REVISTA DE CIENCIAS SOCIALES Y HUMANAS XXI
UNIVERSITAS

https://doi.org/10.17163/uni.n43.2025.03

Web coverage of corruption in Chile: automated analysis with artificial intelligence and linguistics

Cobertura web de la corrupción en Chile: análisis automatizado con inteligencia artificial y lingüística

Camila Rojas-Sánchez
camila.rojas@ufrontera.cl
https://orcid.org/0009-0009-3677-2268
Universidad de La Frontera, Temuco, Chile
https://ror.org/04v0snf24

Luis Cárcamo-Ulloa
lcarcamo@uach.cl
https://orcid.org/0000-0003-0633-9606
Universidad Austral de Chile, Valdivia, Chile
https://ror.org/029ycp228

Matthieu Vernier
mvernier@inf.uach.cl
https://orcid.org/0000-0002-3265-4709
Universidad Austral de Chile, Valdivia, Chile
https://ror.org/029ycp228

Received on: 12/03/2025 **Revised on:** 30/04/2025 **Approved on:** 15/06/2025 **Published on:** 01/09/2025

Suggested citation: Rojas-Sánchez, C., Cárcamo-Ulloa, L. and Vernier, M. (2025). Web coverage of corruption in Chile: automated analysis with artificial intelligence and linguistics. *Universitas XX1, 43,* pp. 63-91. https://doi.org/10.17163/uni.n43.2025.03

Abstract

This study seeks to present and validate analytical techniques for evaluating figures involved in political corruption cases using an automated news classifier that combines artificial intelligence, linguistics, and social and political sciences. To this end, a topic analysis using LDA (Latent Dirichlet allocation) is analyzed using automatic language processing strategies. A large corpus of information extracted from Sophia Search, an automated news classifier, contains 33,699 news items about seven corruption cases considered scandalous from 64 Chilean online news outlets. From this corpus, the analysis of 32 figures involved in seven events considered politically scandalous is obtained. These items were triangulated with 1120 concepts that helped to deepen the evaluation of the figures. Thus, the results reveal differences and similarities across one axis between media considered hegemonic and non-hegemonic, and across another axis between media considered traditional or emerging, highlighting distinct strategies in each media type. Finally, a conclusion is drawn from the results and future projections are made regarding the use of artificial intelligence tools in social and political sciences.

Keywords

Data mining, artificial intelligence, linguistics, social and political sciences, political corruption, news media, web.

Resumen

Este estudio busca presentar y validar técnicas de análisis de evaluación de personajes involucrados en casos de corrupción política mediante un clasificador de noticias automatizado que combina inteligencia artificial, lingüística y ciencias sociales y políticas. Para ello, se realiza, con base en estrategias de tratamiento automático del lenguaje, un análisis de tópicos con LDA (Latent Dirichlet allocation) de un corpus amplio de informaciones extraído de Sophia Search, un clasificador de noticias automatizado, y donde se encuentran 33 699 noticias sobre 07 casos de corrupción considerados escandalosos, provenientes de 64 medios informativos chilenos presentes en la web. Desde este corpus, resulta el análisis de 32 personajes involucrados entre siete hechos considerados como escandalosos políticamente, y que fueron triangulados con 1120 conceptos que ayudaron a ahondar en la evaluación de las y los personajes. Así, los resultados dejan en evidencia diferencias y similitudes en un cruce de un eje entre medios considerados hegemónicos y no hegemónicos, y de otro eje entre medios considerados tradicionales o emergentes, resaltando estrategias propias en cada uno de los tipos de medios. Finalmente se concluye sobre los resultados y se dan futuras proyecciones sobre el uso de herramientas con inteligencia artificial en ciencias sociales y políticas.

Palabras clave

Minería de datos, inteligencia artificial, lingüística, ciencias sociales y políticas, corrupción política, medios informativos, web.

Introduction and state-of-the-art

Since the return to democracy, Chile has been recognized as one of the countries with the highest levels of transparency, competing with Uruguay for first place in Latin America (Transparency International, 2024). However, this score has fallen significantly since 2014, after several high-impact corruption cases involving prominent political figures and institutions in the country came to light, revealing structural deficiencies in the prevention and punishment of corruption (Wilenman and Feddersen, 2023; Heiss, 2023; Matus, 2020; Nash, 2017; CIPER, 2015a, 2015b).

Although it cannot be denied that Chile has made progress in the last two decades in terms of regulations and institutions related to transparency and public probity, such advances have been partial and generally driven by corruption scandals (Nash, 2017; Zalaquett, 2011), which has led to low confidence in public and private institutions (Irarrazabal and Cruz, 2023).

This decline in trust over 20 years is not random, as corruption scandals and the privileges enjoyed by certain figures in Chilean politics are widely reported in the media.

In Thompson's terms (2001), cases of corruption are presented as political scandals and promote the mobilization of moral outrage and public denunciation (Pereyra, 2012). In this sense, corruption cannot be understood without taking into account social networks and multimedia platforms, since, as various studies point out, the use of modern sources of information (the web) influences interest in politics and specifically in issues of corruption (Pastrana, 2019; Nash, 2017). Although greater consumption of digital platforms generates a higher degree of exposure to diverse content, this ultimately strengthens convictions due to the ability of the algorithms used by these platforms to offer personalized and/or *ad hoc* information (Aruguete *et al.,* 2024; Vidal and Olivares, 2020; Waisbord, 2020).

In this way, social networks and technological applications are as important as television. Thus, WhatsApp, social networks, and broadcast television are the most frequently used media for information by Chileans, according to the CADEM survey (2025). It would therefore appear that consumption of traditional media has positive effects on political opinion and participation, while the use of social media, insofar as it is multidirectional, encourages other more critical dimensions of political participation (Navia and Ulriksen, 2017). Political discussions on the web are thus replacing the use of traditional media in democratic processes in Latin America (Rodríguez-Virgili and Fernández, 2017).

Indeed, as Serrano and Salvat (2015) point out, digitization, which began as just another means of communication, has expanded and filtered through all dimensions of our society, leading to the process of digital migration from the Information Society to the current Network Society, where digitization has changed the structure of information, accounting for the emergence of digital media in the press, and where the Internet is recognized as a particular medium characterized by integrating and merging all previously known media. In fact, there is a new configuration of the media system in Chile, where digital versions of traditional media coexist with new media originated for networks (Vernier *et al.,* 2017), and where social networks and technological applications are achieving the same informational importance as formal media, modifying the Chilean information system (López-Rabadán and Mellado, 2018; Ayala, 2016).

Journalism is undergoing substantial changes as a result of technological evolution and the social interaction of this evolution and is experiencing a crisis of credibility and trust (Sixto *et al.,* 2021). For this reason, some argue that educating news consumers from a critical perspective is the most challenging way to guarantee the quality of democratic information ecosystems (Cárcamo *et al.*, 2021; Cárcamo *et al.*, 2023), in the face of this phenomenon where citizens seem vulnerable to poor-quality news or unclear framing (Vernier *et al.*, Scheihing, 2018).

In fact, in the case of Chile, Mellado and Scherman (2020) conducted a study to map the diversity of sources and their quality on news platforms and media in Chile. According to their results, online media do not seem to contribute to greater diversity in the representation of social figures and appear to have a predilection for reporting on celebrities and figures who appear in traditional media, especially on television. This is the same conclusion reached by Couso (2011), who pointed out that, at least in Chile, and probably in other Latin American democracies as well, the new digital media have not radically transformed the public agenda set by traditional media.

In this sense, this type of news coverage reflects a bias typical of sensationalist media coverage: the personalization of the news, which expresses the media's tendency to focus its attention on the actors rather than their actions, increasing public attention through the deployment of explicit dramatic and controversial affective elements (Zunino and Focás, 2021; Palau and Davesa, 2013). Hence, corruption can be understood as an interpretive phenomenon (Sola-Morales and Rivera, 2017).

Hence, this research aims to present and validate techniques for analyzing and evaluating individuals involved in cases of political corruption using an automated news classifier that combines artificial intelligence, linguistics, and social sciences. To this end, this study analyzes, based on automatic language processing strategies, a topic analysis with LDA (Latent Dirichlet allocation) of a large corpus of information extracted from Sophia Search, an automated news classifier, which contains 33,699 news items on seven cases of corruption considered scandalous from media outlets throughout Chile.

Consequently, all the news items are processed to represent them as a list of nouns, key concepts, and entities, and patterns are added to search for their frequencies.

In addition, because frequencies of characters and concepts are obtained, a sentiment analysis is performed within natural language processing (NLP), applying it to moral evaluations of news stories about corruption cases. In this way, a sentiment analysis, also known as polarity analysis, was performed, obtaining various scores with a value between -1 and 1, which indicates at a glance whether the model is positive or negative. Values close to -1 indicate that it is very negative, values close to zero indicate that it is neutral, and values close to 1 indicate that it is very positive.

A triangulation of the data is then generated, both from the analysis of frequency and polarity or sentiments of the individuals linked to the corruption cases, and from the analysis of the conceptual frequency of the news items related to them.

In this regard, it should be noted that research focused on the construction of supervised machine learning models for classifying texts in this language is still scarce, but it is increasing (Sánchez *et al*., 2020), and most studies that classify texts in Spanish originate from the TASS project (Workshop on Sentiment Analysis in the Spanish Society for Natural Language Processing - SEPLN), which aims to promote research and development of new algorithms and techniques, and which is applied to the analysis of opinions in Spanish (García Cumbreras *et al.,* 2016).

In fact, there is research that applies sentiment analysis to investigate moral evaluations and possible biases that could be affecting news texts on the web, a challenging task due to the difficulties of processing some underlying linguistic phenomena such as negation, irony, sarcasm, and the ambiguity of natural language (Xu *et al.*, 2024, Sutriawan *et al.,* 2024), and more broadly, the subjectivity of opinions (García-Vega *et al.,* 2020), which require context for analysis. In fact, linguistics pays attention to technological innovation insofar as it has an impact on the Spanish language and its uses (Lovón and Parini, 2024).

These new approaches, which combine computational methods and big data, are beginning to be addressed in communication research in the social and political sciences (Arcila-Calderón *et al.,* 2016) and open new perspectives for their application in other sectors such as political communication.

This research culminates in the general conclusions of the study, highlighting the experience of using a tool that combines artificial intelligence with the academic disciplines of linguistics and social and political sciences, and indicating projections for future studies that will continue to allow for a deeper analysis of media pluralism in Chile.

Materials and methods

Due to the large volume of information published online, the burden falls on automated approaches to identify these frames within the text. Morstatter *et al.* (2018) test a machine learning model that recognizes patterns and identifies biases in journalistic frames. Similarly, Saez-Trumper *et al.* (2013) examine biases in online news sources and the communities surrounding them. To do so, they present unsupervised methods that consider three types of bias: selection or "control" bias, coverage bias, and statement bias, characterizing each of them through a series of metrics and identifying that these biases increase on platforms linked to social networks. There is also research that points to possible discriminatory biases in the algorithms that enable these analyses (O'Neil, 2017; Martínez *et al.*, 2020).

This type of methodological approach has the difficulty of detecting linguistic phenomena such as irony, sarcasm, and the ambiguity inherent in natural language (Xu *et al.,* 2024, Sutriawan *et al.,* 2024). Therefore, the enormous volumes of text require tools capable of automatically processing these messages without losing reliability. To this end, supervised machine learning processes can be used, which allow context to be taken into account and patterns to be revealed, but researchers must always be given the opportunity to refine the analysis (Baviera, 2017), as automatic analysis methods cannot dispense with rigorous manual coding if they are to be used reliably in research.

The methodological design of this research is presented below, with an emphasis on searching for news using Sophia software (Vernier and Ruiz, 2023) and analysis using open-source programming language due to the high volume of data (Oliva, 2017), and triangulation with data coded by the research team (Otero, 2020).

General Research Design

This study is mixed in nature, with both quantitative and qualitative methodologies, as it seeks to access and track large flows of news from different media outlets with a web presence on cases of political corruption. To this end, an analysis model based on web data mining will be followed, specifically under a set of media outlets indexed by the Sophia software. It also seeks to assess interpretive processes on the issue of political corruption in different media outlets in Chile, both nationally and locally.

In order to explore the news treatment of different topics in all their communication stages, a combination of qualitative and quantitative techniques is needed (Amadeo, 2008). Thus, content analysis is offered as an appropriate and valid technique for collecting and processing large amounts of data and subsequently identifying the topics present in the news (Sádaba *et al.*, 2012; Koziner, 2015).

Sample design and sampling strategy

The sample design was based on a comprehensive logic that depended on the advances that were being made in understanding the problem studied during the course of the research itself. Thus, the search was conducted using the Sophia platform, a tool for collecting large volumes of press news data, which allows social phenomena to be explored using *machine learning* and computational linguistics algorithms, thereby developing computational methodologies for human language analysis applied to the social and political sciences, using artificial intelligence and algorithms.

Sophia is software for the automatic collection and indexing of web documents that develops computational methods and indicators based on the collection of large volumes of press news data in various countries around the world, using *machine learning* and computational linguistics algorithms and developing computational methodologies for the analysis of human language applied to the social sciences.

The set of media outlets monitored is a heterogeneous group in terms of political orientation, geographically distributed throughout the country and with diverse media origins. The period of analysis is related to the creation of milestones in corruption scandals, therefore, all texts related to corruption cases considered scandalous between 2014 and 2022 are analyzed.

It is important to understand how the Sophia platform works. Sophia Search is a search engine that indexes millions of news articles published by the media (Vernier y Ruiz, 2023). Based on a list of keywords, a date range, and a list of media outlets, it returns the most relevant news articles, indicating their publication date, the media that published them, the title and content of the news article, and its URL.

Analysis plan

News items published on various corruption cases considered scandalous were stored. To do this, the following search was performed in the Sophia search engine, based on seven cases of media importance, filtering by dates between December 31, 2013, and January 1, 2023.

Caso Corpesca: ("Corpesca" OR "Jaime Orpis" OR "Marta Isasi" OR "Ximena Chong" OR "Angelini" OR "Ley General de Pesca" OR "Ley de Pesca") AND ("inhabilitación" OR "asesorías" OR "soborno" OR "cohecho" OR "corrupción" OR "caso" OR "infringir" OR "multa" OR "investigación" OR "multas" OR "acusados" OR "acusado" OR "infracciones" OR "infracción" OR "sanción" OR "sanciones")
Caso Cascadas: "cascadas" OR "sqm" OR "Soquimich" OR "Sociedad Química y Minera de Chile" OR "Calichera" OR "Oro blanco" OR "Superintendencia de Valores y Seguros" OR "Ponce Lerou") AND ("corrupción" OR "caso" OR "infringir" OR "multa" OR "investigación" OR "multas" OR "acusados" OR "acusado" OR "infracciones" OR "infracción" OR "sanción" OR "sanciones" OR "corte suprema")
Caso Penta: ("Penta" OR "Pentagate" OR "FUT" OR "Fondo de Utilidades Tributables" OR "Carlos Alberto Délano" OR "Carlos Délano" OR "Délano" OR "Carlos Eugenio Lavín" OR "Carlos Lavín" OR "Hugo Bravo" OR "Jovino Novoa") AND ("corrupción" OR "caso" OR "infringir" OR "multa" OR "investigación" OR "multas" OR "acusados" OR "acusado" OR "infracciones" OR "infracción" OR "sanción" OR "sanciones" OR "soborno" OR "cohecho")
Caso Caval: ("Caval" OR "Sebastián Dávalos" OR "Dávalos" OR "Machalí" OR "Natalia Compagnon" OR "Compagnon" OR "Nueragate" OR "Sociocultural de la Presidencia") AND ("hijo" OR "Michelle Bachelet" OR "crédito" OR "Banco de Chile" OR "Banco Chile" OR "madre" OR "predios" OR "plano regulador" OR "suelo" OR "uso urbano" OR "tráfico de influencias" OR "corrupción" OR "caso" OR "infringir" OR "multa" OR "investigación" OR "multas" OR "acusados" OR "acusado" OR "acusadas" OR "acusada" OR "infracciones" OR "infracción" OR "sanción" OR "sanciones")
Caso Milicogate: ("Milicogate" OR "Ley reservada del cobre" OR "ejército" OR "Corte Marcial" OR "militar" OR "Comando de Apoyo a la Fuerza" OR "defensa" OR "Milico" OR "Milicos" OR "Juan Miguel Fuente-Alba" OR "Juan Miguel Fuente Alba" OR "Juan Miguel Alba" OR "Héctor Ureta" OR "Clovis Montero" OR "Francisco Huincahue") AND ("Codelco" OR "Cobre" OR "datos borrados" OR "uniformado" OR "uniformados" OR "corrupción" OR "caso" OR "infringir" OR "multa" OR "investigación" OR "multas" OR "acusados" OR "acusado" OR "infracciones" OR "infracción" OR "sanción" OR "sanciones" OR "soborno" OR "cohecho")
Caso Pacogate: ("Pacogate" OR "Paco" OR "Carabineros" OR "fuerza policial" OR "Eduardo Gordon" OR "Flavio Echeverría" OR "Felipe Ávila") AND ("documentos faltantes" OR "documentos destruidos" OR "corrupción" OR "caso" OR "infringir" OR "multa" OR "investigación" OR "multas" OR "acusados" OR "acusado" OR "infracciones" OR "infracción" OR "sanción" OR "sanciones" OR "soborno" OR "cohecho")
Caso Rodrigo Rojas Vade: ("Rodrigo Rojas Vade" OR "Rojas Vade" OR "pelado Vade) AND ("cáncer" OR "convención" OR "constituyente" OR "mentira" OR "enfermo" OR "donaciones" OR "tratamiento" OR "exconvencional" OR "pelo" OR "pelado" OR "cejas" OR "estallido social" OR "leucemia" OR "desnudos" OR "protestas" OR "salud" OR "vicepresidencia" OR "sífilis" OR "corrupción" OR "caso" OR "infringir" OR "multa" OR "investigación" OR "multas" OR "acusados" OR "acusado" OR "infracciones" OR "infracción" OR "sanción" OR "sanciones")

These seven outputs from Sophia Research were then combined into a CSV document to be processed and analyzed using Python language, to which some variables were added in order to make a general characterization of the media that report on corruption cases in Chile.

Next, in order to identify and compare the treatment of media coverage of Chilean corruption scandals between 2014 and 2022, all the news items were processed to represent them as a list of nouns (NOUN); key concepts (PATTERN); entities such as (PER) and (ORG); and patterns were added to search for their frequencies.

In addition, because the frequencies of characters and concepts were analyzed, moral evaluations of news stories about corruption cases can be analyzed. In this way, a sentiment analysis, also known as polarity analysis, was performed, obtaining various scores with a value between -1 and 1, which indicates at a glance whether the model is positive or negative. Values close to -1 indicate that it is very negative, values close to zero indicate that it is neutral, and values close to 1 indicate that it is very positive.

Sentiment analysis is a use case of natural language processing (NLP) and is included in the text classification category, which involves classifying a text into various sentiments, such as positive, negative, or neutral. Therefore, the ultimate goal of sentiment analysis is to decipher the underlying sentiment of a text. This is also known as Opinion Mining.

This resulted in an analysis of 32 individuals involved in the seven events considered politically scandalous, which were triangulated with 1,120 concepts that helped to delve deeper into the evaluation of the individuals.

Analysis and results

The analysis presented below was developed to study a set of national news stories about corruption events considered scandalous in Chile between 2014 and 2022, such as: the Cascadas case, also known as the SQM case; the Penta case; the Caval case; Milicogate; Pacogate; the Corpesca case; and the Rodrigo Rojas Vade case.

In total, the analysis was applied to a set of 33,699 news articles, distributed across 64 media outlets throughout Chile, available on the web (Table 1).

Table 1

News coverage of Chilean corruption cases by media outlet on the web between 2014 and 2022

	Media outlet	Posts	Media outlet		Posts
1	emol	4464	33	diariodevaldivia	15
2	hours24	3413	34	lavozdepaillaco	11
3	113	3289	35	newslosrios	10
4	elciudadano	2940	36	key9	10
5	elllanquihue	2822	37	enlanoticia	9
6	CNN Chile	2586	38	enlalinea	8
7	latercera	2562	39	elproa	8
8	elaustral	2154	40	aconcaguadigital	7
9	ahoranoticiasmega	1617	41	work	6
10	the star of Chiloe	1242	42	diariofutrono	5
11	biobiochile	1145	43	radionuevoamanecer	4
12	chvnoticias	992	44	radioestrelladelmar	4
13	ciper	913	45	masnoticia	4
14	eldinamo	630	46	local seminar	3
15	elmostrador	601	47	seconds33	3
16	elvacanudo	544	48	prensadelestuario	3
17	elrepuertero	507	49	elinsular	3
18	lanacion	271	50	elcalbucano	3
19	radiosago	247	51	diariolagoranco	3
20	laquinta	125	52	soypuertomontt	2
21	laserenaonline	96	53	rioenlinea	2
22	the navigable	90	54	radio reception	2
23	laregiontoday	60	55	fresiaahora	2
24	radiopudeto	49	56	diariosanjose	2

	Media outlet	Posts	Media outlet		Posts
25	diariolaguino	35	57	diariolaunion	2
26	elheraldoau stral	30	58	soyosorno	1
27	suractual	29	59	soychiloe	1
28	elquellonino	25	60	elinformador	1
29	ovallehoy	24	61	diarioriobueno	1
30	elhuemul	21	62	Diariopaillaco	1
31	elperiodista	19	63	diariomafil	1
32	elobservador	19	64	diariolanco	1

Likewise, in order to determine which variables are interesting to analyze, an overview of media coverage of corruption in Chile is presented, since it allows us to see the geographical distribution of the media outlets that published at least one news item on a corruption scandal (illustration 1). In this regard, it is important to note that these 33,699 news items, distributed across 64 media outlets, are geographically located in 39 municipalities throughout Chile, and as with most news coverage in Chile, they are concentrated in the Metropolitan Region. On this occasion, slightly more than 70% of news broadcasts were in that region, and in fact, almost 25% of all news broadcasts in the period between 2014 and 2022 were in the municipality of Santiago.

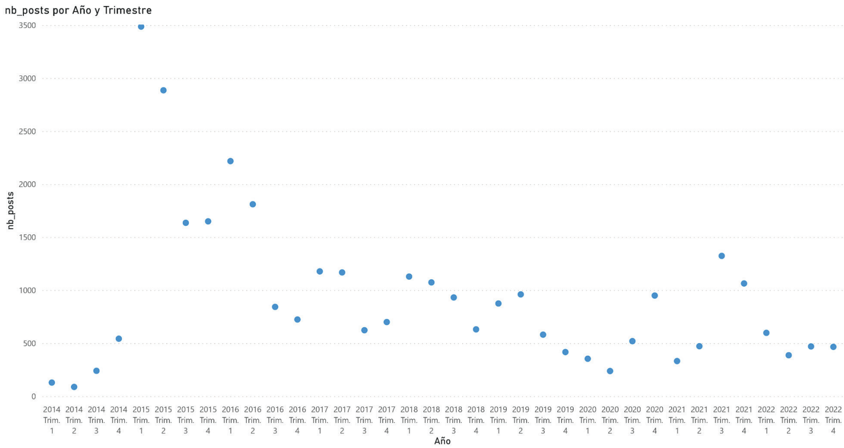
There is also an overview of corruption news broadcasts between 2014 and 2022 over time (Illustration 2). There, the highest *peak* of news posts about corruption coincides with the beginning of Michelle Bachelet's second term in office in the first quarter of 2015, when the issue of illegal financing came to light, and then the second *peak* coincides with the Rodrigo Rojas Vade case in September 2021, when, after an interview with La Tercera, former convention member Rodrigo Rojas admitted that he did not have cancer, despite having obtained donations for his supposed treatment.

It is noteworthy that between these *peaks*, the number of posts about corruption in Chile remained relatively stable, which does not mean that there were no other cases of corruption.

Illustration 1

Geographical distribution of media outlets covering corruption between 2014 and 2022

Illustration 2
Distribution of the number of news items on corruption cases in Chile between 2014 and 2022



Likewise, in order to compare types of media on the web, it is possible to search for characteristics of each media outlet, such as year of origin and legal status, etc.

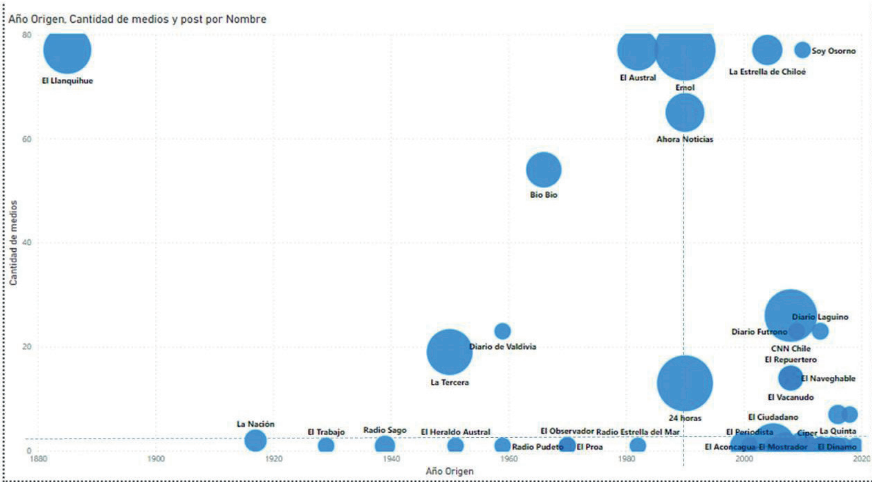
In relation to the year of origin of the media outlet, as established by Couso (2011), in Chile, there is a high degree of digitization, and since its installation in the country, "new media" have begun to emerge, where it is found (1) traditional media, including their digital versions, which were created before 1990 and the professional arrival of the internet in the country; and (2) emerging media, including their digital versions, which were created before 1990 and the professional arrival of the internet in the country.

Regarding the legal status of each media outlet, owners who have more than one media outlet can be categorized as (1) hegemonic, and those who have only one media outlet as (2) non-hegemonic.

In this way, it is possible to categorize the media that cover corruption news into traditional hegemonic media (HT), traditional non-hegemonic media (HNT), emerging hegemonic media (HE), and emerging non-hegemonic media (NHE), as shown in Figure 3.

Illustration 3

Distribution of corruption news coverage from 2014 to 2022 in online media, by year of creation and number of media outlets owned by the owner.



Thus, as seen in Table 2, it is evident that under this first characterization by media outlets covering corruption news, the media outlets that post the most news about events considered political scandals, in descending order, are the emerging hegemonic media outlets (HE), which account for 39.1% of the media analyzed and cover 46.5% of the news in their posts; then there are the Traditional Hegemonic (HT) media, which account for only 9.4% of the media analyzed and cover 36.3% of the news in their posts; then there are the non-hegemonic emerging (NHE) media, which account for 40.6% of the media analyzed and cover 16.1% of the news in their posts; and finally, there are the Traditional Non-Hegemonic (TNH), which account for 10.9% of the media analyzed and cover only 1.1% of the news in their posts.

Table 2
Percentages of media outlets and their coverage, according to categorization

	Post	Media	% of Total Posts	% Media of Total
HE	15,654	25	46.5	39.1
HT	12,243	6	36.3%	9.4%
NHE	5,439	26	16.1	40.6%
TNH	363	7	1.1	10.9%
	33,699	64	100	100

As seen in Figure 3 and Table 2, emerging media outlets, i.e., originating from 1990 onwards, account for 79.7%, almost 80%, of the media outlets analyzed. And if looking at it in terms of hegemony, i.e., if the legal entity of these media outlets has one or more media outlets under their control, 82.8% of the posts covering news related to corruption cases come from hegemonic media outlets. In other words, despite the high emergence of media outlets made possible by digitization, the volume of coverage continues to be in the hands of hegemonic media outlets, whether emerging or traditional.

Now, if the goal is to analyze who, what, and how corruption cases are discussed, in addition to comparing the different treatments of corruption scandals, then it is convenient to analyze each corruption case individually. To do so, the coverage by case is considered, which, as seen in Tables 3.1 and 3.2, varies depending on the news story.

Table 3.1
Percentage of corruption case coverage by type of media

	%Post Corpesca	%Post Cascadas	%Post Pentagate	%Post Caval	%Post Milicogate	%Post Pacogate	%Post Rojas Vade	Total%of Post
HT	30.31	40.90	34.25	39.54	36.75	33.56	44.68%	36.30
HE	42.90	41.76	47.39	49.67	40.47	45.52	41.88	46.50
TNH	2.15	0.78	0.49	0.55	0.27	2.11	1.77	1.10
NHE	24.65	16.56	17.87	10.25	22.51	18.82	11.66	16.10
	100.00%	100.00%	100.00	100.00%	100.00	100.00%	100.00	100.00%

Table 3.2

Percentage of corruption cases covered by type of media

	Post % of Totalb TM Corpesca	Post % of Total TM Cascadas	Post % of Total TM Pentagate	Post % of Total TM Caval	Post % of Total TM Milicogate	Post % of Total TM Pacogate	Post % of Total TM Rojas Vade	Total
HT	14.35	27.96	19.56	19.21	4.24	10.94%	3.74	100.00
HE	16.50	23.19	21.99	19.61	3.79	12.06	2.85	100.00%
NHT	34.30	17.94	9.50	8.97	1.06	23.22%	5.01	100.00%
NHE	24.37	23.64	21.32	10.40	5.42	12.81	2.04	100.00%
	22.38	23.18	18.09	14.55	3.63	14.76	3.41	100.00

Below is a summary of the results that expand on the analysis of 32 individuals involved in the seven events considered politically scandalous, which were triangulated with the 1,120 concepts that helped to deepen the evaluation of the individuals.

It is important to clarify that, as this is an investigation related to cases of corruption, polarity or sentiment assessments tend to lean toward negative scores (García-Vega *et al.,* 2020; Martínez *et al*., Sustaeta, 2020), because these individuals, including prosecutors and whistleblowers, are associated with concepts linked to the crimes mentioned in the news, which, due to their moral aspect, will result in negative averages.

Thus, using a color scale of coverage percentages and the average polarity scored by the characters, it can be noted that the character with the highest media coverage, despite being involved in one of the cases with the least amount of news, is Rodrigo Rojas Vade. This may be because, unlike the other cases, only one person is accused of committing the crime, which means that interest in covering the event is directed solely at that person, triggering a loss of credibility for the first constitutional process.

Something similar occurs in the Milicogate case, in which former commander Fuente-Alba is being prosecuted and investigated by visiting minister Romy Rutherford, both with high percentages of coverage and average negative polarity. Although the other defendants charged by the visiting minister do not receive high coverage, they do have high negative scores on the sentiment scale. It is noteworthy that, although they are evaluated negatively in most types of media, as defendants, they are not evaluated more negatively than Rutherford.

The same is true of prosecutor Ximena Chong in the Corpesca case, who always has a more negative average polarity than Jaime Orpis, one of the people involved in the case and convicted for it.

In this regard, in the case of Cascadas, the low coverage of Julio Ponce Lerou, the main benefactor of the sale of SQM's cascade companies, is striking, with coverage not reaching 7% in several types of media except for emerging non-hegemonic media, which mention him in almost 19% of the news items related to the case.

In contrast, coverage of Natalia Compagnon in the Caval case is high, close to 30% and 40% depending on the type of media, and with it, so is her high negative polarity in almost all types of media.

Regarding the Penta case, the high coverage of the Independent Democratic Union (UDI) is striking. It is the only organization evaluated in this investigation as a character, obtaining the highest negative polarity in almost all types of media, except to the emerging hegemonic media, where Sebastián Piñera reaches that position.

In relation to the Pacogate case, the negative score on the polarity scale of former police officer Bruno Villalobos stands out, even above Eduardo Gordon, who was convicted of embezzlement of public funds and is now under house arrest and required to sign in monthly.

In all cases, the news is personalized, focusing attention on the actors rather than their actions, with explicit emotional elements (Zunino and Focás, 2021; Palau and Davesa, 2013). Tables 4 and 5 summarize all these evaluations statistically.

Table 4

Percentage of coverage and evaluation of individuals in seven corruption cases, by type of media

Corruption case	Person linked to the case	Involvement	Percentage of mentions				Moral assessment			
			HT	HE	TNH	NHE	HT	HE	TNH	NHE
Corpesca	Ximena Chorra	Direct (tax)	34.04	24.51	9.23	18.82	-0.0812	-0.0550	0.0077	-0.0589
	Jaime Orpis	Direct (convicted)	2576	23.16	3.08	19.09	-0.0501	-0.0431	0.0077	0.0422
	Sebastián Piñera	Indirect	5.28	6.85	2.31	18.22	-0.0169	-0.0154	0.0077	0.0683
Waterfalls	Patricio Contesse	Direct (convicted)	19.23	12.95	4.41	13.47	-0.0291	-0.0296	-0.00147	-0.0366
	Pablo Lonaueira	Direct (accused)	15.99	11.61	5.88	16.37	-0.0285	-0.0260	-0.0294	0.0421
	UDI	Indirect	7.69	6.49	4.41	16.92	-0.0266	-0.0173	-0.0147	-0.0539
	Michelle Bachelet	Indirect	6.68	6.02	7.35	16.64	-0.0154	-0.0156	0.0000	0.0539
	Julio Ponce Lerou	Direct (convicted)	6.85	6.24	1.47	18.44	-0.0212	-0.0205	0.0000	0.0698
Pentagate	UDI	Indirect	17.50	14.03	19.44	25.04	-0.0551	-0.0378	0.0000	0.0904
	Carlos Gaiardo	Direct (fiscal)	14.34	12.50	8.33	11.72	-0.0388	-0.0303	-0.0556	0.0375
	Jovino Novoa	Direct (convicted)	13.78	14.61	2.78	15.77	-0.0368	-0.0352	0.0000	-0.0574
	Carlos Delano	Direct (convicted)	12.66	14.73	13.89	21.36	-0.0244	-0.0358	-0.0556	0.0804
	Carlos Lavin	Direct (convicted)	8.75	13.66	5.56	12.94%	-0.0160	-0.0361	0.0000	-0.0436
	Ivan Moreira	Direct (unrestrained)	9.67	11.49	30.56	11.79	-0.0248	-0.0312	-0.0278	0.0421
	Sebastian Pinera	Indirect	9.55	8.55	19.44	29.48	-0.0324	-0.0193	-0.0278	-0.1072
Caval	Natalia Compao non	Direct (convicted)	41.05	40.19	29.41	34.85	-0.1090	-0.1052	0.1176	0.1068
	Michelle Bachelet	Indirect	36.13	30.31	23.53	30.93	-0.0993	-0.0748	-0.0882	-0.1020
	Sebastián Davalos	Direct (accused)	27.26	34.62	23.53	34.54	-0.0765	-0.0965	0.0000	0.0942
	Mauricio Valero	Direct (convicted)	16.56	15.09	8.82	11.77	-0.0382	-0.0405	0.0294	0.0283
	Sergio Moya	Direct (fiscal)	13.10	8.42	5.88	7.69	-0.0325	-0.0207	0.0294	-0.0251
Milicogate	Romv Rutherford	Direct (fiscal)	52.58	31.16	75.00	51.81	-0.1642	-0.0988	0.0000	0.1717
	Juan Miguel Fuente-Alba	Direct (convicted)	62.18	61.81	25.00	57.83	-0.1956	-0.1508	0.0000	-0.1687
	Humberto Oviedo	Direct (accused)	29.89	19.10	25.00	21.39	-0.0941	-0.0653	0.0000	0.0633
	Ricardo Martinez	Direct (accused)	11.81	5.03	25.00	15.96	-0.0277	-0.0184	0.0000	-0.0542
Pacogate	Eduardo Gordon	Direct (convicted)	7.00	4.48	4.55	6.62	-0.0207	-0.0111	0.0000	-0.0217
	Bruno Villalobos	Direct (accused)	18.86	17.85	6.82	12.23	-0.0521	-0.0585	-0.0114	0.0395
	Michelle Bachelet	Indirect	5.57	5.85	10.23	8.41	-0.0157	-0.0211	0.0341	-0.0293
	Sebastián Piñera	Indirect	8.00	2.53	14.77	13.25%	-0.0271	-0.0068	0.0114	0.0535
Rojas Vade	Rodrigo Roías Vade	Direct (convicted)	74.74	74.39	78.95	76.80	-0.2484	-0.1826	-0.1053	-0.2160
	Elisa Loncon	Direct (complainant)	24.43	17.15	21.05	15.20	-0.0605	-0.0178	0.0000	0.0160
	Jaime Bassa	Direct (complainant)	22.13	16.70	5.26	16.80	-0.0689	-0.0245	0.0000	0.0080
	Tomás Ramírez	Direct (lawyer)	5.43	6.90	5.26	8.00	-0.0146	-0.0111	0.0000	-0.0320

Table 5
Concepts related to seven cases of corruption by type of media

Corruption case	Most frequent nouns in the case (in descending order)				Most frequent noun phrases in the case (in descending order)			
	HT	HE	NHT	NHE	HT	HE	NHT	NHE
Corpesca	case; prosecutor, senators crimes; law: years: investigation: company; bribery; millions;	case; prosecutor; law; senator years; investigation; country; fishing; millions; resources;	fishing; resources; resource; law; fishermen; case; year; people; years;	millions; companies; years; company; law; case; project; year; country; investigation;	national prosecutor; tax crimes; regional prosecutor: preventive detention; oral trial; millions of pesos; artisanal fishermen: irregular financing; house arrest; precautionary measure;	artisanal fishermen; tax crimes; artisanal fishing; preventive detention; national prosecutor; millions of pesos; fishing law: bill; regional prosecutor; fishing resources:	artisanal fishermen; artisanal fishing: recreational fishing: hydrobiological resources; illegal fishing; legal origin; fishing law: fishery resources; criminal prosecution; minimum size:	bill: artisanal fishermen: millions of pesos; artisanal fishing; general manager; millions of dollars; economic groups; illegal financing; fishery resources; tax crimes:
Cascadas	case; prosecutor; millions; investigation: senator years; people; crimes; company; minister;	case; prosecutor; investigation; company; millions; senators crimes; years; minister;	years; case; information; sector; hours; people; municipality; millions; country;	millions; company; companies: case; years: year; investigation; satiety: president: country;	national prosecutor; tax crimes; oral trial; regional prosecutor; general manager; irregular financing; presidential candidate; millions of pesos; definitive dismissal: political campaigns;	tax crimes; national prosecutor: general manager: political campaigns; irregular financing; millions of pesos; regional prosecutor; presidential candidate; oral trial; formalization hearing;	last year; regional prosecutor; tax crimes; millions of pesos; national level; weekend; oral trial; social media; national prosecutor: irregular financing;	general manager; tax crimes; illegal financing; millions of dollars; millions of pesos; false receipts; political campaigns; national prosecutor; last year; irregular financing;
Pentagate	case; prosecutor; senator crimes; investigation; millions; deputy; years; minister; president;	investigation; prosecutor: senator; crimes; years; millions; minister; president; financing;	case; years; days; investigation; year; people; prosecutor; millions; parliamentarian: crimes;	millions; case; companies; company; years; investigation; project; society; year; president;	national prosecutor; tax crimes; irregular financing; regional prosecutor; political campaigns; house arrest; secretary general; precautionary measures; general manager; constitutional indictment;	tax crimes; political campaigns; national prosecutor; millions of pesos; preventive detention; irregular financing; house arrest; precautionary measures; regional prosecutor; national confinement;	last year; tax crimes; appeal; preventive detention; millions of pesos; irregular financing; political campaigns; facilitation of ballots: constitutional indictment: national prosecutor:	millions of dollars; general manager; tax crimes: millions of pesos; illegal financing; political campaigns; false ballots: tax havens; confidential contributions; draft

Caval	case; prosecutor; minister; millions; son; company; investigation; people; deputy; country;	case; prosecutor; investigation; millions; company; son; information; country; minister;	case; years; millions; people; country; year; week; hours; days; information;	millions; case; years; company: country; prosecutor; government; year; investigation; project;	regional prosecutor; national prosecutor; influence peddling; tax crimes; oral trial; millions of pesos; sociocultural director; last week; secretary general; investigative commission;	regional prosecutor; millions of pesos: influence peddling; investigative commission; tax crimes; national prosecutor; precautionary measures; national arrest warrant; oral trial; privileged information;	millions of pesos; millions of people; weekend; preventive detention; days in prison; tax crimes: seizure of assets: legal action; district week; economic growth;	millions of pesos; influence peddling: regional prosecutor; national prosecutor; public opinion; millions of dollars; political class; bill; tax crimes: media:
Milicogate	chief; general; millions; expenses: minister; case; commander; institution; investigation;	chief: millions: commander; general; case; minister; investigation; years; expenses; institution;	general; case; chief; money; travel; pesos; information; trip; minister;	millions; chief; general; minister; investigation; case; years; commander; institution; expenses;	confidential expenses; public funds; embezzlement of funds: millions of pesos; military institution; preventive detention; use of expenses; military justice; crime of embezzlement; investigative commission;	confidential expenses; millions of pesos; public funds; embezzlement of funds; preventive detention; military institution; investigative commission; illicit enrichment; military justice; repeated offense:	excess baggage; false invoices; tourism companies; service commission; emails; total amount; chief financial officer; ticket prices; invoice value; million pesos;	confidential expenses; millions of pesos; public funds; embezzlement of funds: millions of dollars; military institution; preventive detention; repeat offense; total amount; public funds;
Pacogate	institution; case; prosecutor; fraud; general; millions; investigation; director; people;	institution; fraud; millions; investigation; years; prosecutor; general; people; director	years; people; case; year; millions; fraud; institution; hours; days;	millions; years; investigation; year; people; company: general; country; institution;	confidential expenses; millions of pesos; regional prosecutor; uniformed police; public funds; embezzlement of funds; preventive detention; national prosecutor: house arrest; national detention:	millions of pesos; confidential expenses; preventive detention; public funds; embezzlement of funds: regional prosecutor; criminal association; house arrest; uniformed police; national prosecutor;	millions of pesos; pretrial detention; last year; weekend; corruption cases: last week; uniformed police; oral trial; social unrest; social media;	millions of pesos; confidential expenses; public funds; preventive detention: social unrest; millions of dollars; embezzlement of funds; last year; organized crime; social media;
Rojas Vade	convention; cancer; table; process; resignation; convention delegates; illness; organ; situation;	convention; cancer; process; case; table; convention delegates: resignation; situation; people; illness;	cancer; convention; situation; resignation; statement; illness; crime; case; body; facts;	convention; cancer; resignation; constituents; process; September; case; work; body; convention delegates:	board of directors; constituent process; constituent body; constitutional reform; social media; constituent convention; bill; weekend; social unrest; millions of pesos;	board of directors; constituent process; social media: constitutional reform; constituent convention; indigenous peoples; social unrest; public trust: point of view; declaration of interests;	constitutional convention; board of directors; constitutional process; constitutional body; declaration of interests; millions of pesos: declaration of assets; regulatory vacuum: formal resignation; immune thrombocytopenia:	constitutional process; social media; constituent body; board of directors; constitutional reform; social unrest; indigenous peoples; declaration of assets; constituent convention; billions of pesos;

The triangulation of the data obtained, both from the frequency analysis of mentions and from that of polarity or sentiments of the figures linked to cases considered politically scandalous in Chile, and the conceptual frequency analysis of news related to them, allows to identify predominant topics and strategies in the journalistic coverage of Chilean corruption scandals that occurred between 2014 and 2022 by Chilean news media present on the web, as well as to compare the different treatments of these corruption scandals.

Furthermore, this type of tool, by providing the most relevant news, made it possible to carry out a content analysis of the news (see Table 6). Thus, this study shows that the hegemonic media, whether traditional or emerging, treat the news in a similar way, with differences only in the emphasis on who and what they talk about. It should be noted that this type of media accounts for almost 83% of all news broadcasts related to the corruption cases studied in this research. Therefore, as pointed out by Mellado and Scherman (2020) and Couso (2011), online media do not seem to contribute to greater diversity in the representation of different social sectors, and even seem to have a predilection for reporting on celebrities and personalities who appear in traditional media.

Table 6
Example of news stories about different figures involved in corruption cases

Personality	Type of media	No. of news items on the case	No. of news items on the figure	Person count	Character score	Examples
Eduardo Gordon (Pacogate)	HT	1400	98	7.00%	-0.0207	"Prosecutor Eugenio Campos, who is leading the investigation in the fraud within the Carabineros, faces an intense day on Wednesday, as a series of witness statements began at 10:00 a.m. at the OS-7 offices of the Carabineros. According to the information provided, at least six people have been summoned, including former Deputy Comptroller Dorothy Pérez, former Comptroller Ramiro Mendoza, and former Carabineros Director General Eduardo Gordon." La Tercera, Nov. 20, 2018.

Humberto Oviedo (Milicogate)	HE	597	114	19.10%	-0.0653	"The decision was made after retired General Humberto Oviedo, who is also under investigation for misusing Army funds amounting to nearly $4.5 billion, who was released on August 1." CHV, Aug. 21, 2019.
Jaime Orpis	NHT	130	4	3.08%	0.0077	"Orpis is terrified of freedom: a final appeal to the Court would delay his entry into prison. 'It is going to be my home.' That is how former UDI senator Jaime Orpis described the prison where he will serve the sentence he received in the Corpesca case. Radio Saqo, Nov. 16, 2021.
UDI	NHE	1448	245	16.92	-0.0539	"This is a problem that affects practically the entire spectrum of political parties and, at this particular moment, has come to the fore on the side of the UDI." El Ciudadano, October 9, 2014.

Regarding the analysis of individuals, the tactic of some media outlets, particularly non-hegemonic traditional media, to omit mention of crimes is striking. They employ the strategy of using concrete nouns that do not carry a negative moral connotation, thus neutralizing the moral weight of the act considered scandalous, and even use positive adjectives such as "freedom," so that several of the characters highlighted in these media outlets have a neutral or even positive evaluation on the polarity or sentiment scale, thus perpetuating the *status quo* of the information ecosystem, without an obvious difference in perspective from the hegemonic media, which is not even desired due to possible connections between political and public authorities and the various media outlets analyzed here.

Likewise, emerging non-hegemonic media outlets differ from hegemonic media outlets in that they place greater emphasis on covering national authorities from across the political spectrum, but with greater perseverance in the sector close to the UDI, relating them to negative words and linking them to the Pinochet dictatorship, so their evaluations tend to be more negative. Also noteworthy in this type of media is the high coverage of figures who are not directly involved in the corruption cases, such as Sebastián Piñera and Michelle Bachelet, both leaders of Chile since the corruption cases began to come to light.

It is important to note that during the analysis process, the research team had to provide context in the case of some ambiguities that arose, since, as suggested by Xu *et al*. (2014) and Sutriawan *et al.* (2024), it is important to make the necessary adjustments to the software's output in the event of lexical ambiguities. This occurred in cases where two people had exactly the same name and the software was unable to differentiate between them (as is the case with company managers where fathers and sons have the same name and both are managers). It also occurred when individuals were not always named in the same way in the news, so the software assumes that if a surname is separated from the first name, it will treat the surname as a separate character (e.g., Piñera and Sebastián Piñera), or if there is an accent in the character's name and some media outlets refer to them without an accent, then the program also assumes that they are two different characters (e.g., Loncon and Loncón).

Conclusions

This study identifies predominant treatments and strategies in the journalistic coverage of Chilean corruption scandals that occurred between 2014 and 2022 by Chilean news media present on the web, in addition to comparing the different treatments of these corruption scandals.

It is also important to note that there are some limitations to using a tool that combines artificial intelligence, linguistics, and social sciences. Three limitations were observed in this study in general terms.

The first occurred in cases where two individuals had the same name, as is the case with company managers where fathers and sons have the same name and both are managers.

The second occurred when the individuals are named differently, either by their first name, last name, or nickname, or if there is an accent or apostrophe in the individual's name, and some media outlets name them with or without an accent, assuming that they are different individuals.

Both of these limitations result from the difficulty of analyzing linguistic phenomena in natural language using automatic language processing techniques, such as ambiguity (Xu, *et al.,* 2024; Sutriawan *et al.,* 2024). In both cases, lexical ambiguities occur where two people have the same name or one person is named differently, which requires in-depth observation by the research team to provide context and generate an in-depth analysis.

The third limitation of the study, which is a little more specific, highlights the fact that those who administer justice, such as Rurherford and Chong, are charged with negative polarity, which suggests that further adjustments are needed to automatic language processing, as to some extent, it is proof of the limitations that these tools may have. However, despite these limitations, the program provided a rewarding guide for metadata analysis.

Given the limitations presented in this study, future studies are planned that are related to the computerized analysis of large volumes of data on the web, in order to identify other possible biases in multimedia texts.

The sentiment analysis used in this study allowed to find different biases, but it failed to delve deeper into others, such as gender bias, for example, by noting that female prosecutors were rated worse than convicted, who were mostly men. It also failed to delve deeper into discriminatory biases toward Elisa Loncon, who was linked to indigenous peoples, connections that could presumably have meaning, whether negative or positive, but that would affect her evaluation on the polarity scale.

Investigating the moral evaluations and possible biases that could be affecting the texts of news articles on the web is a challenging task, due to the difficulties of processing some underlying linguistic phenomena, as seen in this study with the fact of being able to individualize names in a single way, or difficulties inherent in this type of analysis that consist of calculating or measuring irony, sarcasm, and the subjectivity of opinions (García-Vega *et al.,* 2020), as well as calculating and measuring possible discriminatory biases in the algorithms that enable these analyses (O'Neil, 2017; Martínez *et al.*, 2020), in order to generate new algorithms that enable more equitable and non-discriminatory studies.

Acknowledgments

This work was carried out thanks to funding from National Doctorate Grant No. 21201115 and Fondecyt Regular No. 1241726, both from the National Research and Development Agency (ANID), Chile.

References

Amadeo, B. (2008). Framing: Modelo para armar. En M. T. Baquerín (ed.), *Los medios ¿aliados o enemigos del público?* EDUCA.

Arcila-Calderón, C., Barbosa-Caro, E. and Cabezuelo-Lorenzo, F. (2016). Técnicas Big Data: Análisis de textos a gran escala para la investigación científica y periodística. *El Profesional de La Información, 25*(4), 623-631. https://doi.org/10.3145/epi.2016.jul.12

Aruguete, N., Batista, F., Calvo, E., Guizzo-Altube, M., Scartascini, C. and Ventura, T. (2024). Framing fact-checks as a "confirmation" increases engagement with corrections of misinformation: a four-country study. *Sci Rep 14*, 3201. https://doi.org/10.1038/s41598-024-53337-0

Ayala, T. (2016). Redes sociales, poder y participación ciudadana. *Revista Austral de Ciencias Sociales*, *26,* 23-48. https://acortar.link/1ykPo9Baviera, T. (2017). Técnicas para el análisis de sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. *Dígitos, 3*, 33-50. https://doi.org/10.7203/rd.v1i3.74

Cárcamo, L., Cárdenas, C., Sáez, D. and Toural, C. (2021). Fake news en Chile y España: ¿Cómo los medios nos hablan de noticias falsas? *Journal of Iberian and Latin American Research*, 1-18. https://doi.org/10.1080/13260219.2020.1909849

Cárcamo-Ulloa, L., Cárdenas-Neira, C.y Scheihing-García, E., Sáez-Trumper, D., Vernier, M.y Blaña-Romero, C. (2023). On politics and pandemic: how do Chilean media talk about disinformation and fake news in their social networks? *Societies, 13*(25), 1-14. doi.org/10.3390/soc13020025

CIPER. (2015a). *El periodismo que remece a Chile*. Catalonia.

CIPER. (2015b). *Lo mejor de CIPER 3. El periodismo que remece a Chile*. Editorial Catalonia.

Cordero, R. and Marín, C. (2006). Los medios masivos y las transformaciones de la esfera pública en Chile. *Documento de trabajo ICSO, 7*(2). https://acortar.link/1ykPo9

Couso, J. (2011). El mercado como obstáculo a la libertad de expresión: la concentración de la prensa escrita en Chile en la era democrática. En B. Sorj, *Democracia y medios de comunicación. Más allá del estado y el mercado* (pp. 109-142). Catálogos.

García Cumbreras, M. Á., Villena Román, J., Martínez-Cámara, E., Díaz Galiano, M. C., Martín-Valdivia, M. T. and Ureña-López, L. A. (2016). Resumen de TASS 2016. *TASS 2016: Workshop on Sentiment Analysis at SEPLN Proceedings*, 13-21. https://acortar.link/1ykPo9

García-Vega, M., Díaz-Galiano, M., García-Cumbreras, M., Plaza del Arco, F., Montejo-Ráez, A., Jiménez-Zafra, S., Martínez, E., Aguilar, C., Sobrevilla, M. A., Chiruzzo, L. and Moctezuma, D. (2020). Overview of TASS 2020: Introducing Emotion Detection. *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020)*, 163-170. https://acortar.link/1ykPo9

Heiss, C. (2023). De las calles a las urnas. El estallido social tras el fracaso del proceso constituyente. En Batthyány, K. (ed.), *Constituir en la revuelta. Nuevos desafíos a la democracia*. CLACSO. https://acortar.link/1ykPo9

Irrarázabal, I. and Cruz, F. (2023). Confianza institucional en Chile: un desafío para el desarrollo. *Punto de referencia* (682), 1-34. https://acortar.link/1ykPo9

Koziner, N. (2015). El framing: Un programa de investigación para el estudio de las comunicaciones mediáticas. *Animus, 14*(28), 22-45. https://doi.org/10.5902/2175497720203

Lopez-Rabadan, P. and Mellado, C. (2018). Twitter as a space for interaction in political journalism. Dynamics, consequences and proposal of interactivity scale for social media. *Communication & Society 32*(1), 1-18. https://doi.org/10.15581/003.32.37810

Lovón, M. and Parini, A. (2024). Inteligencia artificial, discurso digital y lenguaje: La lingüística en la era de las nuevas tecnologías. *Lengua y Sociedad, 23*(2), 577-598. https://doi.org/10.15381/lengsoc.v23i2.29773

Martínez, C. D., García, P. D. and Sustaeta, P. N. (2020). Sesgos de género ocultos en los macrodatos y revelados mediante redes neurales: ¿hombre es a mujer como trabajo es a madre? *Reis: Revista Española de Investigaciones Sociológicas*, *172*, 41-59. https://doi.org/10.5477/cis/reis.172.41

Matus, A. (2020). The end of the Chilean fantasy. *Dissent, 67*(2), 97-105. https://dx.doi.org/10.1353/dss.2020.0020.

Mellado, C. and Scherman, A. (2020). Mapping source diversity across Chilean news platforms and mediums. *Journalism Practice, 15*(7), 974-993. https://doi.org/10.1080/17512786.2020.1759125

Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. and Liu, H. (2018). Identifying Framing Bias in Online News. ACM Trans. *Soc. Comput. 1*(2), 1-18. https://doi.org/10.1145/3204948

Nash, C. (2017). Dictadura, transición y corrupción. Algunas lecciones del caso chileno. *Revista Electrónica de Investigación y Asesoría Jurídica-REDIAJ, 1*(8), 353-408. https://acortar.link/1ykPo9

Navia, P. and Ulriksen, C. (2017). Tuiteo, luego voto. El efecto del consumo de medios de comunicación y uso de redes sociales en la participación electoral en Chile en 2009 y 2013. *Cuadernos.Info*, (40), 71-88. https://dx.doi.org/10.7764/cdi.40.1049

Oliva, S. (2017). *Introducción a Pandas y Jupyter Notebook de Python*. https://bit.ly/4f9gjWA

O'Neil, C. (2017). *Armas de Destrucción Matemática. Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Capitán Swing.

Otero, V. (2020). *Political media's bias, in a single chart*. Ad Fontes Media.

Palau, A. and Davesa, F. (2013). El impacto de la cobertura mediática de la corrupción en la opinión pública española. *Revista Española de Investigaciones Sociológicas, 144*(1), 97-126. https://doi.org/10.5477/cis/reis.144.97

Pastrana, A. (2019). Estudio sobre la corrupción en América Latina. *Revista mexicana de opinión pública,* (27), 13- 40. https://doi.org/10.22201/fcpys.24484911e.2019.27.68726

Pereyra, S. (2012). La política de los escándalos de corrupción desde los años 90. *Desarrollo Económico, 52*(206), 255-284. https://bit.ly/4lFjIPp

Rodríguez-Virgili, J., López-Escobar, E. and Tolsá A. (2011). La percepción pública de los políticos, los partidos y la política, y uso de medios de comunicación. *Comunicación y Sociedad, 24*(2), 7-39. https://bit.ly/4nQbjKe

Sádaba, T., Rodriguez, J. and Bartolomé, M. (2012). Propuesta de sistematización de la teoría del framing para el estudio y praxis de la comunicación política. *Observatorio Journal, 6*(2), 109-126. https://doi.org/10.15847/obsOBS622012540

Saez-Trumper, D., Castillo, C. and Lalmas, M. (2013). Social media news communities: Gatekeeping, coverage, and statement bias. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. *ACM*, 1679-1684. https://doi.org/10.1145/2505515.2505623

Sánchez-Holgado, P., Martín Merino, M. and Blanco Herrero, D. (2020). Del data-driven al data-feeling: análisis de sentimiento en tiempo real de mensajes en español sobre divulgación científica usando técnicas de aprendizaje automático. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones", 13*(1), 35-58. https://doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.7691

Serrano, V. and Salvat, G. (2015). *Tras el ensayo digital. Una aproximación interdisciplinar a la sociedad de la información*. Ediciones UACh.

Sixto-García, J., Soengas-Pérez, X., Rodríguez-Vázquez, A.-I., Vázquez-Herrero, J. and López-García, X. (2021). Percepción social del periodismo en España. *AdComunica*, (22), 191-210. https://doi.org/10.6035/2174-0992.2021.22.11

Sola-Morales, S. and Rivera Gallardo, R. (2017). El tratamiento periodístico sobre la corrupción política. Análisis comparado del caso SQM en Chile y del caso de Bárcenas en España. *Estudios sobre el Mensaje Periodístico, 23*(1), 647-662. https://doi.org/10.5209/ESMP.55619

Sutriawan, S., Rustad, S., Shidik, G. F., Pujiono, P. and Muljono, M. (2024). Review of ambiguity problem in text summarization using hybrid ACA and SLR. *Intelligent Systems with Applications*, *22*, 200360.

Vernier, M., Cárcamo Ulloa, L. and Scheihing García, E. (2017). Diagnóstico de la estrategia editorial de medios informativos chilenos en Twitter mediante un clasificador de noticias automatizado. *Revista Austral de Ciencias Sociales,* (30), 183-201. https://doi.org/10.4206/rev.austral.cienc.soc.2016.n30-09

Vernier, M., Cárcamo Ulloa, L.y Scheihing García, E. (2018). Pensamiento crítico de los jóvenes ciudadanos frente a las noticias en Chile. *Comunicar, 26*(54), 101-110. https://doi.org/10.3916/C54-2018-10

Vernier, M. and Ruiz, F. (2023). *Manual de usuario. Sophia search: software de recopilación automática e indexación de documentos web para la investigación en comunicación y territorios.* https://bit.ly/4nZzaXW

Vidal-Sepúlveda, M. and Olivares-Rodríguez, C.. (2021). ¿Quién elige mi dieta informativa? La datificación automatizada de personas en el ambiente digital. *Información, cultura y sociedad*, (44), 143-154. https://doi.org/10.34096/ics.i44.8627

Waisbord, S. (2020). ¿Es válido atribuir la polarización política a la comunicación digital? Sobre burbujas, plataformas y polarización afectiva. *SAAP, 14*(2), 249-279. https://bit.ly/44YnbBe

Wilenman, J. and Feddersen, M. (2023). Contesting authority in the crisis of neoliberalism: The Chilean Spring and the mobilization of human rights frames. In Steven A. Boutcher, Corey S. Shdaimah y Michael W. andarbrough (eds.), *Research Handbook on Law, Movements and Social Change*. Edward Elgar. https://doi.org/10.4337/9781789907674.00036

Xu, S., Li, Z., Mei, K. and Zhang, Y. (2024). Core: Llm as interpreter for natural language programming, pseudo-code programming, and flow programming of ai agents. *arXiv preprint arXiv:2405.06907*, *2*.

Zalaquett, J. (2011). Conflictos de intereses: normas y conceptos. *Anuario de Derechos Humanos* (7), 179-189. https://doi.org/10.5354/adh.v0i7.17363

Zunino, E. and Focás, B. (2021). La corrupción como problema público en la televisión argentina: interacción entre medios y audiencias. *Anagramas Rumbos y Sentidos de la Comunicación, 20*(39), 129-155. https://doi.org/10.22395/angr.v20n39a6

Declaration of Authorship - CRediT Taxonomy	
Author	**Contributions**
Camila Rojas-Sánchez	Conceptualization, Data curation, Formal analysis, Research, Visualization, Original draft writing, Writing-revision, and Editing.
Luis Cárcamo-Ulloa	Conceptualization, Fundraising, Supervision, Writing-review and editing.
Matthieu Vernier	Data curation, Methodology, Software, Validation.

Declaration on the Use of Artificial Intelligence
The authors DECLARE that the article ***Web Coverage of Corruption in Chile: Automated Analysis with Artificial Intelligence and Linguistics*** was not supported by Artificial Intelligence (AI).