–194. https://doi.org/10.1177/0894439313506847.

Lidberg, J., and D. Muller. 2018. *In the Name of Security - Secrecy, Surveillance and Journalism*. London: Anthem Press.

Liu, Q., Z. Zheng, J. Zheng, Q. Chen, G. Liu, S. Chen, B. Chu, et al. 2020. "Health Communication through News Media during the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach." *Journal of Medical Internet Research* 22 (4): e19118.

Lu, J., and G. F. Nemet. 2020. "Evidence Map: Topics, Trends, and Policy in the Energy Transitions Literature." *Environmental Research Letter* 15 (12): 123003.

Luscombe, A., and K. Walby. 2022. "Deciphering the Decline: A Computational Analysis of Two Decades of Canadian Newspaper Op-Eds on Freedom of Information." *Canadian Journal of Communication* 47 (2): 313-335.

Maier, D., A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, et al. 2021. "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology." In *Computational Methods for Communication Science*, edited by W. v. Atteveldt, and T.-Q. Peng, 13–38. London: Routledge.

Margaret, K. 2018. "The Other FOIA Requesters." In *Troubling Transparency: The History and Future of Freedom of Information*, edited by D. Prozen and S. Michael, 73–90. New York: Columbia University Press.

Markov, A. A. n. 2006 (original 1913). "An Example of a Statistical Investigation of the Text of 'Eugene Onegin' Illustrating the Dependence between Samples in Chain (Translated into English)." *Science in Context* 19 (4): 591–600. https://www.alpha60.de/research/markov/DavidLink_AnExampleOf Statistical_MarkovTrans_2007.pdf

Mason, R., and N. Watt. 2015. "Freedom of Information Act Misused by Media to Create Stories, says Grayling." *The Guardian*. https://www.theguardian.com/media/2015/oct/29/freedom-of-information-journalists-chris-grayling-foi.

Mazzi, D. 2018. "'By Partially Renouncing Their Sovereignty…': On the Discourse Function(s) of Lexical Bundles in EU-Related Irish Judicial Discourse." In *Phraseology in Legal and Institutional Settings: A Corpus-Based Interdisciplinary Perspective*, edited by S. Goźdź-Roszkowski and G. Pontrandolfo, 189-202. New York: Routledge.

Mellado, C., D. Hallin, L. Carcamo, R. Alfaro, D. Jackson, M. L. Humanes, M. Márquez-Ramírez, et al. 2021. "Sourcing Pandemic News: A Cross-National Computational Analysis of Mainstream Media Coverage of COVID-19 on Facebook, Twitter, and Instagram." *Digital Journalism* 9 (9): 1261–1285.

Michener, G. 2011. "FOI Laws around the World." *Journal of Democracy* 22 (2): 145–159.

Mihailidis, P., and A. Gamwell. 2022. "Designing Engagement in Local News: Using FOIA Requests to Create Inclusive Participatory Journalism Practices." *Journalism Practice* 16 (5): 828–847.

Ministry of Justice. 2011. "Memorandum to the Justice Select Committee." https://assets.publishing.service.gov.uk/media/5a7c12a840f0b63f7572b297/post-legislative-assessment-of-the-foi-act.pdf.

Morrison, H. 2023. "Coronation Protest Sees Journalist Arrested and Press Pass Seized." *The National*. https://www.thenational.scot/news/23508249.coronation-protest-sees-journalist-arrested-press-pass-seized/.

Nastiti, K. R., A. F. Hidayatullah, and A. R. Pratama. 2021. "Discovering Computer Science Research Topic Trends Using Latent Dirichlet Allocation." *JOIN (Jurnal Online Informatika* 6: 17–24.

Ojo, O. E., A. Gelbukh, H. Calvo, and O. O. Adebanji. 2021. "Performance Study of N-Grams in the Analysis of Sentiments." *Journal of the Nigerian Society of Physical Sciences* 3 (4): 477–483.

Pearce, H. 2020. "The (UK) Freedom of Information Act's Disclosure Process Is Broken: Where Do We go from Here?" *Information & Communications Technology Law* 29 (3): 354–390. https://doi.org/10.1080/13600834.2020.1785663.

Ponsford, D. 2016. "Police Media Guidelines Set Out New Wide-ranging Controls on Contact Between Officers and Journalists." *PressGazette*. https://pressgazette.co.uk/news/new-police-media-guidelines-iset-out-new-wide-ranging-controls-contact-between-officers-and/

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54 (1): 209–228.

Rahimian, M., J. L. Warner, S. K. Jain, R. B. Davis, J. A. Zerillo, and R. M. Joyce. 2019. "Significant and Distinctive n-Grams in Oncology Notes: A Text-Mining Method to Analyze the Effect of OpenNotes on Clinical Documentation." *JCO Clinical Cancer Informatics* 3: 1–9. https://ascopubs.org/doi/full/10.1200CCI.19.00012

Reinemann, C., J. Stanyer, S. Scherr, and G. Legnante. 2012. "Hard and Soft News: A Review of Concepts, Operationalizations and key Findings." *Journalism* 13 (2): 221–239.

Relly, J. E., and C. B. Schwalbe. 2013. "Watchdog Journalism: India's Three Largest English-Language Newspapers and the Right to Information Act." *Asian Journal of Communication* 23: 284–301. https://doi.org/10.1080/01292986.2012.729149.

Ricketson, M., and R. Snell. 2002. "FOI: Threatened by Governments, Underused by Journalists—Still a Sharp Tool." In *Journalism—Investigation and Research*, edited by S. Tanners, 150–164. Sydney, Australia: Longman.

Riddell, P. 2013. "Impact of Transparency on Accountability." In *Transparency in Politics and the Media: Accountability and Open Government*, edited by N. Bowles, J. T. Hamilton, and D. A. L. Levy, 19–31. IB Tauris: London and New York.

Roberts, M. E., B. M. Stewart, and D. Tingley. 2019. "stm: R Package for Structural Topic Models." *Journal of Statistical Software* 91: 2. https://doi.org/10.18637/jss.v091.i02.

Rudniy, A., and N. Elliot. 2016. "Collaborative Review in Writing Analytics: N-Gram Analysis of Instructor and Student Comments." Paper Presented at the EDM (Workshops). https://www.norbertelliot.com/wp-content/uploads/2016/03/Rudiny_Elliot_NgramEdm2016.pdf.

Ryan, C. 2015. "Freedom of Information and the Media: A Case of Delay, Deny, Defeat?" In *Ireland and the Freedom of Information Act: Foi@ 15*, edited by R. Kitchin, 102–121. Manchester: Manchester University Press.

Shapira, R. 2018. "Law as Source: How the Legal System Facilitates Investigative Journalism." *YALE L. & POL'y REV* 37 (1): 153–226.

Silver, D. 2019. "The News Media and the FOIA." In *The US Freedom of Information Act at 50*, 63–84. London: Routledge.

Simpson, K., and N. Startin. 2023. "Tabloid Tales: How the British Tabloid Press Shaped the Brexit Vote." *JCMS: Journal of Common Market Studies* 61 (2): 302-322.

Sparks, C. 2000. "Introduction: The Panic over Tabloid News." In *Tabloid Tales: Global Debates over Media Standards*, edited by C. Sparks, and J. Tulloch, 1–40. Lanham, MD: Rowman & Littlefield Publishers. Inc.

Syed, S., and M. Spruit. 2017. "Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation." Paper Presented at the 2017 International Conference on Data Science and Advanced Analytics. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8259775&casa_token=VvypeTssjh0AAAAA:GSz8ui8QG7ls7r62W_nmDIAW8f1KfhXsMm5L6q5Sy2GKJtIeuctB0VW1aP2jlDoIdvxqHgLZ&tag=1.

Tong, J. 2022. "Rethinking the Expertise of Data Journalists: A Case Study." *International Journal of Communication* 16 (20): 2781–2800.

Tong, J. 2025. "Growing Government Secrecy: How and Why Information Access has Become Increasingly Restricted in the UK." *Journalism Practice*.

Turetsky, K. M., and T. A. Riddle. 2018. "Porous Chambers, Echoes of Valence and Stereotypes: A Network Analysis of Online News Coverage Interconnectedness Following a Nationally Polarizing Race-Related Event." *Social Psychological and Personality Science 9* (2), 163–175.

Umbricht, A., and F. Esser. 2016. "The Push to Popularize Politics, Understanding the Audience-Friendly Packaging of Political News in six Media Systems since the 1960s." *Journalism Studies* 17 (1): 100–121.

Wagner, A. J. 2022. "Popular Information: An Analysis of FOI use and Behavior." *Government Information Quarterly* 39 (2): p101677.

Wagner, A. J., and D. Cuillier. 2025. "Tale of two Requesters: How Public Records law Experiences Differ by Requester Types." *Journalism* 26 (2): 325–344.

Wald, P. M. 1984. "The Freedom of Information Act: A Short Case Study in the Perils and Paybacks of Legislating Democratic Values." *Emory Law Journal* 33 (3): 649–684.

Wang, W., Y. Feng, and W. Dai. 2018. "Topic Analysis of Online Reviews for two Competitive Products Using Latent Dirichlet Allocation." *Electronic Commerce Research and Applications* 29: 142–156.

Whiting, A., B. Campbell, and I. Awan. 2021. "The Prevent Duty in UK Higher Education: Insights from Freedom of Information Requests." *The British Journal of Politics and International Relations* 23 (3): 513–532. https://doi.org/10.1177/1369148120968.

Worthy, B. 2008. "The Future of Freedom of Information in the United Kingdom." *The Political Quarterly* 79: 100–108.

Worthy, B. 2013. "'Some Are More Open Than Others': Comparing the Impact of the Freedom of Information Act 2000 on Local and Central Government in the UK." *Journal of Comparative Policy Analysis: Research and Practice* 15 (5): 395–414. https://doi.org/10.1080/13876988.2013.836300.

Worthy, B. 2017. "Freedom of Information and the Media." In *The Routledge Companion to Media and Human Rights*, edited by H. Tumber, and S. Waisbord, 60–72. London: Routledge.

Worthy, B., and R. Hazell. 2013. "The Impact of the Freedom of Information Act in the UK." In *Transparency in Politics and the Media: Accountability and Open Government*, edited by N. Bowles, J. T. Hamilton, and D. Levy, 31–45. London: L.B. Tauris.

Worthy, B., and R. Hazell. 2017. "Disruptive, Dynamic and Democratic? Ten Years of FOI in the UK." *Parliamentary Affairs* 70: 22–42. https://doi.org/10.1093/pa/gsv069.

Wyeth, L. 2014. "A Cheap Shot or a Real Investigation?" *British Journalism Review* 25 (4): 49–55.

Ye, Z. 2017. *The Semantics of Nouns*. Oxford: Oxford University Press.

Žuffová, M. 2023. "Fit for Purpose? Exploring the Role of Freedom of Information Laws and Their Application for Watchdog Journalism." *The International Journal of Press/Politics* 28 (1): 300–322. https://doi.org/10.1177/19401612211006702.

Züll, C., and J. Landmann. 2004. "Computer-Assisted Content Analysis Without Dictionary?" Paper Presented at the Sixth International Conference on Logic and Rationality and Interaction (LORI-VI).