UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE CIÊNCIAS

DEPARTAMENTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

IVANDRO CLAUDINO DE SÁ

DIGITAL LIGHTHOUSE: A PLATFORM FOR MONITORING MISINFORMATION IN WHATSAPP PUBLIC GROUPS

FORTALEZA

2021

IVANDRO CLAUDINO DE SÁ

DIGITAL LIGHTHOUSE: A PLATFORM FOR MONITORING MISINFORMATION IN WHATSAPP PUBLIC GROUPS

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. José Maria da Silva Monteiro Filho

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S112d Sá, Ivandro Claudino de.
Digital lighthouse : a platform for monitoring misinformation in WhatsApp public groups / Ivandro Claudino de Sá. – 2021.
85 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2021.
Orientação: Prof. Dr. José Maria da Silva Monteiro Filho.

1. Sistemas de informação. 2. Desinformação. 3. Mídias sociais. 4. Processamento de linguagem natural. I. Título.

CDD 005

IVANDRO CLAUDINO DE SÁ

DIGITAL LIGHTHOUSE: A PLATFORM FOR MONITORING MISINFORMATION IN WHATSAPP PUBLIC GROUPS

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Ciência da Computação

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. José Maria da Silva Monteiro Filho (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Angelo Roncalli Alencar Brayner
Universidade Federal do Ceará (UFC)

Prof. Dr. José Gilvan Rodrigues Maia
Universidade Federal do Ceará (UFC)

À todos aqueles que fazem o extraordinário se tornar cotidiano.

ACKNOWLEDGEMENTS

Elaborar o agradecimento é um desafio, pois demonstrar nossa gratidão adentra pelos nossos valores e pela nossa história. Humildemente recorremos às nossas lembranças e queremos destacar todos aqueles que caminharam conosco em busca do conhecimento: dos artistas e intelectuais que nos inspiram, até nossos amigos e familiares. Quero agradecer à todos eles!

Agradeço ao povo latino americano, o MST, o Centro Frei Humberto, Armazém do Campo, Cáritas, a Rede de Médicos e Médicas Populares e à todos os companheiros e companheiras que lutam por um mundo melhor e mais justo.

Agradeço aos colegas do Serpro, da DE301 e todas as equipes que tive o prazer de participar.

À todos os que fazem da Universidade Federal do Ceará este centro de excelência, destacando o MDCC, o Laboratório Arida e os que participam do projeto Farol Digital, entre eles o Leonardo, San Diego, Barão, Patrícia, Pedro Mourão, Thiago, Lucas e demais.

Agradeço a André, Jorge, Virgínia, Cristina, Bianca, Gislei e todos os que fazem a Fiocruz e o PSAT.

Agradeço aos meus amigos que socializam sonhos e aperreios: Rangel e Henrique. E os companheiros Flávio, José Eduardo, Mauda, Lourival, Paulo Coelho, Pedro Aníbal, Victor, Daniel, Suderland, Valdijan, Alexandre, todos da Alternativa Progressista, Vlader, Alan, Bruno, Dmitri, Rafael, Gleyson, João Mendes e Andrew.

De forma mais do que especial quero agradecer à minha esposa Ana Paula, meu amor e companheira que há anos partilha sonhos, desejos e esperanças. Aos meus filhos Rafael, Miguel e Theo, por serem esses lindos seres humanos, minha inesgotável fonte de inspiração. Minha mãe Inez e meu pai João Bosco, por todo o carinho, incentivo e exemplo de vida. Ao meu irmão Vinícius, meus sobrinhos, minha cunhada Aline e toda sua família. Ao meu sogro José Neto e a minha sogra Sônia que mesmo em memória segue contribuindo na caminhada. A Netinho e família, Cláudia e Família. Agradeço aos meus avós Silvia e Joãozinho. À Mercês, Benício, Fernando, Érica, Evandro, Silvinha e família, Cristina, Pedro Jorge e todos os meus tios e primos.

Por fim, agradeço o Wellington por todo seu compromisso e companheirismo e o meu orientador José Maria, por acreditar e principalmente ser esse grande ser humano repleto de sabedoria, valores e responsabilidade social.

”Não sois máquina! Homens é que sois! E com o amor da humanidade em vossas almas! Não odieis!”

(Charlie Chaplin)

RESUMO

Atualmente, a disseminação em larga escala de desinformação por meio das mídias sociais tornou-se uma questão crítica, prejudicando a estabilidade social, a democracia e a saúde pública. No Brasil, 48% da população utiliza o WhatsApp para obter notícias. Assim, muitos grupos têm usado esse aplicativo de mensagens instantâneas para espalhar desinformação, especialmente como parte de campanhas políticas ou ideológicas articuladas. Nesse contexto, o WhatsApp oferece um recurso importante: os grupos públicos. Esses grupos são bastante adequados para a disseminação de desinformação. Portanto, o desenvolvimento de plataformas de software para monitorar a disseminação de desinformação em grupos públicos do WhatsApp tornou-se um campo de grande interesse na academia, no governo e na indústria. Neste trabalho apresentamos uma plataforma, denominada Farol Digital, para localizar, integrar, analisar e visualizar os conteúdos que trafegam nos grupos públicos do WhatsApp. Para avaliar nossa metodologia, construímos três conjuntos de dados diferentes. Esperamos que a nossa plataforma possa ajudar jornalistas e pesquisadores a entender a propagação da desinformação no Brasil.

Palavras-chave: Detecção de Desinformação. Processamento de Linguagem Natural. WhatsApp. Mídias Sociais.

ABSTRACT

Actually, the large-scale dissemination of misinformation through social media has become a critical issue, harming social stability, democracy, and public health. In Brazil, 48% of the population use WhatsApp to get news. So, many groups have been used this instant messaging application in order to spread misinformation, especially as part of articulated political or ideological campaigns. In this context, WhatsApp provides an important feature: the public groups. These groups are so suitable for misinformation dissemination. Thus, developing software frameworks to monitor the misinformation spreading in WhatsApp public groups has become a field of high interest both in academia, government and industry. In this work we present a platform, called Digital Lighthouse, for finding, gathering, analyzing, and visualize public groups in WhatsApp. In order to evaluate our methodology, we built three different datasets. We hope that our platform can help journalists and researchers to understand the misinformation propagation in Brazil.

Keywords: Misinformation Detection. Natural Language Processing. WhatsApp. Social Media.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

1.1 Motivation

In the last years, the popularity of instant messaging applications has contributed to the spread of misinformation. Through these systems, misinformation can deceive thousands of people in a short time (due to their appealing nature) and cause significant harm to individuals or society. Such applications allow content to be spread without editorial judgment. In this context, misinformation has been used to change political scenarios, to contribute to the spread of diseases, and even to cause deaths (VOSOUGHI *et al.*, 2018; GUO *et al.*, 2019; SU *et al.*, 2020).

It is important to highlight that misinformation is a wide concept that can be defined in a general way as misrepresented information, including fabricated, misleading, false, fake, deceptive, or distorted information (SU *et al.*, 2020). This broad definition covers a variety of concepts such as fake news (LAZER *et al.*, 2018), rumor(SHU *et al.*, 2017), deception (MAALEJ, 2001) and hoaxes. However, despite describing intentionally misleading information written as journalistic news, the term fake news has become very present in popular culture. It sometimes is used as a misinformation synonym (GUO *et al.*, 2019).

The WhatsApp instant messaging application, created in 2009 by Brian Acton and Jan Koum, is very popular in Brazil, with more than 120 million users in about 210 million people. WhatsApp makes it possible to instantly share different media types, such as images, audio, and videos. In Brazil, 48% of the population use WhatsApp to get, share and discuss news (RESENDE *et al.*, 2019). In December 2019, a survey carried out by the Brazilian Chamber of Deputies and the Senate with 2,400 people concluded that 79% of people use the WhastApp as the main source of information[1]. In February 2020, the Panorama Mobile Time/Opinion Box survey on mobile messaging in Brazil revealed that WhatsApp is installed on 99% of Brazilian smartphones. Among users of the application, 98% said they access it every day or almost every day [2].

[1] DATASENADO. Redes Sociais, Notícias Falsas e Privacidade de Dados na Internet. Brasília, nov. 2019. Available in: www2.camara.leg.br/a-camara/estruturaadm/ouvidoria/dados/pesquisa-nov-2019-relatorio-completo. Accessed in: 26 abr. 2020.

[2] SCHERMANN, Daniela. Panorama Mobile Time/Opinion Box: Mensageria no Brasil. Opinion Box, 2 mar. 2018. Available in https://blog.opinionbox.com/mensageria-no-brasil-sexta-edicao/. Accessed in: 11 mar. 2020.

WhatsApp provides an essential feature: the public groups. These public groups are accessible through invitation links published on popular websites and various social networks, such as Facebook and Twitter. Usually, they have specific topics for discussion, such as politics, soccer, and education. WhatsApp allows each public group to have a maximum of 256 members. In this way, WhatsApp public groups are very similar to social networks. Thus, public groups have been used to spread misinformation, especially as part of articulated political or ideological campaigns. Furthermore, misinformation spreads faster, more profound, and wide than legit information. Further, due to the high volume of information that we are exposed to, we have a limited ability to distinguish true information from misinformation (VOSOUGHI *et al.*, 2018; QIU *et al.*, 2017). Thus, currently, the main tool used to spread misinformation is WhatsApp. A survey carried out by the Oswaldo Cruz Foundation (Fiocruz) showed that 73.7% of the false news about the new coronavirus circulated through WhatsApp. Another 10.5% were published on Instagram and 15.8% on Facebook[3].

In this context, monitoring the content circulating in public WhatsApp groups is a fundamental task to identify and understand the spread of misinformation and get insights to address this problem. However, collecting a database of messages already in circulation in WhatsApp public groups is a complex task. To fill this gap, we built the Digital Lighthouse, an entire platform for finding, gathering, analyzing, and visualize public groups in WhatsApp. Early detection of misinformation could prevent its spread, thus reducing its damage. To evaluate the proposed platform, we used it to build three different WhatsApp' messages datasets, covering relevant themes such as the Brazilian general elections campaign in 2018, the COVID-19 pandemic, and the vaccine COVID-19.

1.2 Objective and Scope

In this work, we are interested in a particular type of disinformation: more precisely, the disinformation contained in messages spreading in WhatsApp public groups. The main question investigated in this work is: is it possible to conceive a mechanism that makes it possible to monitor the misinformation circulating in WhatsApp public groups? Going even further, this solution could:

[3] Available in: https://portal.fiocruz.br/noticia/pesquisa-revela-dados-sobre-fake-news-relacionadas-covid-19. Accessed in: 27 abr. 2020.

- Be compliant with current legislation about the privacy of personal data?
- Could this mechanism runs continuously (or periodically) and automatically?
- Help to identify messages that contain misinformation?
- Assist in the characterization of messages that contain misinformation?
- Assist in the identification of misinformation super-spreader users?
- Assist in the characterization of users who spread misinformation?
- Assist in the identification of sites (or URLs) that operate as a source of misinformation?
- Follow the most commented terms (or topics) in the monitored groups?
- Be made available publicly and free of charge?

This research shows that it is possible to devise a platform capable of continuously monitoring the misinformation that spreads in WhatsApp public groups.

1.3 Main Contributions

The main contributions of this research are:

1. An architecture for monitoring the misinformation that spreads in public WhatsApp groups;
2. An implementation of the designed architecture;
3. A case study involving three different WhatsApp' messages datasets, covering relevant themes such as the Brazilian general elections campaign in 2018, the COVID-19 pandemic, and the vaccine COVID-19.

1.4 Initial Considerations

Personal data protection is a problem that has been widely discussed in several countries. Many countries created specific laws and regulations to protect fundamental rights and privacy, such as the General Data Protection Law (Lei Geral de Proteção de Dados - LGPD) from Brazil. However, the LGPD does not apply for journalistic or academic purposes. For scientific research purposes, the law only recommends that personal data be anonymized whenever possible.

On the other hand, WhatsApp's Privacy Policy states that all user profile information such as username and cell phone is available to any other user who has an interaction through the app, whether it be a private conversation or in a public group. Thus, the information circulating in a given group is available to all group members, whether public or private. Therefore, we do

not violate WhatsApp's privacy policy. However, to maintain users' privacy, the data related to the user's name and cell phone number were duly anonymized, and only the direct dialing codes were kept for geographic analysis.

We hope that the Digital Lighthouse platform can help journalists and researchers understand Brazil's misinformation propagation. Besides, the proposed platform can be used to build misinformation detection systems, which aim to assist users in detecting and filtering out deceptive news. Additionally, the proposed platform aims to inform and anticipate communicators about the type of information shared by Brazilians in public groups. The idea is to provide access to the platform for researchers, journalists, and public security agents, who can check facts and information shared in public groups.

More specifically, the Digital Lighthouse platform can be used as a base for:

- Automatic analysis of the content of the most recurring messages to verify the misinformation that needs to be addressed by public authorities and society in general;
- Analysis of the misinformation spread pattern, to identify the virtualization logic of these content;
- Automatic classification of misinformation;
- Automatic identification of hate speech;
- Automatic identification of fake profiles and bots;
- Automatic search for true content to counter given misinformation.

1.5 Text Organization

The remainder of this dissertation is organized as follows.

- In Chapter 2, the main approaches found in the literature for monitoring and classifying misinformation in social networks will be discussed. A detailed comparative analysis of these approaches is also presented;
- Chapter 3 presents the Digital Lighthouse platform, an entire solution to monitor the misinformation spreading in WhatsApp public groups. Initially, we detail the operation of the main components of the proposed platform architecture. Next, we discuss an implementation of the designed architecture;
- Chapter 4 details a case study involving three different WhatsApp' messages datasets, covering relevant themes such as the Brazilian general elections campaign in 2018, the COVID-19 pandemic, and the vaccine COVID-19. The case study was performed using

the Digital Lighthouse platform;

– In Chapter 5, we point several essential considerations about the compliance with data privacy laws (such as LGPD) and the WhatsApp privacy policy. Besides, we discuss some threats to validity that impact this research;
– Finally, Chapter 6 concludes this work, evaluating the results obtained, the difficulties encountered, and pointing out future work directions.

2 RELATED WORK

Several works attempt to detect misinformation in different languages and social networks. Most of them use texts in English or Chinese languages, extracted from Websites or the Twitter platform. Despite the large number of works investigating the misinformation detection problem, few searches for suitable solutions for the Brazilian Portuguese language (PT-BR) (MONTEIRO *et al.*, 2018; SILVA *et al.*, 2020).

However, it is essential to highlight that WhatsApp is unique in several ways relative to other social media platforms. WhatsApp was developed to allow users to privately and freely send messages to each other through their smartphones. A specific aspect of WhatsApp messaging is the public groups. These are openly accessible groups, frequently publicized on well-known websites, and typically themed around particular topics. Another difference between WhatsApp and other social networks is that membership registration is done exclusively through one's phone number. It is worth mentioning that texts extracted from WhatsApp are pretty different from those collected through Websites, fact-checkers, or other kinds of social media platforms, such as Twitter. WhatsApp messages include conversation, opinions, humorous and satirical texts, prayers, commercial offers, news, short texts, emojis, and others.

Thus, despite the scientific community's efforts, there is still a need for monitoring and identifying misinformation in WhatsApp messages, mainly in Portuguese. The paper presented in (GARIMELLA; TYSON, 2018) is a seminal work in collecting and analyzing WhatsApp' messages. The authors built a dataset by crawling 178 public groups containing 45K users and 454K messages from different countries and languages, such as India, Pakistan, Russia, Brazil, and Colombia.

In (GAGLANI *et al.*, 2020), the authors contextualize the problem of spreading fake news on WhatsApp, especially in India and Brazil, and proposes a strategy for the automatic detection of fake news. A total of 10 public groups were scrapped for one week to get 1000 multilingual messages. After cleaning the data, the multilingual data was translated into English by employing the google translate API.

In (RESENDE *et al.*, 2018), the authors presented a system for gathering, analyzing, and visualize public groups in WhatsApp. Besides describing their methodology, the authors also provide a brief characterization of the 169.154 messages shared by 6,314 users in 127 public groups to help journalists and researchers understand the repercussion of events related to the 2018 Brazilian elections.

In the study presented in (MACHADO *et al.*, 2019), the authors collected and analyzed 298,892 WhatsApp' messages, from 130 public groups, in the period leading up to the two rounds of the 2018 Brazilian presidential elections. Further, they examined a sample of 200 videos and images extracted from these WhatsApp messages and developed a new typology to classify this media content.

In (RESENDE *et al.*, 2019), the authors analyzed different aspects of WhatsApp messages from public political-oriented groups. The messages were collected during major social events in Brazil: a national truck drivers' strike and the Brazilian presidential campaign. The authors analyzed the types of content shared within such groups and the network structures that emerge from user interactions. Besides, they identified misinformation among the shared images using labels provided by journalists and by an automatic procedure based on Google searches. However, none of these works provides an entire public platform for finding, gathering, analyzing, and visualizing public groups in WhatsApp.

Other works propose classifiers to detect misinformation automatically. In (MONTEIRO *et al.*, 2018), the authors presented the first Fake News' corpus in Brazilian Portuguese (PT-BR), called Fake.Br. This corpus was built manually, collecting Fake News from Web sites. The authors also evaluated some machine learning classifiers (Naıve-Bayes, Random Forest, and Multilayer Perceptron). Next, the work presented in (SILVA *et al.*, 2020) investigated the use of different features and algorithms to detect fake news, exploring the Fake.Br corpus.

In (FAUSTINI; COVÕES, 2019), the authors evaluated using the One-Class Classification (OCC) technique to detect fake news automatically. OCC is a two-phases process. First, a model is trained only with fake news data instead of other supervised learning algorithms which use fake and true classes. So, during the test phase, the model must distinguish between what is fake and what is not, based only on the learned fake news data's characteristics. In the experiments, three different datasets were explored: Fake.Br, a Twitter corpus, and a small WhatsApp corpus (with only 177 messages).

In (SHU *et al.*, 2018), the authors investigated the use of complex networks to detect and mitigate fake news on social media. During fake news dissemination, different entities can be categorized into content, social and temporal dimensions. These dimensions have mutual relations and dependencies. So, fake news dissemination has inherent network properties.

In (SHU *et al.*, 2019), the authors explored user profiles to detect fake news. They argue that there are correlations between malicious accounts and fake news. Thus, user fea-

tures, content features, and social network features can be used together to improve fake news classifiers.

In this same way, the paper presented in (HAMDI *et al.*, 2020) proposed a hybrid approach that explores features from the user profile and his social graph (Twitter followers/followees graph) to detect fake news.

In (ZHANG; HARA, 2020), the authors propose a probabilistic model for malicious user and rumor detection (MURD). In contrast to existing approaches, their model not only the behavior but also the intention when a user retweets a rumor or a true story. This approach helps us capture user maliciousness and rumor veracity, especially when the stories are complex and confusing.

Fact-checking requires a ground of contextualized claims labeled with their truth values. In (TCHECHMEDJIEV *et al.*, 2019), the authors presented a public, large-scale, up-to-date, and queryable corpus of structured information about claims and related metadata called ClaimsKG, a knowledge graph (KG) of fact-checked claims. The proposed KG supports structured queries about their truth values, authors, dates, journalistic reviews, and other kinds of metadata. ClaimsKG is generated through a semi-automated pipeline, which harvests data from popular fact-checking websites regularly, annotates claims with related entities from DBpedia, and lifts the data to RDF an RDF/S model that makes use of established vocabularies. ClaimsKG consists of 28,383 claims published since 1996, amounting to 6,606,032 triples.

Melo et al. (MELO *et al.*, 2019) evaluated the dynamics of the spread of (mis)information on a network of public WhatsApp groups. The authors investigated the mass communication features of public chat groups and the forwarding/broadcasting of messages. More specifically, they tried to answer how the forwarding tools contribute to the virality of (mis)information and whether system limitations are capable of preventing the spread of content. Finally, they proposed some hints on how the problem of large-scale dissemination can be countered.

Reis et al. (REIS *et al.*, 2020a) observed that misinformation had been shared on WhatsApp public groups even after popular fact-checking agencies already fact-checked them. They posit that such misinformation content could be prevented if WhatsApp had a means to flag already fact-checked content. To this end, the authors proposed an architecture that WhatsApp could implement to counter such misinformation.

Reis et al. (REIS *et al.*, 2020b) performed an extensive data collection from a large set of WhatsApp public groups and fact-checking agency websites. The authors built two

datasets, composed of 135 and 897 images containing misinformation from Brazil and India. These images circulated on hundreds of publicly accessible WhatsApp groups1 around the 2018 Brazilian national elections and the 2019 Indian national elections and were fact-checked by well-known fact-checking agencies. These datasets are publicly available.

In (VASCONCELOS *et al.*, 2020), the authors studied YouTube videos shared by political-oriented public groups on Whatsapp for a month during the COVID-19 pandemic. Through a careful analysis of the topical distribution and the lexicon present in the videos shared, the authors shed light on the COVID-19 debate happening in these groups. Moreover, the authors compare COVID-19 related videos with other political videos being shared in these groups. They observed that videos that discuss political themes have more emotional attributes as well as topics related to typical right-wing concerns such as family, work, and religion than videos discussing both pandemic and politics.

In (MONDAL *et al.*, 2020), the authors released a public dataset of more than 89 million posts from an anonymous social media site called Whisper. The dataset was collected via continuously crawling the "latest" section of Whisper from June 2014 to June 2016. This is the first large dataset from an anonymous social media, and this dataset has the potential to provide a solid ground to understand the needs and pitfalls of anonymous communication.

Maros et al. (MAROS *et al.*, 2020) presented a methodology to analyze audio messages shared in WhatsApp groups, characterizing content properties (e.g, topics and language characteristics), their propagation dynamics and the impact of different types of audios (e.g., speech versus music) on such dynamics.

Table 1 presents a comparative analysis between datasets that use WhatsApp messages in Brazilian Portuguese. This table details the datasets described by the related literature and those collected by the Digital Lighthouse platform. The Work column identifies the authors of the works. The Labeled column tells you whether messages were labeled, not labeled or partially labeled. The Total of Text Messages column highlights the number of text messages analyzed. The Group column reports the total number of groups where messages were collected. The Users column details the total number of users who are part of these groups. The Monitor column indicates whether the research has a monitoring platform or whether data was collected by another means. Finally, the last column indicates whether the dataset or monitoring system is publicly available.

Table 1 – Datasets of WhatsApp Messages in Brazilian Portuguese. Hyphen (-) means that the information could not be found in the work.

Work	**Labeled**	**Total of Text Messages**	**Groups**	**Users**	**Monitor**	**Publicly Available**
(RESENDE *et al.*, 2018)	No	169,154	127	6,314	Yes	No
(RESENDE *et al.*, 2019)/ Truck Drivers' Strike	No	95,424	141	5,272	No	No
(RESENDE *et al.*, 2019)/ Election Campaign	No	591,162	136	18,725	No	No
(FAUSTINI; COVÕES, 2019)	Yes	177	-	-	No	Yes
(MACHADO *et al.*, 2019)	No	298,892	130	-	No	No
Digital Lighthouse/ Election Campaign	Partially	282,601	59	5,364	Yes	Yes
Digital Lighthouse/ COVID-19 Pandemic	Partially	228,061	59	10,495	Yes	Yes
Digital Lighthouse/ COVID-19 Vaccine	No	16,056	175	1,857	Yes	Yes

3 THE DIGITAL LIGHTHOUSE PLATFORM

This chapter will present the main components of the Digital Lighthouse platform, which aims to find, gather, analyze, and visualize public groups in WhatsApp. The proposed platform architecture comprises four modules, as illustrated in Figure 1. Next, we will discuss in detail each one of these components.

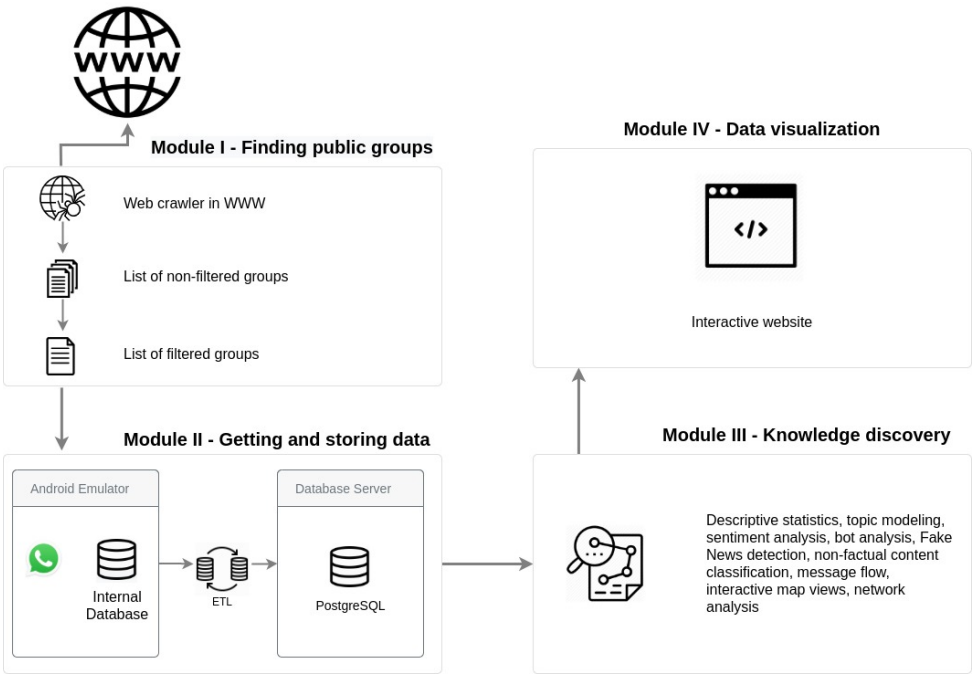

Figure 1 – The Digital Lighthouse Platform Architecture

3.1 Module I: Finding Public Groups

Whatsapp allows you to join public groups through the use of links without the authorization of the administrator. The links are URLs containing the domain 'chat.whatsapp.com' and a group identification code, allowing them to be publicized through websites or social networks. In this way, groups can be found through simple queries on search engines like Google, Bing, or Yahoo, or simply by accessing sites created for this specific purpose. This work used two different strategies for finding public groups: accessing specific sites and using the Google search engine.

So, the first strategy used to find WhatsApp public groups' invitation links was to search for well-known sites for containing these links. Table 2 illustrates some specific sites used in our first strategy.

Table 2 – Specific Sites used to Finding Invitation Links.

Site
https://zapbolsonaro.com/2019/03/29/grupos-zap-bolsonaro/
https://gruposdezap.com/
https://www.grupowhats.online/
https://zapzapgrupos.com/
https://azpop.com.br/grupos
https://gruposwhats.app/

The second strategy to find public groups uses a web crawler that sends queries to Google search engine. The main idea is to build a query containing the text 'chat.whatsapp.com' (WhatsApp domain) together with a set of keywords. This strategy will be detailed next. However, it is essential to highlight that if the group owner (or spreader) uses a URL shortener, the new URL will not contain the text 'chat.whatsapp.com'. So, this action can embarrass the operation of the Finding Public Groups Module.

3.1.1 Web Crawler

A web crawler (sometimes called spider or spiderbot) can be defined as a program or software which systematically browses the World Wide Web and, typically, download web pages in a methodical and automated manner. To find WhatsApp public groups, we develop a web crawler using the Python programming language through the Google search engine. Its task is to build queries, send them to the search engine and receive the result. To set up a particular query, the Crawler receives a series of input parameters, such as: the WhatsApp domain, a set of keywords, and the used language. Listing 1 illustrates de crawler setup. After a given query is executed, the Crawler receives metadata, including references to the web pages where the invite links were found. These web page links are stored in a file called search_links.csv. Figure 2 shows an example of the file search_links.csv.

Source Code 1 – Crawler Setup

```
1 tags_input = ['covid','coronavirus']
2 domain = 'chat.whatsapp.com'
3 language = 'portuguese'
4 crawler.start(tags_input, domain, language)
```

id	title	link	grepped	date-time
1	Ministério de Fé & Política -Porto Alegre - पोस्ट | F	https://hi-in.facebook.com/mfppoa/posts/?ref=page_internal	True	Sat Apr 10 17:05:26 2021
2	INTRODUÇÃO À CITOLOGIA - Citologia | Biologi	https://kzhome.info/crone/s2NpoYeKq42bho0/introdu-o-citologia-citologia-biologia-com-samuel-cunha	True	Sat Apr 10 17:05:26 2021
3	Uma ponte sob o Atlântico e outra sobre o Guaíb	https://bgpost.info/post/q9el3JjbpqJjnHU/uma-ponte-sob-o-atl-ntico-e-outra-sobre-o-gua-ba	True	Sat Apr 10 17:05:26 2021
4	#COVID19 GRUPO DE... - Ministério das Relaçõ	https://www.facebook.com/ItamaratyGovBr/posts/covid19-grupo-de-whatsapp-para-divulga%C3%A7%C3%A3o-de-	True	Sat Apr 10 17:07:55 2021
5	ZAP CoVida e Telegram: Rede amplia canais de	https://redecovida.org/2020/06/10/zap-covida-o-canal-de-informacoes-da-rede-covida-no-whatsapp/	True	Sat Apr 10 17:07:55 2021
6	Grupo de WhatsApp dos PAIS – Campus Assis C	https://assis.ifpr.edu.br/grupo-de-whatsapp-dos-pais/	True	Sat Apr 10 17:07:55 2021
7	"Saúde na palma da mão" tem grande procura e n	https://www.paranavai.pr.gov.br/noticias/1412782	True	Sat Apr 10 17:07:55 2021
8	Coronavirus - Townhill Surgery	https://www.townhillsurgery.nhs.uk/coronavirus	True	Sat Apr 10 17:07:55 2021
9	UnB on Twitter: "O Programa de Educação Tutori	https://twitter.com/unb_oficial/status/1247263683304206337	True	Sat Apr 10 17:07:55 2021
10	Grupo virtual de acolhimento de estudantes está	https://www.ifpb.edu.br/noticias/2020/03/grupo-virtual-de-acolhimento-de-estudantes-esta-no-ar	True	Sat Apr 10 17:07:55 2021
11	PROBEX da UFCG – Campus Cuité-PB convida	http://www.ces.ufcg.edu.br/portal/index.php/noticias/2094-probex-da-ufcg-campus-cuite-pb-convida-a-comunidade-	True	Sat Apr 10 17:07:55 2021
12	Covid -19 em Serro - Prefeitura de Serro - MG	https://www.serro.mg.gov.br/portal/servicos/203/covid--19-em-serro/	True	Sat Apr 10 17:07:55 2021
13	CANAL SUBCULT PARA ARTISTAS CABISTAS	https://www.arraial.rj.gov.br/portal/noticias/0/3/1199/canal-subcult-para-artistas-cabistas-/	True	Sat Apr 10 17:08:01 2021
14	Separados agora, juntos para sempre. Novo. - IF	https://prc.ifsp.edu.br/index.php/ultimas-noticias/1930-separados-agora-juntos-para-sempre	True	Sat Apr 10 17:08:01 2021
15	Prefeitura lança serviço de primeiros cuidados ps	https://linhares.es.gov.br/2020/04/06/prefeitura-lanca-servico-de-primeiros-cuidados-psicologicos-para-enfrentar-a-	True	Sat Apr 10 17:08:01 2021
16	Estudantes da UFSM-FW podem participar de gr	https://www.ufsm.br/unidades-universitarias/frederico-westphalen/2020/11/11/estudantes-da-ufsm-fw-podem-partic	True	Sat Apr 10 17:08:01 2021
17	Ações da UFPel – UFPEL COVID-19	https://wp.ufpel.edu.br/covid19/acoes_da_ufpel/	True	Sat Apr 10 17:08:01 2021
18	Boletim Epidemiológico COVID-19	https://caparao.mg.gov.br/coronavirus/be/12/2872-301220/file	True	Sat Apr 10 17:08:01 2021
19	Boletim Epidemiológico COVID-19	https://caparao.mg.gov.br/coronavirus/be/2021-1/01/2894-060121/file	True	Sat Apr 10 17:08:01 2021
20	#SolidarizaGoiânia: saiba como ajudar quem mais	https://sagresonline.com.br/solidarizagoiania-saiba-como-ajudar-quem-mais-precisa-durante-a-pandemia-na-capital/	True	Sat Apr 10 17:08:01 2021
21	Atendimentos no Câmpus Cuiabá - UFMT	https://www.ufmt.br/unidade/covid19/pagina/atendimento-ufmt/3225	True	Sat Apr 10 17:08:01 2021
22	Ações de Extensão no Enfrentamento do Covid-1	http://www.uftm.edu.br/proext/acoes-extensao-covid19	True	Sat Apr 10 17:08:01 2021
23	Boletim Médico Diário – - - HRMS	https://www.hospitalregional.ms.gov.br/boletim-medico-diario/	True	Sat Apr 10 17:08:08 2021
24	Prevenção à Covid-19 - Instituto Legado	https://institutolegado.org/blog/instituto-legado-entra-para-movimento-de-prevencao-a-covid-19/	True	Sat Apr 10 17:08:08 2021
25	Butantan tira dúvida - CRF-SP - Conselho Region	http://www.crfsp.org.br/noticias/11621-butantan-tira-d%C3%BAvida.html	True	Sat Apr 10 17:08:08 2021

Figure 2 – An Illustration of the File search_links.csv

For different reasons, some search engines have an anti-crawling mechanism, which can block crawlers. Bots can scrape at a very fast pace. However, making quick requests can overload the search engine. Consequently, if the crawler sends requests faster than a human possibly can, it will be blocked by the search engine. So, to avoid this mistake, our crawler sleeps programmatically for a random period of around 10 to 20 seconds between two requests. Furthermore, if the crawler uses the same IP for a certain period, it can be blocked by the search engine.So, we have used a pool of 10 different IP addresses. To choose an IP address for a certain HTTP request, we have used the clock algorithm. So, we have used a pool of 10 different IP addresses. To choose an IP address for an HTTP request, we have used the clock algorithm.

3.1.2 List of Non-filtered Groups

The next step consists of requesting each web page found previously (and stored in the search_links.csv file) and parse it seeking WhatsApp invite links. More specifically, the crawler sends a HTTP request to the URL of the webpage. Then, the search engine responds to its request by returning the content of the webpage. After, a parser will create a tree structure with the HTML content of the web page. This tree structure will help the bot to search for invite links.

Data Scraping is something that has to be done quite responsibly. The crawler has to be very cautious to don't generate negative effects on the website. Finally, the crawler produces as output a list of invite links stored in a file called group_links.csv or yet list of non-filtered groups. Figure 3 shows an example of the file group_links.csv.

id	id	id	id	id	source	date-time
1	https://chat.whatsapp.com/HJ9mrSfHfQwFXTXstZHESw	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.facebook.com/ItamaratyGovBr/posts/	Sat Apr 10 17:09:12 2021
2	https://chat.whatsapp.com/Ez2rKRAJKx8D4NSr2bfjeE	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://redecovida.org/2020/06/10/zap-covida-o-can	Sat Apr 10 17:09:16 2021
3	https://chat.whatsapp.com/CbiSv1BpHFt2gRz1RVxyzS	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://assis.ifpr.edu.br/grupo-de-whatsapp-dos-pais	Sat Apr 10 17:09:18 2021
4	https://chat.whatsapp.com/BfIKCG0o8yd0QzegrN9wbY	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.paranavai.pr.gov.br/noticias/1412782	Sat Apr 10 17:09:21 2021
5	https://chat.whatsapp.com/HWN6hJTNvxLAF8t8SBYMaR	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.paranavai.pr.gov.br/noticias/1412782	Sat Apr 10 17:09:21 2021
6	https://chat.whatsapp.com/LIIK6jlfzgJ2wlgDGT3eBp	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
7	https://chat.whatsapp.com/BEXihAcyhFi20JIGHlMniR	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
8	https://chat.whatsapp.com/DHZ2JDVmZPvEPMAFvTDqA7	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
9	https://chat.whatsapp.com/BylbxHEVeWUKf6wu4mhncv	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
10	https://chat.whatsapp.com/DVXRpXSzUHz9rEqToEkvHa	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
11	https://chat.whatsapp.com/EodHfJS9qj93VY3Tm3hfjH	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
12	https://chat.whatsapp.com/EcvyifurXaRJszJ5hilk0D	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
13	https://chat.whatsapp.com/C7NL0fc8ZkFEnaFKvgEiJr	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.townhillsurgery.nhs.uk/coronavirus	Sat Apr 10 17:09:25 2021
14	https://chat.whatsapp.com/EsMh2qDLDtf066jRryxyc2	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.ifpb.edu.br/noticias/2020/03/grupo-virtu	Sat Apr 10 17:09:27 2021
15	https://chat.whatsapp.com/IYr3Bu8sI2r0CfV2r2mrWu	['Corona', 'Coronavírus', 'Covid 19']	None	None	http://www.ces.ufcg.edu.br/portal/index.php/noticia	Sat Apr 10 17:09:31 2021
16	https://chat.whatsapp.com/EaUiZhJIAfIDoXF8UOZKte	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.serro.mg.gov.br/portal/servicos/203/cov	Sat Apr 10 17:09:34 2021
17	https://chat.whatsapp.com/Ido0HiI7bs9IAknDOPvpLp	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://prc.ifsp.edu.br/index.php/ultimas-noticias/1	Sat Apr 10 17:09:39 2021
18	https://chat.whatsapp.com/LBh2HQlNwpw1igZtpr8KTI	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://prc.ifsp.edu.br/index.php/ultimas-noticias/1	Sat Apr 10 17:09:39 2021
19	https://chat.whatsapp.com/D7PgrGQibSD6KCVQJS57xo	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://linhares.es.gov.br/2020/04/06/prefeitura-lanc	Sat Apr 10 17:09:48 2021
20	https://chat.whatsapp.com/Hgo8Mjd8ZIE3W6rIMslZI0	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.ufsm.br/unidades-universitarias/frederi	Sat Apr 10 17:09:51 2021
21	https://chat.whatsapp.com/DSAtPwtxGs6LEtbrpaYo9I	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.ufsm.br/unidades-universitarias/frederi	Sat Apr 10 17:09:51 2021
22	https://chat.whatsapp.com/D1KzS1z2EjR0WJEFcOMZZ3Facebod	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://www.ufsm.br/unidades-universitarias/frederi	Sat Apr 10 17:09:51 2021
23	https://chat.whatsapp.com/LxzPE9S3iiU0AL4hmWPzFp	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://sagresonline.com.br/solidarizagoiania-saiba-	Sat Apr 10 17:10:06 2021
24	https://chat.whatsapp.com/EM59OknBt0f8UHb0QAtwHq	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://sagresonline.com.br/solidarizagoiania-saiba-	Sat Apr 10 17:10:06 2021
25	https://chat.whatsapp.com/JOoAdui4JhF7e39Q6pkN3Q	['Corona', 'Coronavírus', 'Covid 19']	None	None	https://sagresonline.com.br/solidarizagoiania-saiba-	Sat Apr 10 17:10:06 2021

Figure 3 – An Illustration of the File group_links.csv

It is essential to highlight that we need to take a lot of care to avoid duplicate links. Besides, humans don't perform repetitive tasks as they browse through a website. However, web scraping bots will crawl in the same pattern because they are programmed to do so. Some websites have significant anti-scraping mechanisms. They will catch the crawler and will ban it permanently. To avoid this, the crawler needs to incorporate some random clicks on the page, mouse movements, and random actions to make the crawler look like a human. Another problem is many websites change their layouts frequently for many reasons, and due to this, the scraper will fail to bring the expecting data. For this, the crawler should have a monitoring system that detects changes in their layouts and then generate alerts about this.

3.1.3 List of Filtered Groups

However, find a set of invite links is not sufficient. Some groups no longer exist, several links have been disabled, and a few groups have a minimal number of participants. Thus, it is necessary to check the status of each invite link. After this checking process, a new file, called a list of filtered groups, is generated containing only the valid links. Finally, it is possible to join public groups using a cell phone chip and a web browser automatically or manually with valid links. In this work, we manually joined the groups to don't violate WhatsApp politics.

It is essential to highlight that WhatsApp also has an anti-bot mechanism, which can block cell phone chips. So, if a bot joins public groups faster than a human possibly can, the bot's cell phone chip will be blocked by WhatsApp. To avoid this mistake, the bot needs to sleep programmatically for a random period of around 10 to 60 seconds between two join requests. This made it look more human to the WhatsApp anti-bot mechanism. Besides, it will not harm WhatsApp performance.

Additionally, it is worth mentioning that, after our bot joins a specific group, it remains there indefinitely. Furthermore, the bot acts only as an observer. That is, it does not interact with any member of the group. From the instant in which the bot joins the group, all posted messages are captured. Therefore, we have no way of getting the messages posted in the group before the bot joined it.

3.2 Module II: Getting and Storing Data

One of WhatsApp's key features is end-to-end encryption, which means the messages, audio, images, and videos sent using WhatsApp are visible only to the message's sender and receiver. However, all notices from a WhatsApp user are stored on your mobile device. So, the user can do a backup and extract the files from your mobile phone. Another alternative to obtain messages circulating in the groups in which a given user participates is to use the WhatsApp Web tool and automate the collection through Selenium, an integrated development environment for automated test scripts. A third possibility is to use an Android emulator on which the WhatsApp application is installed and the Selenium tool. The fourth option consists of access the WhatsApp internal database, where all user content is stored. This database can be accessed using SQL queries. However, it is worth mentioning that there are other alternatives to obtain messages that spread in WhatsApp groups. Although, these four approaches are the most popular.

3.2.1 Android Emulator

Unlike other social media, such as Twitter and Facebook, and due to its private chat nature, there is no public API to collect data from WhatsApp in an automated manner. Thus, monitoring WhatsApp public groups poses a technical and even ethical challenge. To tackle this issue, we take an approach similar to (GARIMELLA; TYSON, 2018; RESENDE *et al.*, 2018).

To obtain the content (text messages, audios, images, and videos) circulating in the groups in which the Digital Lighthouse joined, we used an Android emulator installed in a virtual machine. In this emulator, we had installed the WhatsApp application. Finally, we used the Selenium Web Driver to manipulate the Android emulator automatically, access the user content, and store it in a SqlLite database. In this way, it is not necessary to hack WhatsApp internal files or databases, not violating its privacy policies.

3.2.2 Database Server

To store the messages extracted from WhatsApp in an integrated manner, we used a PostgreSQL database, installed in a different virtual machine. Knowledge discovery and data visualization mechanisms will use the data stored in the PostgreSQL database. Figure 4 illustrates the PostgreSQL database schema, that is, its tables and columns. It is important to highlight that the audios, images, and videos are stored in the file system. The PostgreSQL database stores only the path to these files.

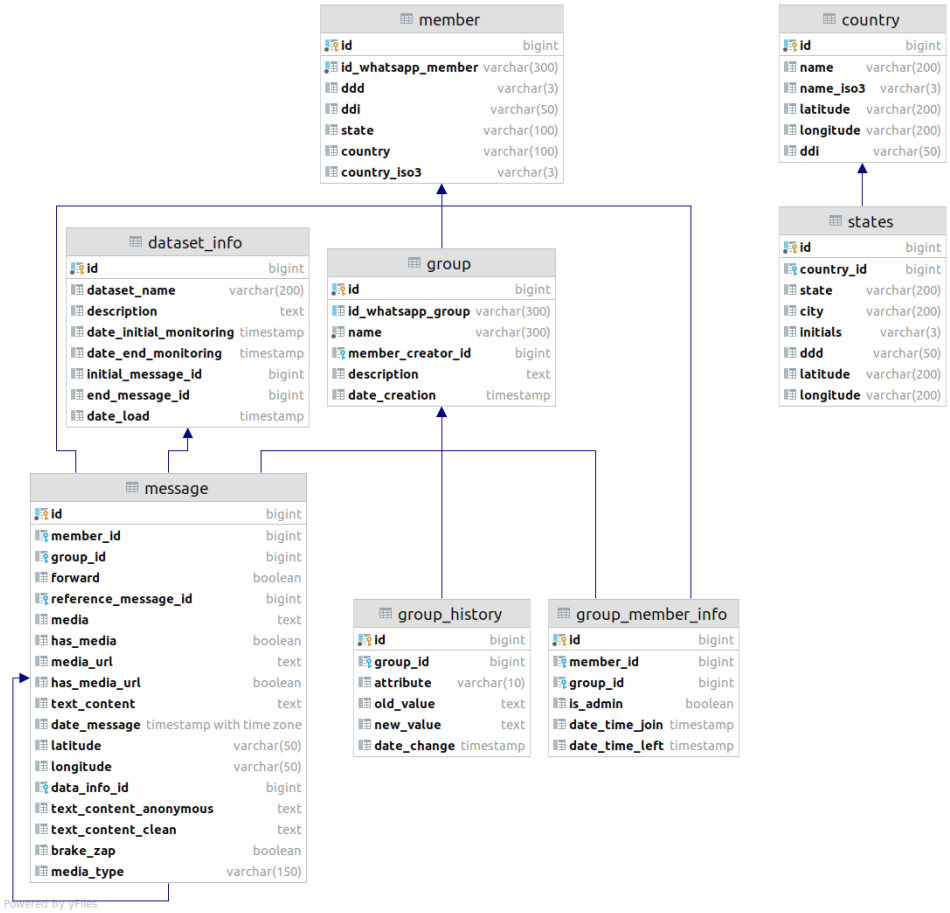

Figure 4 – The PostgreSQL Database Schema

3.2.3 ETL

As mentioned earlier, Selenium Web Driver captures messages that are spread across monitored public groups and stores this information in an SQLite database hosted on the android emulator. Thus, it is necessary to periodically perform a process of extracting, transforming, and loading (ETL) the data stored in SQLite to a PostgreSQL database, installed on the same virtual machine but outside the android emulator. This ETL step is necessary because the use of SQLite is not recommended for applications that manage large databases or that use the client/server architecture. Thus, we chose to use PostgreSQL to store the data already consolidated centrally. The choice of PostgreSQL is due to its popularity, performance, security, and free of charge. The data stored in PostgreSQL is used by the knowledge discovery and data visualization modules.

A series of rules are applied during the ETL process to avoid data redundancy, ensure data anonymization and preserve data integrity. The process of anonymizing personal information uses two different approaches: replacing characters and applying a hash function, precisely the MD5 hash specified in RFC 1321[1]. Hash is a function that maps input data (of variable length) in output data (of fixed length), usually a sequence of hexadecimal characters. Although there are more modern hash algorithms, the MD5 algorithm was chosen for the following reasons: simplicity, consolidation, for having implementation in several programming languages, for having PostgreSQL support, and for offering the possibility to represent the data through a fixed length of only 32 characters. So, user cell phone numbers and public group identifiers are anonymized using the MD5 hash. The personal information that occurs in the message body is identified using regex expressions. In this way, we use regex expressions to find phone numbers, cpf, cnpj and references to a specific user (made using the character @). Then, each character in the found term is replaced by the character “X”. Since the groups are public, our approach does not violate WhatsApp’s privacy policy [2].

The ETL process runs daily, from the execution of two scripts developed in Python language. Initially, the first script performs the SQLite backup. SQLite stores its information in a single file. Thus, this first script performs the copying, compression, and storage of this file in a secure repository. The compressed file is named containing the timestamp of the local machine. In this way, we can quickly identify the file creation date and time. The second script performs extracting, transforming, and load data from the backup file to PostgreSQL.

---

1 https://tools.ietf.org/html/rfc1321

2 https://www.whatsapp.com/legal/privacy-policy

In order to automate the ETL process, we evaluated three different tools: Nifi[3], Kafka[4], and Airflow[5]. These three investigated tools are open-source and maintained by the Apache Foundation. All have good documentation and a wide community. Apache Nifi and Kafka are tools developed in Java and suitable for ETL, and can be executed from Docker containers. Apache Airflow, in turn, was developed in Python and has the purpose of automating general tasks.

The analyzed tools allow the necessary automation for the Farol Digital platform. Besides, they enable the execution of ETL processes in real-time and over large volumes of data. However, due to the platform's initial requirements, including flexibility and programming simplicity, we opted for the manual execution of these two scripts mentioned previously. However, the adoption of one of these tools in future works is not ruled out.

The ETL process consists of a series of steps, which are illustrated in Figure 5 and described next.

- Initial configuration and metadata update. In this step, we define some parameters and metadata that will be used in the following steps. For example, we identify the date and time of the last time the ETL process was run. Consequently, we can identify the data that has not yet been exported to PostgreSQL, perform the ETL process incrementally and ensure that we will not store messages redundantly;
- Loading data about Whatsapp public groups. In this step, for each group found in the backup file, we apply the MD5 hash function over the group's name. Next, we verify if the group has been previously inserted in the database using the hash value. If not, the new group is inserted in the appropriate table using the anonymized data;
- Loading data about Whatsapp group members. In this step, for each member found in the backup file, we apply the MD5 hash function over the member's identifier. Next, we verify if the member has been previously inserted in the database using the hash value. If not, the new member is inserted in the appropriate table using the anonymized data. Finally, we verify if exist in the database relationships between a member and its groups. If some relationship does not exist, we add a new relationship between the member and the specific group in the database;
- Loading messages from Whatsapp public groups. This step comprises three different tasks:

[3] https://nifi.apache.org/
[4] kafka.apache.org
[5] https://airflow.apache.org/

– Check if the message contains “zap-lock” or “contact bomb”, which are particular texts to exploit vulnerabilities in the Whatsapp application that lead to the mobile device’s blocking;
– Anonymize personal information using the two approaches presented previously;
– Perform cleaning and tokenization of messages, standardizing them in lower case, removing accentuation, punctuation, emojis, and stopwords.

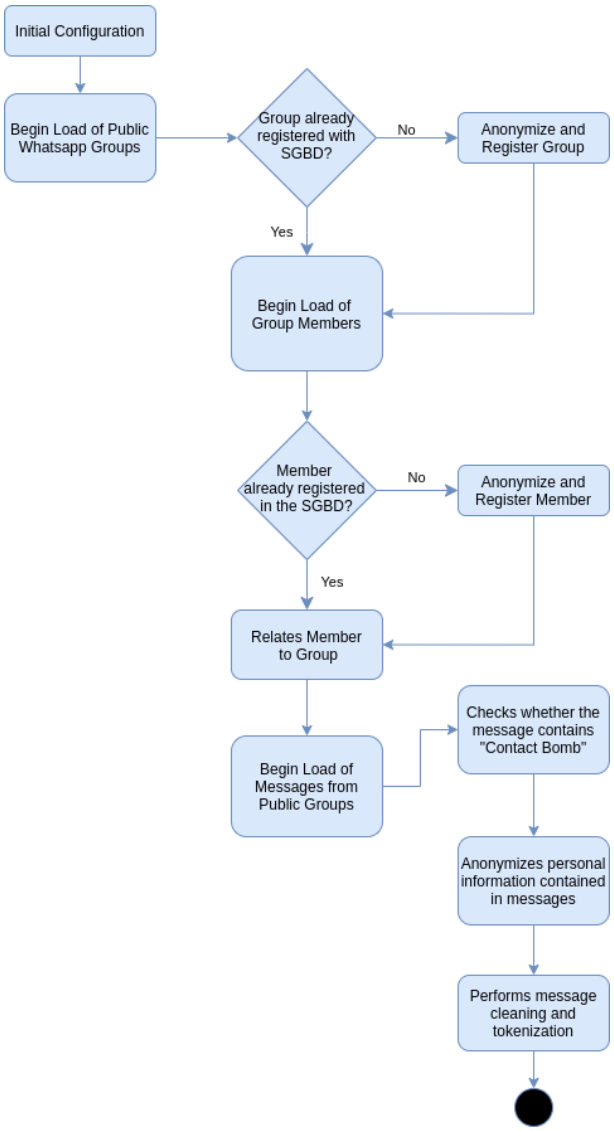

Figure 5 – The ETL Process

3.3 Module III - Knowledge Discovery

Knowledge Discovery can be defined as extracting useful information from large datasets using sophisticated algorithms/methods. Thus, this module explores the data stored in the PostgreSQL to finding implicit, previously unknown, and potentially useful patterns. The Knowledge Discovery module's main component is the Misinformation Detector, a machine learning classifier once trained and tested. This component receives a text as input and returns as output if the text is or not the misinformation. Besides, two other components are under development: a misinformation super-spreader users classifier and a bot detector.

It is important to emphasize that the development of predictive models or classifiers is not the focus of this work. Thus, we use existing misinformation classifiers. In the current version of the Digital Lighthouse platform, adding or replacing predictive models is hard-coded. As future work, we intend to design mechanisms that make the addition of new models more flexible.

3.4 Module IV - Data Visualization

Knowledge Discovery can be defined as extracting useful information from large datasets using sophisticated algorithms/methods. Thus, this module explores the data stored in the PostgreSQL to finding implicit, previously unknown, and potentially useful patterns. The Knowledge Discovery module's main component is the Misinformation Detector, a machine learning classifier once trained and tested. This component receives a text as input and returns as output if the text is misinformation. Besides, two other components are under development: a misinformation super-spreader users classifier and a bot detector.

4 CASE STUDY

To evaluate the platform proposed in this paper, we performed an exploratory case study using three different WhatsApp' messages datasets, covering relevant themes such as the Brazilian general elections campaign in 2018, the COVID-19 pandemic the vaccine for COVID-19. This case study was influenced by (JEDLITSCHKA; PFAHL, 2005; KITCHENHAM *et al.*, 2008; ROBSON; MCCARTAN, 2016; RUNESON; HÖST, 2009). Then, many data analysis techniques were applied to this dataset to get insights about misinformation spread.

Next, we will describe these three datasets in detail:

- Brazilian general elections: This dataset contains 282,601 messages, obtained from 5,364 users (cell phone chips), which participated in 59 WhatsApp public groups from August to October 2018;
- COVID-19 pandemic: This dataset contains 228,061 messages, obtained from 10,495 users (cell phone chips), which participated in 236 WhatsApp public groups in the period from March to June 2020;
- Vaccine for COVID-19: This dataset contains 16,056 messages, obtained from 1,857 users (cell phone chips), which participated in 175 WhatsApp public groups from December 2020 to January 2021.

Using the Data Visualization Module from the Lighthouse Platform, the user can choose a specific dataset or all data from all datasets. For simplicity, from this point onwards, all graphs will be illustrated using the COVID-19 dataset. Figure 6 shows the main screen of the Data Visualization component of the Lighthouse Platform.

4.1 Messages Characterization

Initially, the Lighthouse Platform shows some visualizations to characterize the used dataset. Figure 7 shows the proportion between messages with and without URL. In general, messages created to spread misinformation include a URL, often from a little-known website or blog, intending to give it credibility. Therefore, the presence of a URL can be a criterion for selecting messages to be analyzed by fact-checkers. As you can observe in Figure 7, a significant proportion of the caught messages (9.4%) involves some media file.

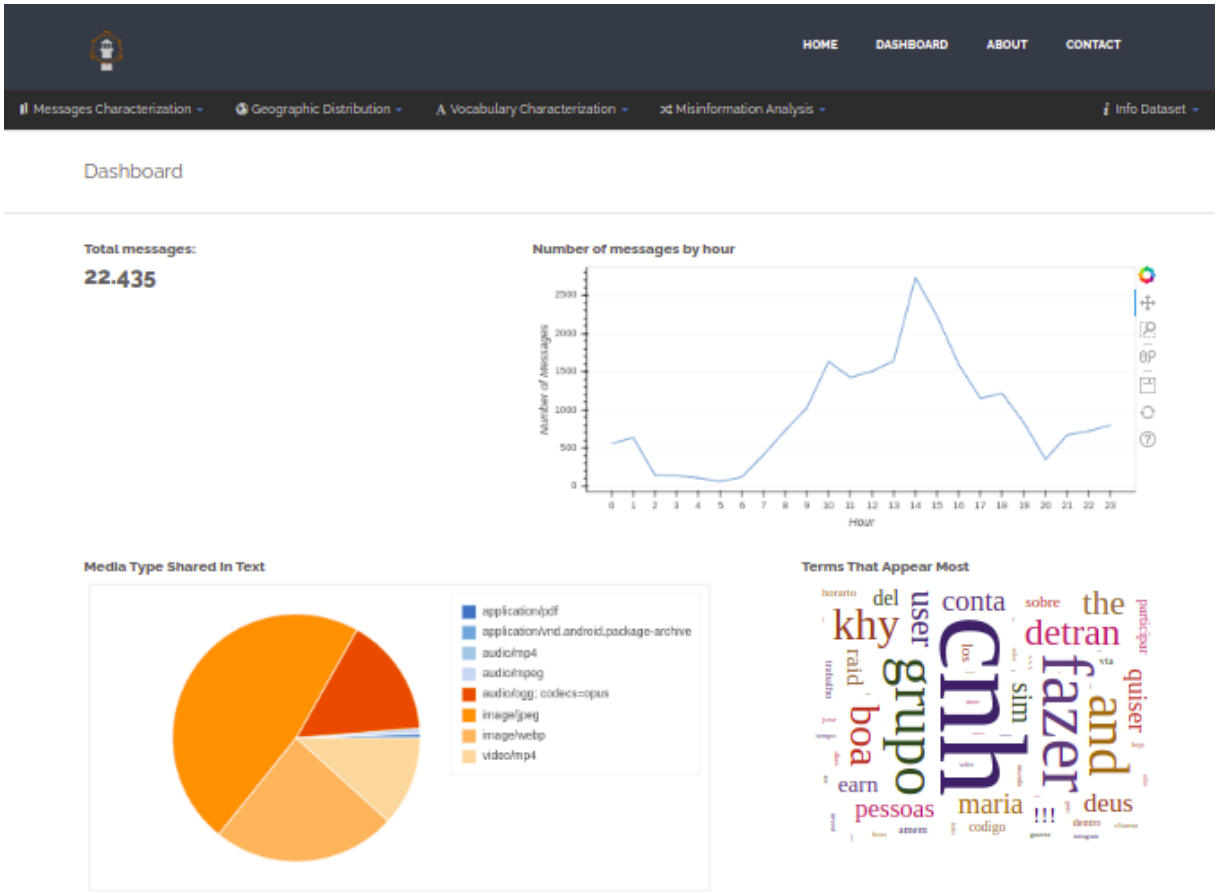

Figure 6 – The Visualization Module Main Screen

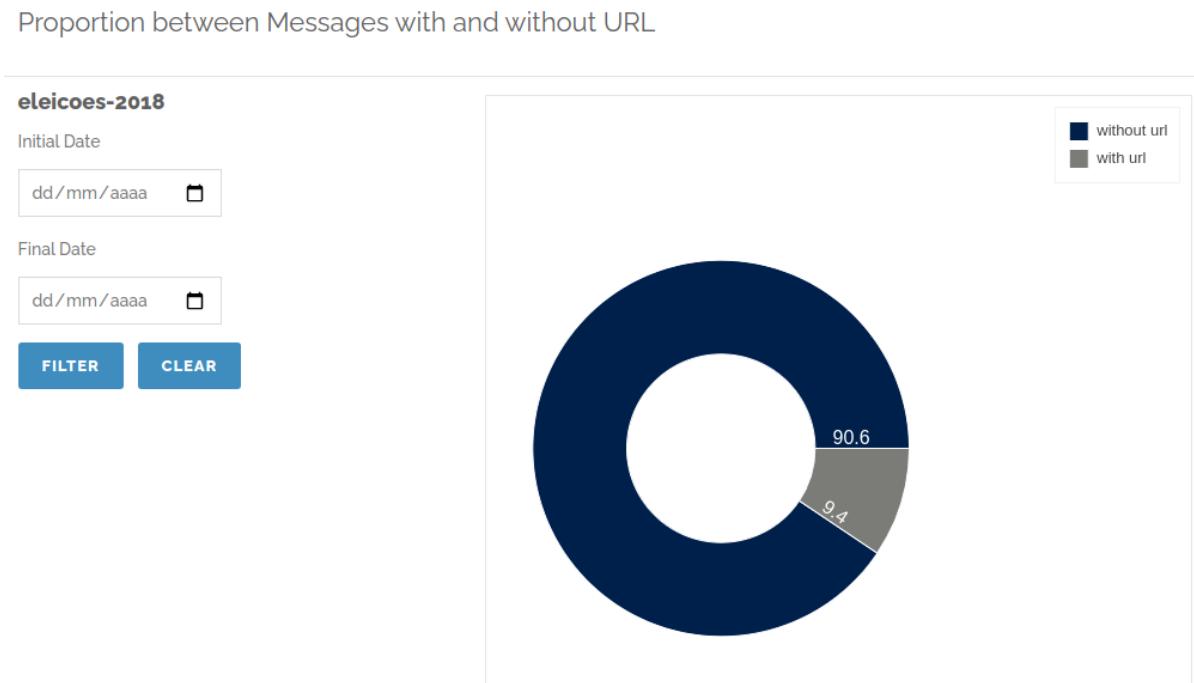

Figure 7 – Proportion between Messages with and without URL

Currently, audio, images, and videos are commonly used to spread misinformation. Therefore, the messages associated with these files are potential candidates to undergo a verifica-

tion process. Figure 8 shows the proportion between messages with and without media. As you can note, a significant proportion of the caught messages (32.90%) involves some media file.

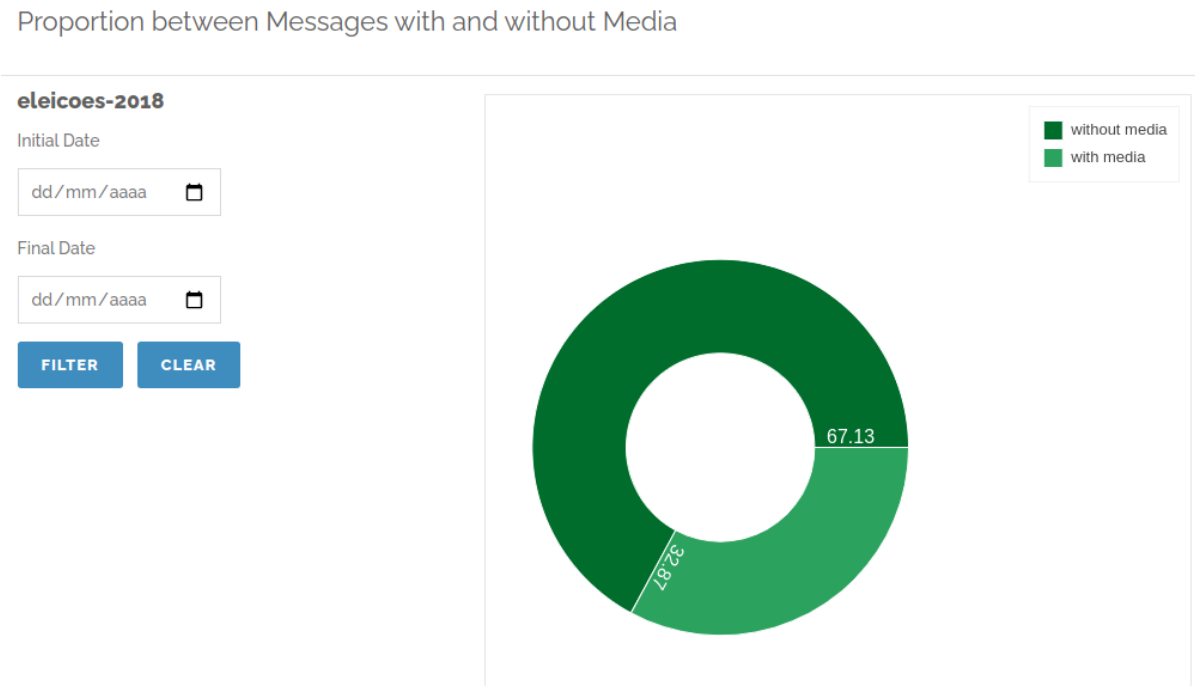

Figure 8 – Proportion between Messages with and without Media

Figure 8 illustrates the types of media most commonly encountered in the monitored WhatsApp public groups. Note that images are still the most shared media. However, the number of audio files shared already outperforms the number of videos, which shows a WhatsApp trend.

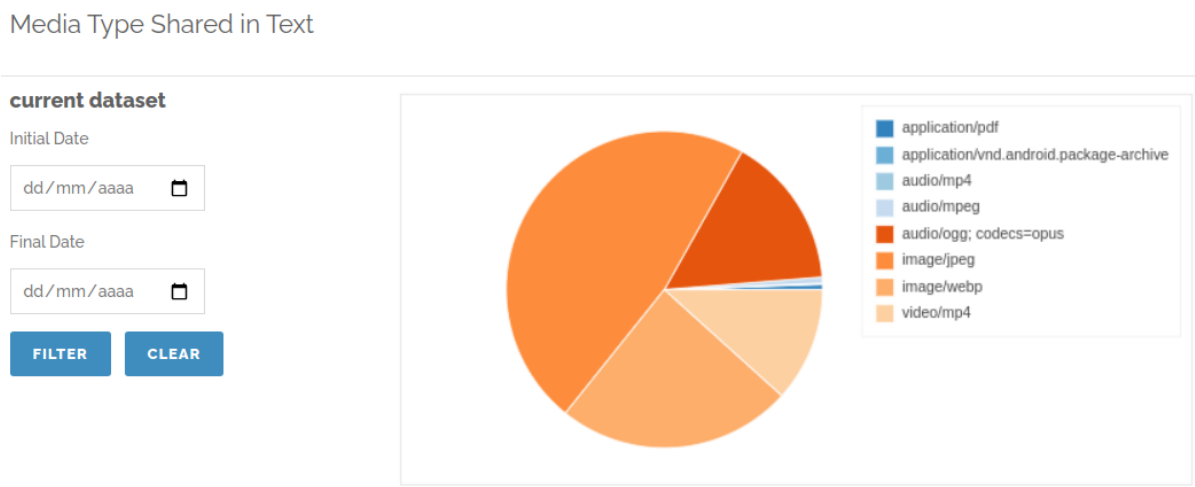

Figure 9 – Proportion between Media Types

In the 2018 Brazilian elections, many cell phone chips from foreign countries were used to massive messaging with electoral advertisement. Thus, monitor the messages sent by

these chips is an important task to identify misinformation spreading. Figure 10 illustrates the proportion of foreign countries messages.

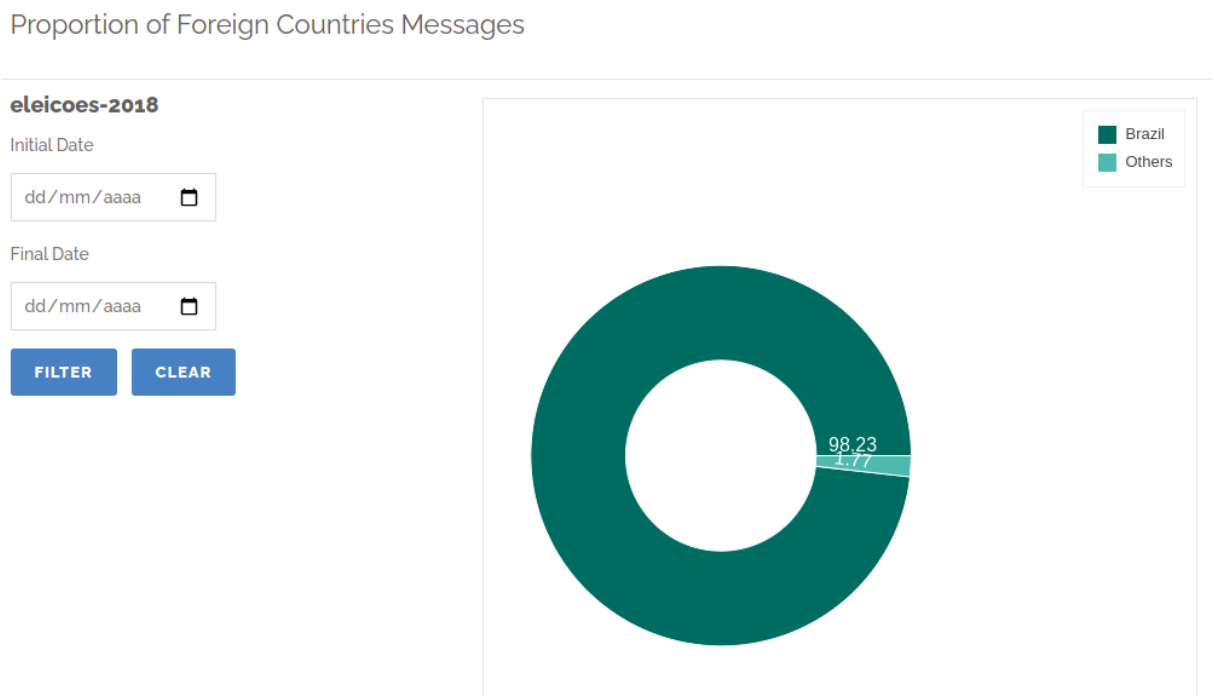

Figure 10 – Proportion of Foreign Countries Messages

Figure 11 shows the distribution messages sending time by the day hours. As we can imagine, the peak of sending messages occurs when reserved for lunch (between 12 and 14 hours) and in the early evening, just after work hours.

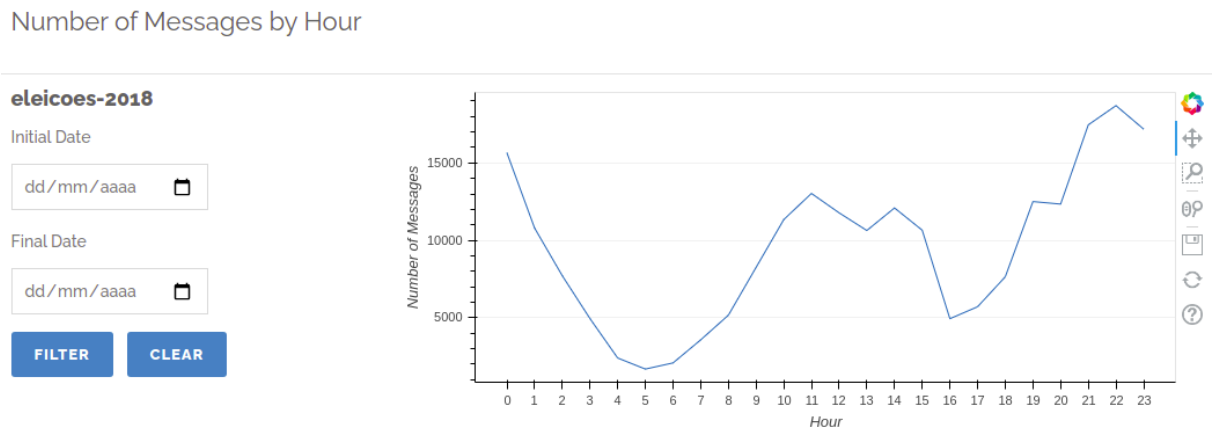

Figure 11 – Number of Messages by Hour

4.2 Geographic Distribution

Another relevant aspect to observe in the monitored groups is the geographic location of users (cell phone chips), both Brazilians and foreigners, besides these users' activity level. Figure 12 shows the Brazilian states with more quantity of messages. As might be expected, the most populous states have the most significant amount of messages sent.

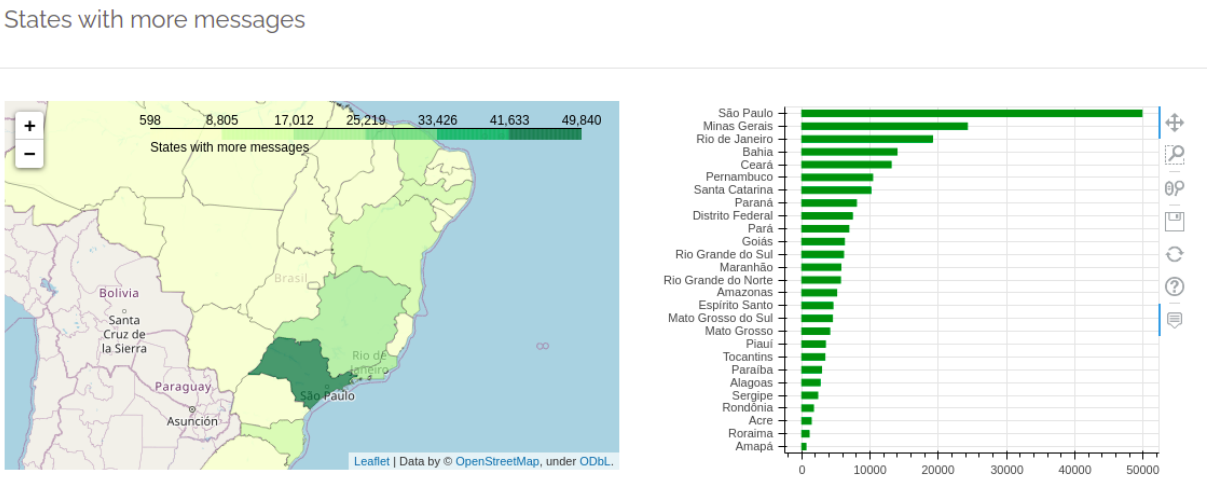

Figure 12 – States with more Messages

Figure 13 illustrates the Brazilian states with more users (cell phone chips). As might be expected, the most populous states have the largest amount of users.

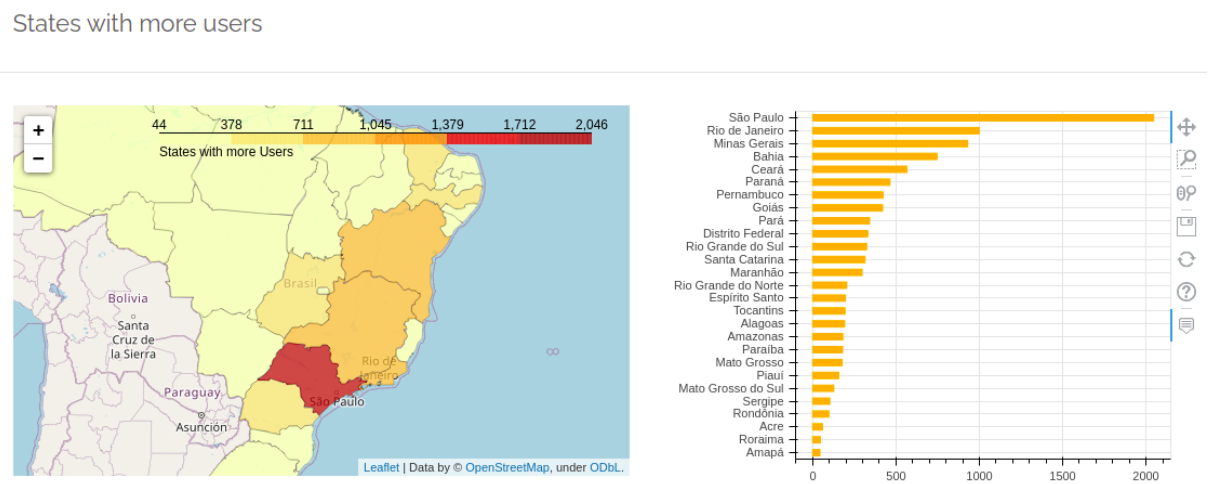

Figure 13 – States with more Users

However, when analyzing the states with more messages per user (Figure 14), we can observe that not so populous states such as Mato Grosso do Sul, Santa Catarina, and Amazonas,

have the most active users.

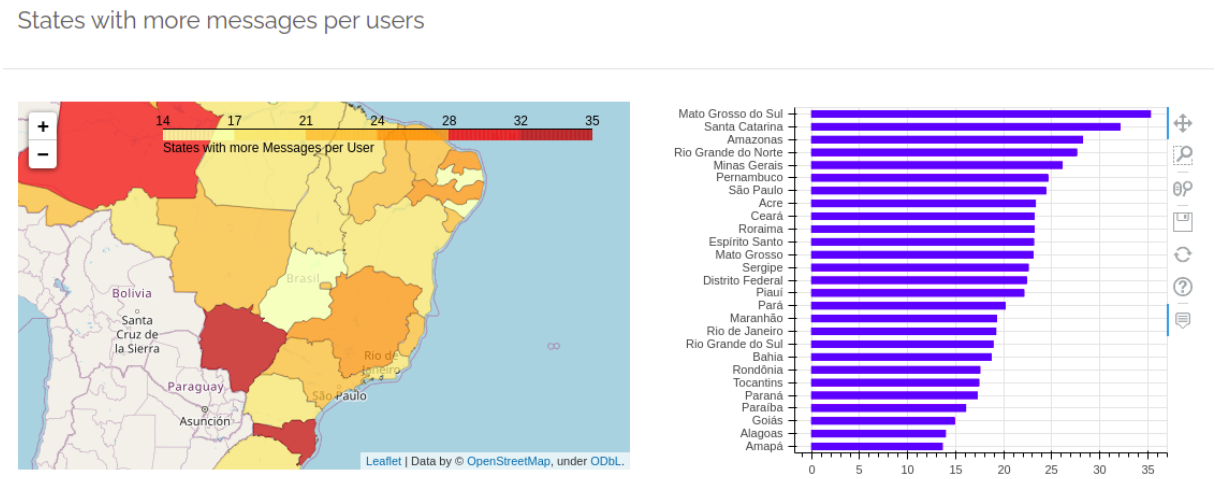

Figure 14 – States with more Messages per User

As previously mentioned, cell phone chips from foreign countries have been used in Brazil for massive messaging, many times spreading misinformation. Figure 15 illustrates the quantity of messages sent by foreign countries cell phone chips by country, while Figure 16 shows the countries with the lagers ratio between sent messages and the number of users.

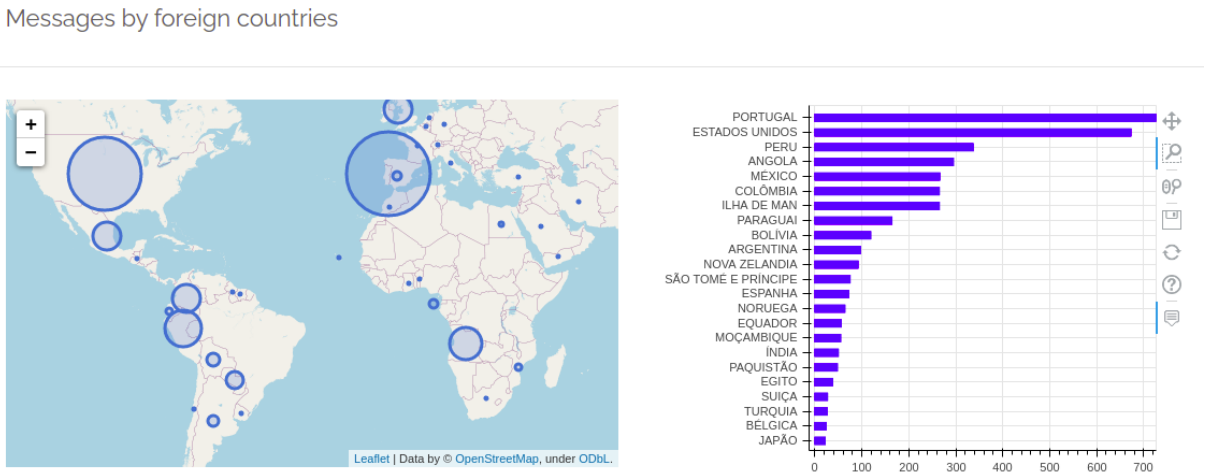

Figure 15 – Messages by Foreign Countries

4.3 Vocabulary Characterization

Another aspect that needs to be analyzed is the characteristics of the vocabulary used in the text messages, since there is a strong relationship between the used vocabulary and the social network, in this case, WhatsApp. Figure 17 shows the number of messages

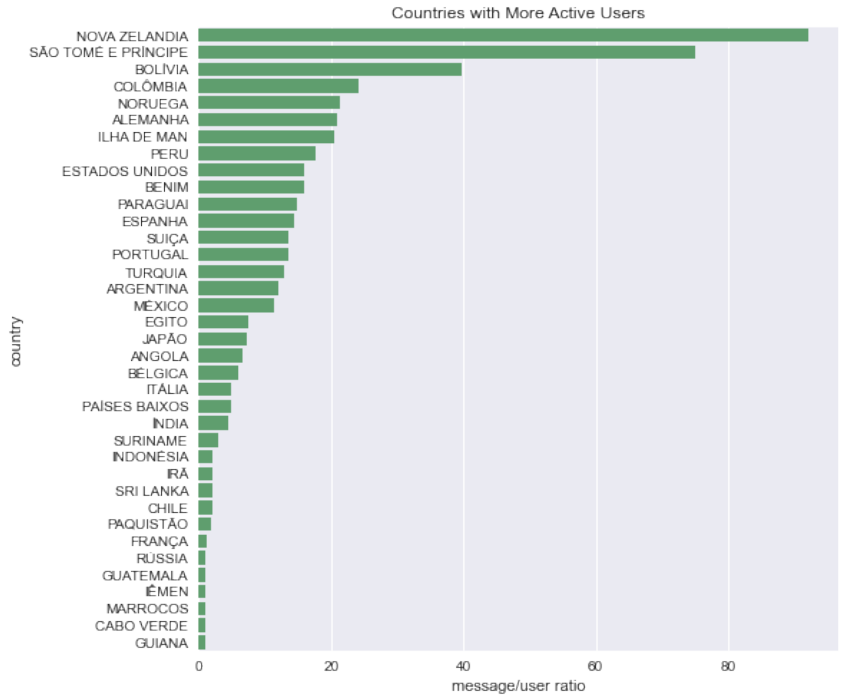

Figure 16 – Countries with More Active Users

by the number of words contained in the message. As we can note, there are few messages with a large number of words and a high number of messages with few words. Figure 18 shows the word cloud highlighting the most popular terms. Finally, Figure 19 illustrates a word network, highlighting the connection between the words in the messages. Next, we detach some examples: a) [querem derrubar presidente jair bolsonaro governo federal brasil ninguém consegue roubar poder executivo], b) [combate mortes pandemia novo coronavirus] and c) [trump remédio bolsonaro].

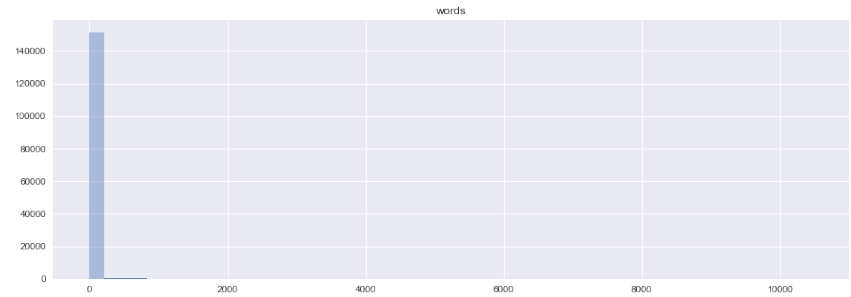

Figure 17 – Number of Messages by the Number of Words in the Message

Figure 18 – Word Cloud

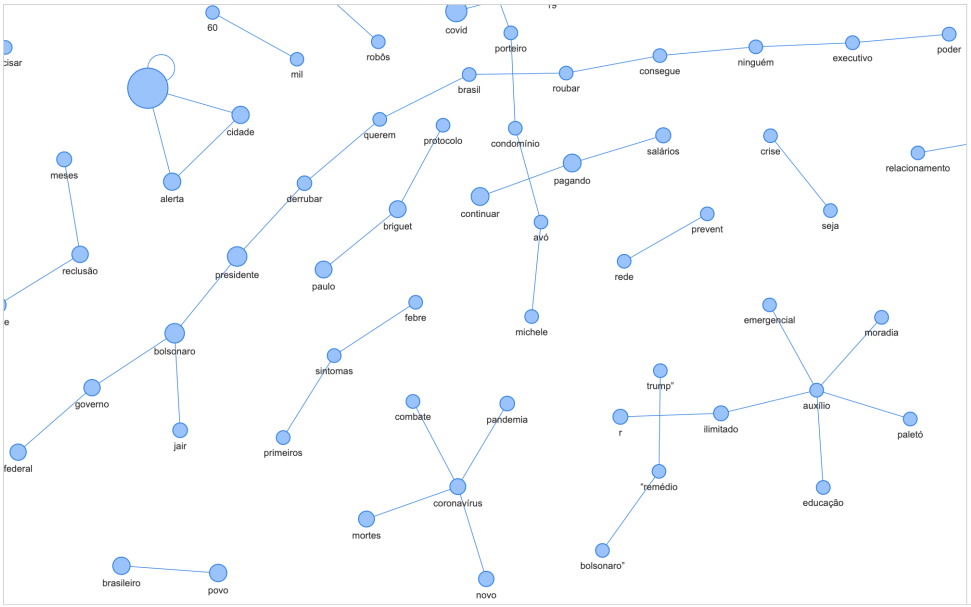

Figure 19 – Interactive Word Network

4.4 Misinformation Analysis

The last aspect to be explored using the Data Visualization component of the Lighthouse Platform is the misinformation analysis. In this context, different information about messages and users are explored to identify text messages containing misinformation and superspreaders, that is, the users that most spread misinformation.

To exemplify the Data Visualization component's main features, we built a new dataset from the COVID-19 dataset, adding only the messages with at least one of these words: "*covid*", "*corona*", "*coronga*" or "*virus*". The resulting dataset had 3,014 messages. Table 3 contains the seven most shared messages. The "*Sharings*" column indicates how many times the message was shared. The "*Mis*" column indicates if the messages contain or not the misinformation. Finally, the column "*Groups*" denotes the number of distinct groups where the

message was shared. It is important to highlight that all the seven most shared messages contain misinformation.

Table 3 – Most Shared Messages

Sharings	**Text**	**Mis**	**Groups**
43	"PATRIOTA! *VAMOS ACORDAR BRASIL!!!! E VOCE AINDA ACREDITANDO NESTA FARSA DE COVID19, É UM GOLPE QUE FOI ARQUITETADO PARA ENGANAR OS BRASILEIROS, MENOS ESCLARECIDOS...* #NAOFIQUEEMCASA #VAMOSTRABALHAR #BOLSONAROESTACERTO	Yes	40
26	"Pesquisa com mais de 6.000 médicos em 30 países diz que hidroxicloroquina é o tratamento mais eficaz para coronavírus."	Yes	23
23	"Dra. Nise Yamaguchi integra gabinete de crise e propõe a cloroquina como tratamento imediato nos casos de coronavírus."	Yes	23
23	"Herança maldita: Mandetta renova contratos de publicidade de R$ 1 bilhão firmados no governo Dilma..."	Yes	14
22	"Organização Mundial de Saúde: O aborto é “essencial” durante a pandemia de coronavírus chinês."	Yes	22
18	"Prezados amigos.. vocês sabiam que, todos os problemas da humanidade foram curados com esse pânico fake do covid19?????? Vejam?? Sempre morreram milhares de pessoas de H1N1, POIS, NUNCA FOI ERRADICADA ESTA GRIPE, DE AIDS que NUNCA FOI ERRADICADA, DE TUBERCULOSE, DE INFARTO, DE BRIGAS DOMÉSTICAS, DE IDADE, DE INSUFICIÊNCIA RESPIRATÓRIA, DE CÂNCER, DE DIVERSAS OUTRAS DOENÇAS E MALES... TUDO ACABOU..."	Yes	18
16	"*Atenção*: Isso a Globo não mostra. Banco Mundial acaba de lançar um documento que ressalta o papel do comércio internacional na mitigação dos impactos do coronavírus. A instituição argumenta que a manutenção dos fluxos de comércio será crucial para o suprimento de itens médicos e alimentos — e portanto limitar impactos negativos sobre empregos e nível de pobreza em escala global. O trabalho do Banco Mundial coloca o Brasil como “Exemplo 1” no quadro “Melhores Práticas em Lidar com a Covid-19”. #BolsonaroTemRazão"	Yes	11

Table 4 contains the 10 most active users together with the number of messages shared by each user. How you can see, the user identification is anonymized.

Let us take a particular user, for example, the user with Id -9126362355320474072, which sent 67 messages. Table 5 contains all messages shared by the user -9126362355320474072. The “*Sharings*” column indicates how many times the message was shared. The “*Mis*” column indicates if the messages contain or not the misinformation. Note that all 67 messages shared by

Table 4 – Most Active Users

User Id	Number of Messages
3346599479176653344	110
8121536360444460807	102
-9126362355320474072	67
8900877460624761918	62
1721737435325801397	60
600984133124592046	58
4318942240836460215	47
3089133927534067747	44
-2525857733931129209	42
-7724257649060270690	33

this user contain misinformation. Besides, some messages were shared many times, probably in different groups.

Table 5 – Most Shared Messages of User Id -9126362355320474072

Sharings	Text	Mis
22	"Pesquisa com mais de 6.000 médicos em 30 países diz que hidroxicloroquina é o tratamento mais eficaz para coronavírus."	Yes
22	"Dra. Nise Yamaguchi integra gabinete de crise e propõe a cloroquina como tratamento imediato nos casos de coronavírus."	Yes
22	"Organização Mundial de Saúde: O aborto é “essencial” durante a pandemia de coronavírus chinês."	Yes
1	"ENTENDA COMO FOMOS IMPEDIDOS DE VOTAR O FUNDÃO PARA O COMBATE AO CORONAVÍRUS. O deputado Rodrigo Maia inadmitiu a emenda do NOVO que destinaria os recursos do fundão e fundo partidário para a saúde e também a emenda da redução dos salários de políticos."	Yes

Now, let us take a specific message of the user -9126362355320474072, for example, the message in the first row of Table 5. Table 6 contains the date and time of each sharing of the selected message, besides the group in which the message was shared. Note that the selected message was shared 22 times in 22 different groups. Besides, all messages were sent in four minutes. So, we can classify the user -9126362355320474072 as a misinformation super-spreader.

Finally, we can query the sites most used in the messages. Table 7 contains the five most shared sites together with the number of messages that refer to each site.

Table 6 – Details of the Most Shared Message of User Id -9126362355320474072

Date	**Time**	**Group Id**
2020/04/06	18:36	2020_117
2020/04/06	18:36	2020_133
2020/04/06	18:36	2020_153
2020/04/06	18:36	2020_187
2020/04/06	18:36	2020_243
2020/04/06	18:36	2020_26
2020/04/06	18:36	2020_96
2020/04/06	18:37	2020_128
2020/04/06	18:37	2020_131
2020/04/06	18:37	2020_174
2020/04/06	18:37	2020_84
2020/04/06	18:38	2020_146
2020/04/06	18:38	2020_170
2020/04/06	18:38	2020_171
2020/04/06	18:38	2020_22
2020/04/06	18:38	2020_225
2020/04/06	18:38	2020_229
2020/04/06	18:38	2020_233
2020/04/06	18:38	2020_73
2020/04/06	18:38	2020_99
2020/04/06	18:39	2020_105
2020/04/06	18:39	2020_226

Table 7 – Most Shared Sites

Qtd	**Site**
106	GazetaBrasil
73	ConexaoPolitica
31	PortalNovoNorte
24	AgoraParana
12	UOL

5 CONSIDERATIONS

5.1 LGPD Compliance

Personal data protection is a problem that has been widely discussed in several countries. Several countries created specific regulations and laws to protect fundamental rights and privacy, such as the General Data Protection Regulation (GDPR) from European Union (EU) and the General Data Protection Law (LGPD) in Brazil. These laws propose many standards for the security and privacy of personal data. Besides, they regulate how personal data should be protected and how data may be shared between other companies, institutions or countries. These regutaions encourage the use of anonymization (and its approaches, processes, and methods) in order to reach personal data security. The full text of LGPD can be found in Annex A.

LGPD (Law No. 13,709) was approved on August 14, 2018. It is relevant to highlight that GDPR was the basis for the conception of LGPD. Besides, with LGPD, Brazil became part of the countries with specific legislation to protect data and citizens' privacy. About 120 countries around the world currently have data protection laws (SILVA *et al.*, 2019). The General Data Protection Law (LGPD) aims to protect personal data and the free movement of such data. In addition, LGPD regulates the activities of processing personal data, including in digital media, applied to any person, physical or legal, that performs the processing of personal data, online and/or offline.

LGPD is composed of 65 articles, organized into ten chapters. LGPD defines a set of rights for data holders and establishes rules and limits for companies regarding personal data collection, storage, processing, and sharing, especially in digital media, in order to protect the fundamental rights of freedom, privacy, and the free formation of the personality of each individual. LGPD can be seen as a general guideline for data protection in Brazil. Thus, LGPD does not seek to replace those currently existing but to establish general principles and rules to be fulfilled in a much more beneficial way for the personal data holders.

LGPD defines concepts like Data Holder, Personal Data, Sensitive Data, Public Data and Anonymous Data. A data holder is any natural person to whom the personal data referred to are treated. Personal Data is information that allows you to identify, directly or indirectly, an individual. Sensitive Data is personal data that reveals the racial or ethnic origin, religious or philosophical beliefs, political opinions, union membership, genetic, biometric, and health issues or sexual life of a person, that is, they are data that can lead to discrimination of a person. Public

Data is data published and collected from a public source. Anonymous Data is that data relating to the holder that cannot be identified, considering the use of reasonable technical means and available at the time of your treatment.

On the other hand, data processing is any operation carried out with personal data. In this sense, the LGPD specifies processing agents, such as the Data Controller, Data Operator, and Data Supervisor. Data Controller (legal entity or individual) is responsible for decisions related to personal data processing. Data Operator is the one (natural or legal person) who performs data processing on behalf of the Controller. A supervisor is a person designated by the Controller and the Operator to act as a communication channel between the Controller, the data holders personal data, and ANPD (National Data Protection Authority), a public administration body responsible for ensuring, implementing, and supervising compliance with this Law throughout the national territory.

The principles described in Article 6 of the LGPD impose new rules and limitations on how personal data should be treated. Thus, personal data processing activities will observe, in addition to good faith, the following principles: Purpose, Adequacy, Necessity, Transparency, Security, Open Access, Prevention, Non-Discrimination, Data Quality, and Accountability.

With regard to consent, the LGPD establishes that it is a free manifestation, which authorizes the processing of personal data for a certain purpose. Such consent will be considered a temporary authorization since it can be revoked at any time by the data subject. Therefore, consent for the processing of sensitive personal data must be provided in a specific and detached manner, that is, the processing agent responsible for obtaining the consent must be concerned with obtaining a special authorization for the processing of sensitive personal data.

Related to the sanctions, in addition to the responsibility to indemnify the data subject, the LGPD provides for administrative penalties in the event of non-compliance. These sanctions, applicable by the national authority, can be a warning until the imposition of financial sanctions, which can reach 2% of the billing, limited to R$ 50 million per infraction, with the possibility of blocking and eliminating the data involved in the infraction.

Concerning exceptions, the law does not apply to the processing of personal data carried out by natural persons for private and non-economic purposes or for exclusively journalistic, artistic, academic, public security, national defense, or investigative purposes.

Thus, LGPD aims to bring greater legal certainty to the holders of personal data while providing greater legal compliance to organizations that demonstrate that they comply

with current legislation. LGPD regulates how any individual (whether natural or legal, under the public or private law) must store and process Brazilian citizens' data while respecting citizens' privacy. Furthermore, individuals that do not comply with the LGPD can be monetarily penalized, which underscores the seriousness of ensuring LGPD compliance.

However, the LGPD does not apply for journalistic or academic purposes. For scientific research purposes, the law only recommends that personal data be anonymized whenever possible. Next, we reproduce part of LGPD abou this aspect.

> "Art. 4 This Law does not apply to the processing of personal data:
> ...
> II - carried out for the sole purpose of:
> a) journalistic and artistic; or
> b) academics, applying arts. 7th and 11th of this Law;
> ...
> Art. 7 The processing of personal data can only be carried out in the following cases:
> ...
> IV - to carry out studies by a research body, guaranteeing, whenever possible, the anonymization of personal data;
> ...
> Art. 11. The processing of sensitive personal data can only occur in the following cases:
> ...
> c) carrying out studies by a research body, guaranteeing, whenever possible, the anonymization of sensitive personal data;".

5.2 WhatsApp Privacy Policy Compliance

The last version of WhatsApp's Privacy Policy was released on 11th of Jannuary, 2021 that enabled the encrypted messaging App to share a significant amount of data with parent company Facebook in order to enhance business usage amongst group companies. However, this does not involves personal chats between users but allows the sharing of data on business interactions across the group. In simple terms, many businesses rely on WhatsApp to communicate, for instance, if you purchase a flight ticket, it will be sent to your WhatsApp. In this context, WhatsApp helps the business to communicate with their client or customers. WhatsApp works with these business entities that use Facebook or third parties to help store and manage better communication with users on WhatsApp. With the present privacy policy, WhatsApp will require consent to sharing transaction data, IP address, mobile device information and data on how they interact with business with Facebook group companies. This will help Facebook to personalise contents and display relevant advertisements across the group multiple social platforms. This will also enable users to interlink services like Facebook Pay account to pay for

things on a messaging app. The full text of WhatsApp's Privacy Policy can be found in Annex B.

It is essential to highlight that WhatsApp user's data is public for its contacts: any person or businesses with whom it communicates. Next, we reproduce part of WhatsApp's privacy policy.

> "Your Contacts And Others. Users, including businesses, with whom you communicate can store or reshare your information (including your phone number or messages) with others on and off our Services. You can use your Services settings and the "block" feature in our Services to manage who you communicate with on our Services and certain information you share".

5.3 Threats to Validity

Even with the careful case study design, this research can be affected by different factors which, while extraneous to the concerns of our work, can invalidate its main findings. The executed actions in order to mitigate the impact of these factors on the research results are described according to Wohlin *et al.* (2012), as follows.

5.3.1 Internal Validity

It's important to highlight that if the group owner (or spreader) uses a URL shortener, probably, the new URL will not contain the text 'chat.whatsapp.com'. So, this action can embarrass the operation of the Finding Public Groups Module.

Additionally, it is worth mentioning that, after our bot joins a specific group, it remains there indefinitely. Furthermore, the bot acts only as an observer. That is, it does not interact with any member of the group. From the instant in which the bot joins the group, all posted messages are captured. Therefore, we have no way of getting the messages posted in the group before the bot joined it.

Web Crawler and the Data Scraper operate periodically using intervals between 10 and 20 seconds. However, as they are querying the Web, no data is lost. Besides, the Selenium Web Driver operates periodically. Although, if it runs one time per day, no data is lost.

Finally, the proposed architecture uses an Android Emulator. Thus, this architecture is strong related with the Android Platform. However, the monitoring is independent of the user's device.

5.3.2 Construct Validity

Construct validity is concerned with the relationship between theory and observation. In this context, our main concern is that ee are monitoring only a few public groups. Many other groups indeed exist. Thus, the snapshot studied may not represent the general reality of public WhatsApp groups.

5.3.3 External Validity

External validity is the extent that the obtained results of a case study can be generalized to other relevant research scenarios. The results of an externally valid case study can be generalized and applied safely to the software engineering practice and be recommended as patterns. To enhance external validity, we collect a large set of information, including data about users (cell phone number, state, country, etc.), public groups (id, name, number of users, creation date, etc.), and messages (text, date and time, whether they contain media and the media type, if they contain URL, if they are forwarded, etc).

5.3.4 Conclusion Validity

Conclusion validity is the extent to which the conclusions about the presence of a statistically significant relationship between the treatments and the outcomes are valid. To mitigate threats to conclusion validity and increase the reliability of this exploratory case study, we built three different datasets. This validation is important to evaluate how the proposed approach will generalize to other messaging applications as Telegram, for example.

6 CONCLUSIONS

The fast spread of misinformation through WhatsApp messages poses a significant social problem. In this work, we presented a Digital Lighthouse platform for finding, gathering, analyzing, and visualizing public groups in WhatsApp. To evaluate the proposed platform, we used it to build three different WhatsApp' messages datasets, covering relevant themes such as the Brazilian general elections campaign in 2018, the COVID-19 pandemic, and the vaccine COVID-19. Besides, we presented a case study using the proposed platform. Initially, we characterize the used dataset. Next, we explored the geografic distribution of the messages. Then, we performed a vocabulary characterization. Finally, we performed a misinformation analysis, and we identified a misinformation super-spreader. Thus, we hope that our platform can help journalists and researchers understand misinformation propagation in Brazil.

The main issue investigated in this work was the possibility of designing a mechanism to monitor the misinformation circulating in public WhatsApp groups. In this sense, the proposed platform achieved its objective. It is worth noting that during the work, there was a special concern to comply with the legislation in force, the General Data Protection Law in Brazil. One of the modules implemented by this platform comprises the extraction, transformation and loading of collected data. At this moment, this process occurs manually, but the research showed the possibility of automation and even the guarantee that it takes place in real-time. The platform has also proved useful in identifying and characterizing messages and users who diffuse misinformation, as well as in the identification and classification of websites that act as a source of misinformation. Making this platform public is important to ensure the advancement of the research and the fight against misinformation. To achieve this goal, we provide access to the platform from a website. Through it, it is possible to visualize the data collected and processed. This site was built using a modular architecture so that it can continue to evolve, allowing the inclusion of new features in the future.

The main contributions of this research are:

1. An architecture for monitoring the misinformation that spreads in public WhatsApp groups;
2. An implementation of the designed architecture;
3. A case study involving three different WhatsApp' messages datasets, covering relevant themes such as the Brazilian general elections campaign in 2018, the COVID-19 pandemic, and the vaccine COVID-19.

As future work we intend to:

- Find out the main topics covered in each group.
- Conceive a strategy that allows the labeling of messages in a collaborative manner.
- Address the problem of shortened links.
- Investigate messages that arrive or depart from a particular state or country.
- Search for a way for the Digital Lighthouse platform to become independent of Android.
- Conceive a flexible mechanism to add and change predictive models easily.

BIBLIOGRAPHY

FAUSTINI, P.; COVÕES, T. F. Fake news detection using one-class classification. In: **8th Brazilian Conference on Intelligent Systems, BRACIS 2019, Salvador, Brazil, October 15-18, 2019**. IEEE, 2019. p. 592–597. Disponível em: https://doi.org/10.1109/BRACIS.2019.00109.

GAGLANI, J.; GANDHI, Y.; GOGATE, S.; HALBE, A. Unsupervised whatsapp fake news detection using semantic search. In: IEEE. **2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)**. [*S. l.*], 2020. p. 285–289.

GARIMELLA, K.; TYSON, G. Whatsapp, doc? A first look at whatsapp public group data. **CoRR**, abs/1804.01473, 2018. Disponível em: http://arxiv.org/abs/1804.01473.

GUO, B.; DING, Y.; YAO, L.; LIANG, Y.; YU, Z. The future of misinformation detection: New perspectives and trends. **CoRR**, abs/1909.03654, 2019. Disponível em: http://arxiv.org/abs/1909.03654.

HAMDI, T.; SLIMI, H.; BOUNHAS, I.; SLIMANI, Y. A hybrid approach for fake news detection in twitter based on user features and graph embedding. In: HUNG, D. V.; D'SOUZA, M. (Ed.). **Distributed Computing and Internet Technology - 16th International Conference, ICDCIT 2020, Bhubaneswar, India, January 9-12, 2020, Proceedings**. Springer, 2020. (Lecture Notes in Computer Science, v. 11969), p. 266–280. Disponível em: https://doi.org/10.1007/978-3-030-36987-3_17.

JEDLITSCHKA, A.; PFAHL, D. Reporting guidelines for controlled experiments in software engineering. In: IEEE. **Empirical Software Engineering, 2005. 2005 International Symposium on**. [*S. l.*], 2005. p. 10–pp.

KITCHENHAM, B.; AL-KHILIDAR, H.; BABAR, M. A.; BERRY, M.; COX, K.; KEUNG, J.; KURNIAWATI, F.; STAPLES, M.; ZHANG, H.; ZHU, L. Evaluating guidelines for reporting empirical software engineering studies. **Empirical Software Engineering**, Springer, v. 13, n. 1, p. 97–121, 2008.

LAZER, D. M. J.; BAUM, M. A.; BENKLER, Y.; BERINSKY, A. J.; GREENHILL, K. M.; MENCZER, F.; METZGER, M. J.; NYHAN, B.; PENNYCOOK, G.; ROTHSCHILD, D.; SCHUDSON, M.; SLOMAN, S. A.; SUNSTEIN, C. R.; THORSON, E. A.; WATTS, D. J.; ZITTRAIN, J. L. The science of fake news. **Science**, American Association for the Advancement of Science, v. 359, n. 6380, p. 1094–1096, 2018. ISSN 0036-8075. Disponível em: https://science.sciencemag.org/content/359/6380/1094.

MAALEJ, Z. The language of deception: A discourse analytical study by dariusz galasinski. **Discourse Studies**, Sage Publications, Ltd., v. 3, n. 3, p. 376–378, 2001. ISSN 14614456, 14617080. Disponível em: http://www.jstor.org/stable/24047513.

MACHADO, C.; KIRA, B.; NARAYANAN, V.; KOLLANYI, B.; HOWARD, P. A study of misinformation in whatsapp groups with a focus on the brazilian presidential elections. In: . New York, NY, USA: Association for Computing Machinery, 2019. (WWW '19), p. 1013–1019. ISBN 9781450366755. Disponível em: https://doi.org/10.1145/3308560.3316738.

MAROS, A.; ALMEIDA, J.; BENEVENUTO, F.; VASCONCELOS, M. Analyzing the use of audio messages in whatsapp groups. In: **Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan**. ACM, 2020. Disponível em: https://doi.org/10.1145/3366423.3380070.

MELO, P. de F.; VIEIRA, C. C.; GARIMELLA, K.; MELO, P. O. S. V. de; BENEVENUTO, F. Can whatsapp counter misinformation by limiting message forwarding? **CoRR**, abs/1909.08740, 2019. Disponível em: http://arxiv.org/abs/1909.08740.

MONDAL, M.; CORREA, D.; BENEVENUTO, F. Anonymity effects: A large-scale dataset from an anonymous social media platform. In: GADIRAJU, U. (Ed.). **HT '20: 31st ACM Conference on Hypertext and Social Media, Virtual Event, USA, July 13-15, 2020**. ACM, 2020. p. 69–74. Disponível em: https://doi.org/10.1145/3372923.3404792.

MONTEIRO, R. A.; SANTOS, R. L. S.; PARDO, T. A. S.; ALMEIDA, T. A. de; RUIZ, E. E. S.; VALE, O. A. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In: VILLAVICENCIO, A.; MOREIRA, V. P.; ABAD, A.; CASELI, H. de M.; GAMALLO, P.; RAMISCH, C.; OLIVEIRA, H. G.; PAETZOLD, G. H. (Ed.). **Computational Processing of the Portuguese Language - 13th International Conference, PROPOR 2018, Canela, Brazil, September 24-26, 2018, Proceedings**. Springer, 2018. (Lecture Notes in Computer Science, v. 11122), p. 324–334. Disponível em: https://doi.org/10.1007/978-3-319-99722-3_33.

QIU, X.; OLIVEIRA, D. F.; SHIRAZI, A. S.; FLAMMINI, A.; MENCZER, F. Limited individual attention and online virality of low-quality information. **Nature Human Behaviour**, Nature Publishing Group, v. 1, n. 7, p. 0132, 2017.

REIS, J. C. S.; MELO, P. d. F.; GARIMELLA, K.; BENEVENUTO, F. Can whatsapp benefit from debunked fact-checked stories to reduce misinformation? In: . [*S. n.*], 2020. Disponível em: https://www.acemap.info/paper/494527547.

REIS, J. C. S.; MELO, P. F.; GARIMELLA, K.; ALMEIDA, J. M.; ECKLES, D.; BENEVENUTO, F. A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In: CHOUDHURY, M. D.; CHUNARA, R.; CULOTTA, A.; WELLES, B. F. (Ed.). **Proceedings of the Fourteenth International AAAI Conference on Web and Social Media, ICWSM 2020, Held Virtually, Original Venue: Atlanta, Georgia, USA, June 8-11, 2020**. AAAI Press, 2020. p. 903–908. Disponível em: https://aaai.org/ojs/index.php/ICWSM/article/view/7356.

RESENDE, G.; MELO, P. F.; SOUSA, H.; MESSIAS, J.; VASCONCELOS, M.; ALMEIDA, J. M.; BENEVENUTO, F. (mis)information dissemination in whatsapp: Gathering, analyzing and countermeasures. In: LIU, L.; WHITE, R. W.; MANTRACH, A.; SILVESTRI, F.; MCAULEY, J. J.; BAEZA-YATES, R.; ZIA, L. (Ed.). **The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019**. ACM, 2019. p. 818–828. Disponível em: https://doi.org/10.1145/3308558.3313688.

RESENDE, G.; MESSIAS, J.; SILVA, M.; ALMEIDA, J.; VASCONCELOS, M.; BENEVENUTO, F. A system for monitoring public political groups in whatsapp. In: **Proceedings of the 24th Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: Association for Computing Machinery, 2018. (WebMedia '18), p. 387–390. ISBN 9781450358675. Disponível em: https://doi.org/10.1145/3243082.3264662.

ROBSON, C.; MCCARTAN, K. **Real world research**. [*S. l.*]: Wiley, 2016.

RUNESON, P.; HÖST, M. Guidelines for conducting and reporting case study research in software engineering. **Empir. Softw. Eng.**, v. 14, n. 2, p. 131–164, 2009. Disponível em: https://doi.org/10.1007/s10664-008-9102-8.

SHU, K.; BERNARD, H. R.; LIU, H. Studying fake news via network analysis: Detection and mitigation. **CoRR**, abs/1804.10233, 2018. Disponível em: http://arxiv.org/abs/1804.10233.

SHU, K.; SLIVA, A.; WANG, S.; TANG, J.; LIU, H. Fake news detection on social media: A data mining perspective. **CoRR**, abs/1708.01967, 2017. Disponível em: http://arxiv.org/abs/1708.01967.

SHU, K.; ZHOU, X.; WANG, S.; ZAFARANI, R.; LIU, H. The role of user profiles for fake news detection. In: **Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining**. New York, NY, USA: Association for Computing Machinery, 2019. (ASONAM '19), p. 436–439. ISBN 9781450368681. Disponível em: https://doi.org/10.1145/3341161.3342927.

SILVA, J.; CALEGARI, N.; GOMES, E. Após a lei geral de proteção de dados do brasil: Autorização em aplicativos da web descentralizados. In: **Anais complementares da conferência da World Wide Web 2019**. [*S. l.*: *s. n.*], 2019. p. 819–822.

SILVA, R. M.; SANTOS, R. L. S.; ALMEIDA, T. A.; PARDO, T. A. S. Towards automatically filtering fake news in portuguese. **Expert Syst. Appl.**, v. 146, p. 113199, 2020. Disponível em: https://doi.org/10.1016/j.eswa.2020.113199.

SU, Q.; WAN, M.; LIU, X.; HUANG, C.-R. Motivations, methods and metrics of misinformation detection: An nlp perspective. **Natural Language Processing Research**, v. 1, p. 1–13, 2020. ISSN 2666-0512. Disponível em: https://doi.org/10.2991/nlpr.d.200522.001.

TCHECHMEDJIEV, A.; FAFALIOS, P.; BOLAND, K.; GASQUET, M.; ZLOCH, M.; ZAPILKO, B.; DIETZE, S.; TODOROV, K. Claimskg: A knowledge graph of fact-checked claims. In: GHIDINI, C.; HARTIG, O.; MALESHKOVA, M.; SVÁTEK, V.; CRUZ, I. F.; HOGAN, A.; SONG, J.; LEFRANÇOIS, M.; GANDON, F. (Ed.). **The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II**. Springer, 2019. (Lecture Notes in Computer Science, v. 11779), p. 309–324. Disponível em: https://doi.org/10.1007/978-3-030-30796-7_20.

VASCONCELOS, M.; PEREIRA, E.; aES, S. G.; RIBEIRO, M. H.; MELO, P.; BENEVENUTO, F. Analyzing youtube videos shared on whatsapp in the early covid-19 crisis. In: **Proceedings of the Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: Association for Computing Machinery, 2020. (WebMedia '20), p. 25–28. ISBN 9781450381963. Disponível em: https://doi.org/10.1145/3428658.3431090.

VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. **Science**, v. 359, p. 1146–1151, 03 2018.

WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M. C.; REGNELL, B. **Experimentation in Software Engineering**. Springer, 2012. ISBN 978-3-642-29043-5. Disponível em: https://doi.org/10.1007/978-3-642-29044-2.

ZHANG, Y.; HARA, T. A probabilistic model for malicious user and rumor detection on social media. In: **53rd Hawaii International Conference on System Sciences, HICSS 2020, Maui, Hawaii, USA, January 7-10, 2020**. ScholarSpace, 2020. p. 1–10. Disponível em: http://hdl.handle.net/10125/64051.

ANNEX A – LGPD

The General Data Protection Law (LGPD), Law No. 13,709, approved on August 14, 2018, aims to protect personal data and the free movement of such data. In addition, LGPD regulates the activities of processing personal data, including in digital media, applied to any person, physical or legal, that performs the processing of personal data, online and/or offline.

Presidência da República
Casa Civil
Subchefia para Assuntos Jurídicos

LEI Nº 13.709, DE 14 DE AGOSTO DE 2018.

Mensagem de veto

Vigência

Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet).

O PRESIDENTE DA REPÚBLICA Faço saber que o Congresso Nacional decreta e eu sanciono a seguinte Lei:

CAPÍTULO I
DISPOSIÇÕES PRELIMINARES

Art. 1º Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

Art. 2º A disciplina da proteção de dados pessoais tem como fundamentos:

I - o respeito à privacidade;

II - a autodeterminação informativa;

III - a liberdade de expressão, de informação, de comunicação e de opinião;

IV - a inviolabilidade da intimidade, da honra e da imagem;

V - o desenvolvimento econômico e tecnológico e a inovação;

VI - a livre iniciativa, a livre concorrência e a defesa do consumidor; e

VII - os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais.

Art. 3º Esta Lei aplica-se a qualquer operação de tratamento realizada por pessoa natural ou por pessoa jurídica de direito público ou privado, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados, desde que:

I - a operação de tratamento seja realizada no território nacional;

II - a atividade de tratamento tenha por objetivo a oferta ou o fornecimento de bens ou serviços ou o tratamento de dados de indivíduos localizados no território nacional;

III - os dados pessoais objeto do tratamento tenham sido coletados no território nacional.

§ 1º Consideram-se coletados no território nacional os dados pessoais cujo titular nele se encontre no momento da coleta.

§ 2º Excetua-se do disposto no inciso I deste artigo o tratamento de dados previsto no inciso IV do caput do

art. 4º desta Lei.

Art. 4º Esta Lei não se aplica ao tratamento de dados pessoais:

I - realizado por pessoa natural para fins exclusivamente particulares e não econômicos;

II - realizado para fins exclusivamente:

a) jornalístico e artísticos; ou

b) acadêmicos, aplicando-se a esta hipótese os arts. 7º e 11 desta Lei;

III - realizado para fins exclusivos de:

a) segurança pública;

b) defesa nacional;

c) segurança do Estado; ou

d) atividades de investigação e repressão de infrações penais; ou

IV - provenientes de fora do território nacional e que não sejam objeto de comunicação, uso compartilhado de dados com agentes de tratamento brasileiros ou objeto de transferência internacional de dados com outro país que não o de proveniência, desde que o país de proveniência proporcione grau de proteção de dados pessoais adequado ao previsto nesta Lei.

§ 1º O tratamento de dados pessoais previsto no inciso III será regido por legislação específica, que deverá prever medidas proporcionais e estritamente necessárias ao atendimento do interesse público, observados o devido processo legal, os princípios gerais de proteção e os direitos do titular previstos nesta Lei.

§ 2º É vedado o tratamento dos dados a que se refere o inciso III do caput deste artigo por pessoa de direito privado, exceto em procedimentos sob tutela de pessoa jurídica de direito público, que serão objeto de informe específico à autoridade nacional e que deverão observar a limitação imposta no § 4º deste artigo.

§ 3º A autoridade nacional emitirá opiniões técnicas ou recomendações referentes às exceções previstas no inciso III do caput deste artigo e deverá solicitar aos responsáveis relatórios de impacto à proteção de dados pessoais.

§ 4º Em nenhum caso a totalidade dos dados pessoais de banco de dados de que trata o inciso III do caput deste artigo poderá ser tratada por pessoa de direito privado.

Art. 5º Para os fins desta Lei, considera-se:

I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;

II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;

III - dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;

IV - banco de dados: conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em

suporte eletrônico ou físico;

V - titular: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;

VI - controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;

VII - operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;

VIII - encarregado: pessoa natural, indicada pelo controlador, que atua como canal de comunicação entre o controlador e os titulares e a autoridade nacional;

IX - agentes de tratamento: o controlador e o operador;

X - tratamento: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;

XI - anonimização: utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo;

XII - consentimento: manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para uma finalidade determinada;

XIII - bloqueio: suspensão temporária de qualquer operação de tratamento, mediante guarda do dado pessoal ou do banco de dados;

XIV - eliminação: exclusão de dado ou de conjunto de dados armazenados em banco de dados, independentemente do procedimento empregado;

XV - transferência internacional de dados: transferência de dados pessoais para país estrangeiro ou organismo internacional do qual o país seja membro;

XVI - uso compartilhado de dados: comunicação, difusão, transferência internacional, interconexão de dados pessoais ou tratamento compartilhado de bancos de dados pessoais por órgãos e entidades públicos no cumprimento de suas competências legais, ou entre esses e entes privados, reciprocamente, com autorização específica, para uma ou mais modalidades de tratamento permitidas por esses entes públicos, ou entre entes privados;

XVII - relatório de impacto à proteção de dados pessoais: documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco;

XVIII - órgão de pesquisa: órgão ou entidade da administração pública direta ou indireta ou pessoa jurídica de direito privado sem fins lucrativos legalmente constituída sob as leis brasileiras, com sede e foro no País, que inclua em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico;

XIX - autoridade nacional: órgão da administração pública indireta responsável por zelar, implementar e fiscalizar o cumprimento desta Lei.

Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:

I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular,

sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;

II - adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;

III - necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;

IV - livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;

V - qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;

VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;

VII - segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;

VIII - prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;

IX - não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;

X - responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

CAPÍTULO II
DO TRATAMENTO DE DADOS PESSOAIS

Seção I
Dos Requisitos para o Tratamento de Dados Pessoais

Art. 7º O tratamento de dados pessoais somente poderá ser realizado nas seguintes hipóteses:

I - mediante o fornecimento de consentimento pelo titular;

II - para o cumprimento de obrigação legal ou regulatória pelo controlador;

III - pela administração pública, para o tratamento e uso compartilhado de dados necessários à execução de políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres, observadas as disposições do Capítulo IV desta Lei;

IV - para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;

V - quando necessário para a execução de contrato ou de procedimentos preliminares relacionados a contrato do qual seja parte o titular, a pedido do titular dos dados;

VI - para o exercício regular de direitos em processo judicial, administrativo ou arbitral, esse último nos termos

da Lei nº 9.307, de 23 de setembro de 1996 (Lei de Arbitragem);

VII - para a proteção da vida ou da incolumidade física do titular ou de terceiro;

VIII - para a tutela da saúde, em procedimento realizado por profissionais da área da saúde ou por entidades sanitárias;

IX - quando necessário para atender aos interesses legítimos do controlador ou de terceiro, exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais; ou

X - para a proteção do crédito, inclusive quanto ao disposto na legislação pertinente.

§ 1º Nos casos de aplicação do disposto nos incisos II e III do caput deste artigo e excetuadas as hipóteses previstas no art. 4º desta Lei, o titular será informado das hipóteses em que será admitido o tratamento de seus dados.

§ 2º A forma de disponibilização das informações previstas no § 1º e no inciso I do caput do art. 23 desta Lei poderá ser especificada pela autoridade nacional.

§ 3º O tratamento de dados pessoais cujo acesso é público deve considerar a finalidade, a boa-fé e o interesse público que justificaram sua disponibilização.

§ 4º É dispensada a exigência do consentimento previsto no caput deste artigo para os dados tornados manifestamente públicos pelo titular, resguardados os direitos do titular e os princípios previstos nesta Lei.

§ 5º O controlador que obteve o consentimento referido no inciso I do caput deste artigo que necessitar comunicar ou compartilhar dados pessoais com outros controladores deverá obter consentimento específico do titular para esse fim, ressalvadas as hipóteses de dispensa do consentimento previstas nesta Lei.

§ 6º A eventual dispensa da exigência do consentimento não desobriga os agentes de tratamento das demais obrigações previstas nesta Lei, especialmente da observância dos princípios gerais e da garantia dos direitos do titular.

Art. 8º O consentimento previsto no inciso I do art. 7º desta Lei deverá ser fornecido por escrito ou por outro meio que demonstre a manifestação de vontade do titular.

§ 1º Caso o consentimento seja fornecido por escrito, esse deverá constar de cláusula destacada das demais cláusulas contratuais.

§ 2º Cabe ao controlador o ônus da prova de que o consentimento foi obtido em conformidade com o disposto nesta Lei.

§ 3º É vedado o tratamento de dados pessoais mediante vício de consentimento.

§ 4º O consentimento deverá referir-se a finalidades determinadas, e as autorizações genéricas para o tratamento de dados pessoais serão nulas.

§ 5º O consentimento pode ser revogado a qualquer momento mediante manifestação expressa do titular, por procedimento gratuito e facilitado, ratificados os tratamentos realizados sob amparo do consentimento anteriormente manifestado enquanto não houver requerimento de eliminação, nos termos do inciso VI do caput do art. 18 desta Lei.

§ 6º Em caso de alteração de informação referida nos incisos I, II, III ou V do art. 9º desta Lei, o controlador deverá informar ao titular, com destaque de forma específica do teor das alterações, podendo o titular, nos casos em que o seu consentimento é exigido, revogá-lo caso discorde da alteração.

Art. 9º O titular tem direito ao acesso facilitado às informações sobre o tratamento de seus dados, que deverão ser disponibilizadas de forma clara, adequada e ostensiva acerca de, entre outras características previstas em regulamentação para o atendimento do princípio do livre acesso:

I - finalidade específica do tratamento;

II - forma e duração do tratamento, observados os segredos comercial e industrial;

III - identificação do controlador;

IV - informações de contato do controlador;

V - informações acerca do uso compartilhado de dados pelo controlador e a finalidade;

VI - responsabilidades dos agentes que realizarão o tratamento; e

VII - direitos do titular, com menção explícita aos direitos contidos no art. 18 desta Lei.

§ 1º Na hipótese em que o consentimento é requerido, esse será considerado nulo caso as informações fornecidas ao titular tenham conteúdo enganoso ou abusivo ou não tenham sido apresentadas previamente com transparência, de forma clara e inequívoca.

§ 2º Na hipótese em que o consentimento é requerido, se houver mudanças da finalidade para o tratamento de dados pessoais não compatíveis com o consentimento original, o controlador deverá informar previamente o titular sobre as mudanças de finalidade, podendo o titular revogar o consentimento, caso discorde das alterações.

§ 3º Quando o tratamento de dados pessoais for condição para o fornecimento de produto ou de serviço ou para o exercício de direito, o titular será informado com destaque sobre esse fato e sobre os meios pelos quais poderá exercer os direitos do titular elencados no art. 18 desta Lei.

Art. 10. O legítimo interesse do controlador somente poderá fundamentar tratamento de dados pessoais para finalidades legítimas, consideradas a partir de situações concretas, que incluem, mas não se limitam a:

I - apoio e promoção de atividades do controlador; e

II - proteção, em relação ao titular, do exercício regular de seus direitos ou prestação de serviços que o beneficiem, respeitadas as legítimas expectativas dele e os direitos e liberdades fundamentais, nos termos desta Lei.

§ 1º Quando o tratamento for baseado no legítimo interesse do controlador, somente os dados pessoais estritamente necessários para a finalidade pretendida poderão ser tratados.

§ 2º O controlador deverá adotar medidas para garantir a transparência do tratamento de dados baseado em seu legítimo interesse.

§ 3º A autoridade nacional poderá solicitar ao controlador relatório de impacto à proteção de dados pessoais, quando o tratamento tiver como fundamento seu interesse legítimo, observados os segredos comercial e industrial.

Seção II
Do Tratamento de Dados Pessoais Sensíveis

Art. 11. O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses:

I - quando o titular ou seu responsável legal consentir, de forma específica e destacada, para finalidades específicas;

II - sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para:

a) cumprimento de obrigação legal ou regulatória pelo controlador;

b) tratamento compartilhado de dados necessários à execução, pela administração pública, de políticas públicas previstas em leis ou regulamentos;

c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis;

d) exercício regular de direitos, inclusive em contrato e em processo judicial, administrativo e arbitral, este último nos termos da Lei nº 9.307, de 23 de setembro de 1996 (Lei de Arbitragem);

e) proteção da vida ou da incolumidade física do titular ou de terceiro;

f) tutela da saúde, em procedimento realizado por profissionais da área da saúde ou por entidades sanitárias; ou

g) garantia da prevenção à fraude e à segurança do titular, nos processos de identificação e autenticação de cadastro em sistemas eletrônicos, resguardados os direitos mencionados no art. 9º desta Lei e exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais.

§ 1º Aplica-se o disposto neste artigo a qualquer tratamento de dados pessoais que revele dados pessoais sensíveis e que possa causar dano ao titular, ressalvado o disposto em legislação específica.

§ 2º Nos casos de aplicação do disposto nas alíneas “a” e “b” do inciso II do caput deste artigo pelos órgãos e pelas entidades públicas, será dada publicidade à referida dispensa de consentimento, nos termos do inciso I do caput do art. 23 desta Lei.

§ 3º A comunicação ou o uso compartilhado de dados pessoais sensíveis entre controladores com objetivo de obter vantagem econômica poderá ser objeto de vedação ou de regulamentação por parte da autoridade nacional, ouvidos os órgãos setoriais do Poder Público, no âmbito de suas competências.

§ 4º É vedada a comunicação ou o uso compartilhado entre controladores de dados pessoais sensíveis referentes à saúde com objetivo de obter vantagem econômica, exceto nos casos de portabilidade de dados quando consentido pelo titular.

Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.

§ 1º A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o processo de anonimização, de acordo com as tecnologias disponíveis, e a utilização exclusiva de meios próprios.

§ 2º Poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada.

§ 3º A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.

Art. 13. Na realização de estudos em saúde pública, os órgãos de pesquisa poderão ter acesso a bases de dados pessoais, que serão tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de estudos e pesquisas e mantidos em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudonimização dos dados, bem

como considerem os devidos padrões éticos relacionados a estudos e pesquisas.

§ 1º A divulgação dos resultados ou de qualquer excerto do estudo ou da pesquisa de que trata o caput deste artigo em nenhuma hipótese poderá revelar dados pessoais.

§ 2º O órgão de pesquisa será o responsável pela segurança da informação prevista no caput deste artigo, não permitida, em circunstância alguma, a transferência dos dados a terceiro.

§ 3º O acesso aos dados de que trata este artigo será objeto de regulamentação por parte da autoridade nacional e das autoridades da área de saúde e sanitárias, no âmbito de suas competências.

§ 4º Para os efeitos deste artigo, a pseudonimização é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro.

Seção III
Do Tratamento de Dados Pessoais de Crianças e de Adolescentes

Art. 14. O tratamento de dados pessoais de crianças e de adolescentes deverá ser realizado em seu melhor interesse, nos termos deste artigo e da legislação pertinente.

§ 1º O tratamento de dados pessoais de crianças deverá ser realizado com o consentimento específico e em destaque dado por pelo menos um dos pais ou pelo responsável legal.

§ 2º No tratamento de dados de que trata o § 1º deste artigo, os controladores deverão manter pública a informação sobre os tipos de dados coletados, a forma de sua utilização e os procedimentos para o exercício dos direitos a que se refere o art. 18 desta Lei.

§ 3º Poderão ser coletados dados pessoais de crianças sem o consentimento a que se refere o § 1º deste artigo quando a coleta for necessária para contatar os pais ou o responsável legal, utilizados uma única vez e sem armazenamento, ou para sua proteção, e em nenhum caso poderão ser repassados a terceiro sem o consentimento de que trata o § 1º deste artigo.

§ 4º Os controladores não deverão condicionar a participação dos titulares de que trata o § 1º deste artigo em jogos, aplicações de internet ou outras atividades ao fornecimento de informações pessoais além das estritamente necessárias à atividade.

§ 5º O controlador deve realizar todos os esforços razoáveis para verificar que o consentimento a que se refere o § 1º deste artigo foi dado pelo responsável pela criança, consideradas as tecnologias disponíveis.

§ 6º As informações sobre o tratamento de dados referidas neste artigo deverão ser fornecidas de maneira simples, clara e acessível, consideradas as características físico-motoras, perceptivas, sensoriais, intelectuais e mentais do usuário, com uso de recursos audiovisuais quando adequado, de forma a proporcionar a informação necessária aos pais ou ao responsável legal e adequada ao entendimento da criança.

Seção IV
Do Término do Tratamento de Dados

Art. 15. O término do tratamento de dados pessoais ocorrerá nas seguintes hipóteses:

I - verificação de que a finalidade foi alcançada ou de que os dados deixaram de ser necessários ou pertinentes ao alcance da finalidade específica almejada;

II - fim do período de tratamento;

III - comunicação do titular, inclusive no exercício de seu direito de revogação do consentimento conforme disposto no § 5º do art. 8º desta Lei, resguardado o interesse público; ou

IV - determinação da autoridade nacional, quando houver violação ao disposto nesta Lei.

Art. 16. Os dados pessoais serão eliminados após o término de seu tratamento, no âmbito e nos limites técnicos das atividades, autorizada a conservação para as seguintes finalidades:

I - cumprimento de obrigação legal ou regulatória pelo controlador;

II - estudo por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;

III - transferência a terceiro, desde que respeitados os requisitos de tratamento de dados dispostos nesta Lei; ou

IV - uso exclusivo do controlador, vedado seu acesso por terceiro, e desde que anonimizados os dados.

CAPÍTULO III
DOS DIREITOS DO TITULAR

Art. 17. Toda pessoa natural tem assegurada a titularidade de seus dados pessoais e garantidos os direitos fundamentais de liberdade, de intimidade e de privacidade, nos termos desta Lei.

Art. 18. O titular dos dados pessoais tem direito a obter do controlador, em relação aos dados do titular por ele tratados, a qualquer momento e mediante requisição:

I - confirmação da existência de tratamento;

II - acesso aos dados;

III - correção de dados incompletos, inexatos ou desatualizados;

IV - anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou tratados em desconformidade com o disposto nesta Lei;

V - portabilidade dos dados a outro fornecedor de serviço ou produto, mediante requisição expressa e observados os segredos comercial e industrial, de acordo com a regulamentação do órgão controlador;

VI - eliminação dos dados pessoais tratados com o consentimento do titular, exceto nas hipóteses previstas no art. 16 desta Lei;

VII - informação das entidades públicas e privadas com as quais o controlador realizou uso compartilhado de dados;

VIII - informação sobre a possibilidade de não fornecer consentimento e sobre as consequências da negativa;

IX - revogação do consentimento, nos termos do § 5º do art. 8º desta Lei.

§ 1º O titular dos dados pessoais tem o direito de peticionar em relação aos seus dados contra o controlador perante a autoridade nacional.

§ 2º O titular pode opor-se a tratamento realizado com fundamento em uma das hipóteses de dispensa de consentimento, em caso de descumprimento ao disposto nesta Lei.

§ 3º Os direitos previstos neste artigo serão exercidos mediante requerimento expresso do titular ou de representante legalmente constituído, a agente de tratamento.

§ 4º Em caso de impossibilidade de adoção imediata da providência de que trata o § 3º deste artigo, o controlador enviará ao titular resposta em que poderá:

I - comunicar que não é agente de tratamento dos dados e indicar, sempre que possível, o agente; ou

II - indicar as razões de fato ou de direito que impedem a adoção imediata da providência.

§ 5º O requerimento referido no § 3º deste artigo será atendido sem custos para o titular, nos prazos e nos termos previstos em regulamento.

§ 6º O responsável deverá informar de maneira imediata aos agentes de tratamento com os quais tenha realizado uso compartilhado de dados a correção, a eliminação, a anonimização ou o bloqueio dos dados, para que repitam idêntico procedimento.

§ 7º A portabilidade dos dados pessoais a que se refere o inciso V do caput deste artigo não inclui dados que já tenham sido anonimizados pelo controlador.

§ 8º O direito a que se refere o § 1º deste artigo também poderá ser exercido perante os organismos de defesa do consumidor.

Art. 19. A confirmação de existência ou o acesso a dados pessoais serão providenciados, mediante requisição do titular:

I - em formato simplificado, imediatamente; ou

II - por meio de declaração clara e completa, que indique a origem dos dados, a inexistência de registro, os critérios utilizados e a finalidade do tratamento, observados os segredos comercial e industrial, fornecida no prazo de até 15 (quinze) dias, contado da data do requerimento do titular.

§ 1º Os dados pessoais serão armazenados em formato que favoreça o exercício do direito de acesso.

§ 2º As informações e os dados poderão ser fornecidos, a critério do titular:

I - por meio eletrônico, seguro e idôneo para esse fim; ou

II - sob forma impressa.

§ 3º Quando o tratamento tiver origem no consentimento do titular ou em contrato, o titular poderá solicitar cópia eletrônica integral de seus dados pessoais, observados os segredos comercial e industrial, nos termos de regulamentação da autoridade nacional, em formato que permita a sua utilização subsequente, inclusive em outras operações de tratamento.

§ 4º A autoridade nacional poderá dispor de forma diferenciada acerca dos prazos previstos nos incisos I e II do caput deste artigo para os setores específicos.

Art. 20. O titular dos dados tem direito a solicitar revisão, por pessoa natural, de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, inclusive de decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

§ 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.

§ 2º Em caso de não oferecimento de informações de que trata o § 1º deste artigo baseado na observância de segredo comercial e industrial, a autoridade nacional poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais.

Art. 21. Os dados pessoais referentes ao exercício regular de direitos pelo titular não podem ser utilizados em seu prejuízo.

Art. 22. A defesa dos interesses e dos direitos dos titulares de dados poderá ser exercida em juízo, individual ou coletivamente, na forma do disposto na legislação pertinente, acerca dos instrumentos de tutela individual e coletiva.

CAPÍTULO IV
DO TRATAMENTO DE DADOS PESSOAIS PELO PODER PÚBLICO

Seção I
Das Regras

Art. 23. O tratamento de dados pessoais pelas pessoas jurídicas de direito público referidas no parágrafo único do art. 1º da Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), deverá ser realizado para o atendimento de sua finalidade pública, na persecução do interesse público, com o objetivo de executar as competências legais ou cumprir as atribuições legais do serviço público, desde que:

I - sejam informadas as hipóteses em que, no exercício de suas competências, realizam o tratamento de dados pessoais, fornecendo informações claras e atualizadas sobre a previsão legal, a finalidade, os procedimentos e as práticas utilizadas para a execução dessas atividades, em veículos de fácil acesso, preferencialmente em seus sítios eletrônicos;

II - (VETADO); e

III - seja indicado um encarregado quando realizarem operações de tratamento de dados pessoais, nos termos do art. 39 desta Lei.

§ 1º A autoridade nacional poderá dispor sobre as formas de publicidade das operações de tratamento.

§ 2º O disposto nesta Lei não dispensa as pessoas jurídicas mencionadas no caput deste artigo de instituir as autoridades de que trata a Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação).

§ 3º Os prazos e procedimentos para exercício dos direitos do titular perante o Poder Público observarão o disposto em legislação específica, em especial as disposições constantes da Lei nº 9.507, de 12 de novembro de 1997 (Lei do Habeas Data), da Lei nº 9.784, de 29 de janeiro de 1999 (Lei Geral do Processo Administrativo), e da Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação).

§ 4º Os serviços notariais e de registro exercidos em caráter privado, por delegação do Poder Público, terão o mesmo tratamento dispensado às pessoas jurídicas referidas no caput deste artigo, nos termos desta Lei.

§ 5º Os órgãos notariais e de registro devem fornecer acesso aos dados por meio eletrônico para a administração pública, tendo em vista as finalidades de que trata o caput deste artigo.

Art. 24. As empresas públicas e as sociedades de economia mista que atuam em regime de concorrência, sujeitas ao disposto no art. 173 da Constituição Federal, terão o mesmo tratamento dispensado às pessoas jurídicas de direito privado particulares, nos termos desta Lei.

Parágrafo único. As empresas públicas e as sociedades de economia mista, quando estiverem operacionalizando políticas públicas e no âmbito da execução delas, terão o mesmo tratamento dispensado aos órgãos e às entidades do Poder Público, nos termos deste Capítulo.

Art. 25. Os dados deverão ser mantidos em formato interoperável e estruturado para o uso compartilhado, com vistas à execução de políticas públicas, à prestação de serviços públicos, à descentralização da atividade pública e à disseminação e ao acesso das informações pelo público em geral.

Art. 26. O uso compartilhado de dados pessoais pelo Poder Público deve atender a finalidades específicas de execução de políticas públicas e atribuição legal pelos órgãos e pelas entidades públicas, respeitados os princípios de proteção de dados pessoais elencados no art. 6º desta Lei.

§ 1º É vedado ao Poder Público transferir a entidades privadas dados pessoais constantes de bases de dados a que tenha acesso, exceto:

I - em casos de execução descentralizada de atividade pública que exija a transferência, exclusivamente para esse fim específico e determinado, observado o disposto na Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação);

II - (VETADO);

III - nos casos em que os dados forem acessíveis publicamente, observadas as disposições desta Lei.

§ 2º Os contratos e convênios de que trata o § 1º deste artigo deverão ser comunicados à autoridade nacional.

Art. 27. A comunicação ou o uso compartilhado de dados pessoais de pessoa jurídica de direito público a pessoa de direito privado será informado à autoridade nacional e dependerá de consentimento do titular, exceto:

I - nas hipóteses de dispensa de consentimento previstas nesta Lei;

II - nos casos de uso compartilhado de dados, em que será dada publicidade nos termos do inciso I do caput do art. 23 desta Lei; ou

III - nas exceções constantes do § 1º do art. 26 desta Lei.

Art. 28. (VETADO).

Art. 29. A autoridade nacional poderá solicitar, a qualquer momento, às entidades do Poder Público, a realização de operações de tratamento de dados pessoais, informe específico sobre o âmbito e a natureza dos dados e demais detalhes do tratamento realizado e poderá emitir parecer técnico complementar para garantir o cumprimento desta Lei.

Art. 30. A autoridade nacional poderá estabelecer normas complementares para as atividades de comunicação e de uso compartilhado de dados pessoais.

Seção II
Da Responsabilidade

Art. 31. Quando houver infração a esta Lei em decorrência do tratamento de dados pessoais por órgãos públicos, a autoridade nacional poderá enviar informe com medidas cabíveis para fazer cessar a violação.

Art. 32. A autoridade nacional poderá solicitar a agentes do Poder Público a publicação de relatórios de impacto à proteção de dados pessoais e sugerir a adoção de padrões e de boas práticas para os tratamentos de dados pessoais pelo Poder Público.

CAPÍTULO V
DA TRANSFERÊNCIA INTERNACIONAL DE DADOS

Art. 33. A transferência internacional de dados pessoais somente é permitida nos seguintes casos:

I - para países ou organismos internacionais que proporcionem grau de proteção de dados pessoais adequado ao previsto nesta Lei;

II - quando o controlador oferecer e comprovar garantias de cumprimento dos princípios, dos direitos do titular e do regime de proteção de dados previstos nesta Lei, na forma de:

a) cláusulas contratuais específicas para determinada transferência;

b) cláusulas-padrão contratuais;

c) normas corporativas globais;

d) selos, certificados e códigos de conduta regularmente emitidos;

III - quando a transferência for necessária para a cooperação jurídica internacional entre órgãos públicos de inteligência, de investigação e de persecução, de acordo com os instrumentos de direito internacional;

IV - quando a transferência for necessária para a proteção da vida ou da incolumidade física do titular ou de terceiro;

V - quando a autoridade nacional autorizar a transferência;

VI - quando a transferência resultar em compromisso assumido em acordo de cooperação internacional;

VII - quando a transferência for necessária para a execução de política pública ou atribuição legal do serviço público, sendo dada publicidade nos termos do inciso I do caput do art. 23 desta Lei;

VIII - quando o titular tiver fornecido o seu consentimento específico e em destaque para a transferência, com informação prévia sobre o caráter internacional da operação, distinguindo claramente esta de outras finalidades; ou

IX - quando necessário para atender as hipóteses previstas nos incisos II, V e VI do art. 7º desta Lei.

Parágrafo único. Para os fins do inciso I deste artigo, as pessoas jurídicas de direito público referidas no parágrafo único do art. 1º da Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), no âmbito de suas competências legais, e responsáveis, no âmbito de suas atividades, poderão requerer à autoridade nacional a avaliação do nível de proteção a dados pessoais conferido por país ou organismo internacional.

Art. 34. O nível de proteção de dados do país estrangeiro ou do organismo internacional mencionado no inciso I do caput do art. 33 desta Lei será avaliado pela autoridade nacional, que levará em consideração:

I - as normas gerais e setoriais da legislação em vigor no país de destino ou no organismo internacional;

II - a natureza dos dados;

III - a observância dos princípios gerais de proteção de dados pessoais e direitos dos titulares previstos nesta Lei;

IV - a adoção de medidas de segurança previstas em regulamento;

V - a existência de garantias judiciais e institucionais para o respeito aos direitos de proteção de dados pessoais; e

VI - outras circunstâncias específicas relativas à transferência.

Art. 35. A definição do conteúdo de cláusulas-padrão contratuais, bem como a verificação de cláusulas contratuais específicas para uma determinada transferência, normas corporativas globais ou selos, certificados e códigos de conduta, a que se refere o inciso II do caput do art. 33 desta Lei, será realizada pela autoridade nacional.

§ 1º Para a verificação do disposto no caput deste artigo, deverão ser considerados os requisitos, as condições e as garantias mínimas para a transferência que observem os direitos, as garantias e os princípios desta Lei.

§ 2º Na análise de cláusulas contratuais, de documentos ou de normas corporativas globais submetidas à aprovação da autoridade nacional, poderão ser requeridas informações suplementares ou realizadas diligências de verificação quanto às operações de tratamento, quando necessário.

§ 3º A autoridade nacional poderá designar organismos de certificação para a realização do previsto no caput deste artigo, que permanecerão sob sua fiscalização nos termos definidos em regulamento.

§ 4º Os atos realizados por organismo de certificação poderão ser revistos pela autoridade nacional e, caso em desconformidade com esta Lei, submetidos a revisão ou anulados.

§ 5º As garantias suficientes de observância dos princípios gerais de proteção e dos direitos do titular referidas no caput deste artigo serão também analisadas de acordo com as medidas técnicas e organizacionais adotadas pelo operador, de acordo com o previsto nos §§ 1º e 2º do art. 46 desta Lei.

Art. 36. As alterações nas garantias apresentadas como suficientes de observância dos princípios gerais de proteção e dos direitos do titular referidas no inciso II do art. 33 desta Lei deverão ser comunicadas à autoridade nacional.

CAPÍTULO VI
DOS AGENTES DE TRATAMENTO DE DADOS PESSOAIS

Seção I
Do Controlador e do Operador

Art. 37. O controlador e o operador devem manter registro das operações de tratamento de dados pessoais que realizarem, especialmente quando baseado no legítimo interesse.

Art. 38. A autoridade nacional poderá determinar ao controlador que elabore relatório de impacto à proteção de dados pessoais, inclusive de dados sensíveis, referente a suas operações de tratamento de dados, nos termos de regulamento, observados os segredos comercial e industrial.

Parágrafo único. Observado o disposto no caput deste artigo, o relatório deverá conter, no mínimo, a descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados.

Art. 39. O operador deverá realizar o tratamento segundo as instruções fornecidas pelo controlador, que verificará a observância das próprias instruções e das normas sobre a matéria.

Art. 40. A autoridade nacional poderá dispor sobre padrões de interoperabilidade para fins de portabilidade, livre acesso aos dados e segurança, assim como sobre o tempo de guarda dos registros, tendo em vista especialmente a necessidade e a transparência.

Seção II
Do Encarregado pelo Tratamento de Dados Pessoais

Art. 41. O controlador deverá indicar encarregado pelo tratamento de dados pessoais.

§ 1º A identidade e as informações de contato do encarregado deverão ser divulgadas publicamente, de forma clara e objetiva, preferencialmente no sítio eletrônico do controlador.

§ 2º As atividades do encarregado consistem em:

I - aceitar reclamações e comunicações dos titulares, prestar esclarecimentos e adotar providências;

II - receber comunicações da autoridade nacional e adotar providências;

III - orientar os funcionários e os contratados da entidade a respeito das práticas a serem tomadas em relação à proteção de dados pessoais; e

IV - executar as demais atribuições determinadas pelo controlador ou estabelecidas em normas complementares.

§ 3º A autoridade nacional poderá estabelecer normas complementares sobre a definição e as atribuições do encarregado, inclusive hipóteses de dispensa da necessidade de sua indicação, conforme a natureza e o porte da entidade ou o volume de operações de tratamento de dados.

Seção III
Da Responsabilidade e do Ressarcimento de Danos

Art. 42. O controlador ou o operador que, em razão do exercício de atividade de tratamento de dados pessoais, causar a outrem dano patrimonial, moral, individual ou coletivo, em violação à legislação de proteção de dados pessoais, é obrigado a repará-lo.

§ 1º A fim de assegurar a efetiva indenização ao titular dos dados:

I - o operador responde solidariamente pelos danos causados pelo tratamento quando descumprir as obrigações da legislação de proteção de dados ou quando não tiver seguido as instruções lícitas do controlador, hipótese em que o operador equipara-se ao controlador, salvo nos casos de exclusão previstos no art. 43 desta Lei;

II - os controladores que estiverem diretamente envolvidos no tratamento do qual decorreram danos ao titular dos dados respondem solidariamente, salvo nos casos de exclusão previstos no art. 43 desta Lei.

§ 2º O juiz, no processo civil, poderá inverter o ônus da prova a favor do titular dos dados quando, a seu juízo, for verossímil a alegação, houver hipossuficiência para fins de produção de prova ou quando a produção de prova pelo titular resultar-lhe excessivamente onerosa.

§ 3º As ações de reparação por danos coletivos que tenham por objeto a responsabilização nos termos do caput deste artigo podem ser exercidas coletivamente em juízo, observado o disposto na legislação pertinente.

§ 4º Aquele que reparar o dano ao titular tem direito de regresso contra os demais responsáveis, na medida de sua participação no evento danoso.

Art. 43. Os agentes de tratamento só não serão responsabilizados quando provarem:

I - que não realizaram o tratamento de dados pessoais que lhes é atribuído;

II - que, embora tenham realizado o tratamento de dados pessoais que lhes é atribuído, não houve violação à legislação de proteção de dados; ou

III - que o dano é decorrente de culpa exclusiva do titular dos dados ou de terceiro.

Art. 44. O tratamento de dados pessoais será irregular quando deixar de observar a legislação ou quando não fornecer a segurança que o titular dele pode esperar, consideradas as circunstâncias relevantes, entre as quais:

I - o modo pelo qual é realizado;

II - o resultado e os riscos que razoavelmente dele se esperam;

III - as técnicas de tratamento de dados pessoais disponíveis à época em que foi realizado.

Parágrafo único. Responde pelos danos decorrentes da violação da segurança dos dados o controlador ou o operador que, ao deixar de adotar as medidas de segurança previstas no art. 46 desta Lei, der causa ao dano.

Art. 45. As hipóteses de violação do direito do titular no âmbito das relações de consumo permanecem sujeitas às regras de responsabilidade previstas na legislação pertinente.

CAPÍTULO VII
DA SEGURANÇA E DAS BOAS PRÁTICAS

Seção I
Da Segurança e do Sigilo de Dados

Art. 46. Os agentes de tratamento devem adotar medidas de segurança, técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito.

§ 1º A autoridade nacional poderá dispor sobre padrões técnicos mínimos para tornar aplicável o disposto no caput deste artigo, considerados a natureza das informações tratadas, as características específicas do tratamento e o estado atual da tecnologia, especialmente no caso de dados pessoais sensíveis, assim como os princípios previstos no caput do art. 6º desta Lei.

§ 2º As medidas de que trata o caput deste artigo deverão ser observadas desde a fase de concepção do produto ou do serviço até a sua execução.

Art. 47. Os agentes de tratamento ou qualquer outra pessoa que intervenha em uma das fases do tratamento obriga-se a garantir a segurança da informação prevista nesta Lei em relação aos dados pessoais, mesmo após o seu término.

Art. 48. O controlador deverá comunicar à autoridade nacional e ao titular a ocorrência de incidente de segurança que possa acarretar risco ou dano relevante aos titulares.

§ 1º A comunicação será feita em prazo razoável, conforme definido pela autoridade nacional, e deverá mencionar, no mínimo:

I - a descrição da natureza dos dados pessoais afetados;

II - as informações sobre os titulares envolvidos;

III - a indicação das medidas técnicas e de segurança utilizadas para a proteção dos dados, observados os segredos comercial e industrial;

IV - os riscos relacionados ao incidente;

V - os motivos da demora, no caso de a comunicação não ter sido imediata; e

VI - as medidas que foram ou que serão adotadas para reverter ou mitigar os efeitos do prejuízo.

§ 2º A autoridade nacional verificará a gravidade do incidente e poderá, caso necessário para a salvaguarda dos direitos dos titulares, determinar ao controlador a adoção de providências, tais como:

I - ampla divulgação do fato em meios de comunicação; e

II - medidas para reverter ou mitigar os efeitos do incidente.

§ 3º No juízo de gravidade do incidente, será avaliada eventual comprovação de que foram adotadas medidas técnicas adequadas que tornem os dados pessoais afetados ininteligíveis, no âmbito e nos limites técnicos de seus serviços, para terceiros não autorizados a acessá-los.

Art. 49. Os sistemas utilizados para o tratamento de dados pessoais devem ser estruturados de forma a atender aos requisitos de segurança, aos padrões de boas práticas e de governança e aos princípios gerais previstos nesta Lei e às demais normas regulamentares.

Seção II
Das Boas Práticas e da Governança

Art. 50. Os controladores e operadores, no âmbito de suas competências, pelo tratamento de dados pessoais, individualmente ou por meio de associações, poderão formular regras de boas práticas e de governança que estabeleçam as condições de organização, o regime de funcionamento, os procedimentos, incluindo reclamações e petições de titulares, as normas de segurança, os padrões técnicos, as obrigações específicas para os diversos envolvidos no tratamento, as ações educativas, os mecanismos internos de supervisão e de mitigação de riscos e outros aspectos relacionados ao tratamento de dados pessoais.

§ 1º Ao estabelecer regras de boas práticas, o controlador e o operador levarão em consideração, em relação ao tratamento e aos dados, a natureza, o escopo, a finalidade e a probabilidade e a gravidade dos riscos e dos benefícios decorrentes de tratamento de dados do titular.

§ 2º Na aplicação dos princípios indicados nos incisos VII e VIII do caput do art. 6º desta Lei, o controlador, observados a estrutura, a escala e o volume de suas operações, bem como a sensibilidade dos dados tratados e a probabilidade e a gravidade dos danos para os titulares dos dados, poderá:

I - implementar programa de governança em privacidade que, no mínimo:

a) demonstre o comprometimento do controlador em adotar processos e políticas internas que assegurem o cumprimento, de forma abrangente, de normas e boas práticas relativas à proteção de dados pessoais;

b) seja aplicável a todo o conjunto de dados pessoais que estejam sob seu controle, independentemente do modo como se realizou sua coleta;

c) seja adaptado à estrutura, à escala e ao volume de suas operações, bem como à sensibilidade dos dados tratados;

d) estabeleça políticas e salvaguardas adequadas com base em processo de avaliação sistemática de impactos e riscos à privacidade;

e) tenha o objetivo de estabelecer relação de confiança com o titular, por meio de atuação transparente e que assegure mecanismos de participação do titular;

f) esteja integrado a sua estrutura geral de governança e estabeleça e aplique mecanismos de supervisão internos e externos;

g) conte com planos de resposta a incidentes e remediação; e

h) seja atualizado constantemente com base em informações obtidas a partir de monitoramento contínuo e avaliações periódicas;

II - demonstrar a efetividade de seu programa de governança em privacidade quando apropriado e, em especial, a pedido da autoridade nacional ou de outra entidade responsável por promover o cumprimento de boas práticas ou códigos de conduta, os quais, de forma independente, promovam o cumprimento desta Lei.

§ 3º As regras de boas práticas e de governança deverão ser publicadas e atualizadas periodicamente e poderão ser reconhecidas e divulgadas pela autoridade nacional.

Art. 51. A autoridade nacional estimulará a adoção de padrões técnicos que facilitem o controle pelos titulares dos seus dados pessoais.

CAPÍTULO VIII
DA FISCALIZAÇÃO

Seção I
Das Sanções Administrativas

Art. 52. Os agentes de tratamento de dados, em razão das infrações cometidas às normas previstas nesta Lei, ficam sujeitos às seguintes sanções administrativas aplicáveis pela autoridade nacional:

I - advertência, com indicação de prazo para adoção de medidas corretivas;

II - multa simples, de até 2% (dois por cento) do faturamento da pessoa jurídica de direito privado, grupo ou conglomerado no Brasil no seu último exercício, excluídos os tributos, limitada, no total, a R$ 50.000.000,00 (cinquenta milhões de reais) por infração;

III - multa diária, observado o limite total a que se refere o inciso II;

IV - publicização da infração após devidamente apurada e confirmada a sua ocorrência;

V - bloqueio dos dados pessoais a que se refere a infração até a sua regularização;

VI - eliminação dos dados pessoais a que se refere a infração;

VII - (VETADO);

VIII - (VETADO);

IX - (VETADO).

§ 1º As sanções serão aplicadas após procedimento administrativo que possibilite a oportunidade da ampla defesa, de forma gradativa, isolada ou cumulativa, de acordo com as peculiaridades do caso concreto e considerados os seguintes parâmetros e critérios:

I - a gravidade e a natureza das infrações e dos direitos pessoais afetados;

II - a boa-fé do infrator;

III - a vantagem auferida ou pretendida pelo infrator;

IV - a condição econômica do infrator;

V - a reincidência;

VI - o grau do dano;

VII - a cooperação do infrator;

VIII - a adoção reiterada e demonstrada de mecanismos e procedimentos internos capazes de minimizar o dano, voltados ao tratamento seguro e adequado de dados, em consonância com o disposto no inciso II do § 2º do art. 48 desta Lei;

IX - a adoção de política de boas práticas e governança;

X - a pronta adoção de medidas corretivas; e

XI - a proporcionalidade entre a gravidade da falta e a intensidade da sanção.

§ 2º O disposto neste artigo não substitui a aplicação de sanções administrativas, civis ou penais definidas em legislação específica.

§ 3º O disposto nos incisos I, IV, V, VI, VII, VIII e IX do caput deste artigo poderá ser aplicado às entidades e aos órgãos públicos, sem prejuízo do disposto na Lei nº 8.112, de 11 de dezembro de 1990 (Estatuto do Servidor Público Federal), na Lei nº 8.429, de 2 de junho de 1992 (Lei de Improbidade Administrativa), e na Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação).

§ 4º No cálculo do valor da multa de que trata o inciso II do caput deste artigo, a autoridade nacional poderá considerar o faturamento total da empresa ou grupo de empresas, quando não dispuser do valor do faturamento no ramo de atividade empresarial em que ocorreu a infração, definido pela autoridade nacional, ou quando o valor for apresentado de forma incompleta ou não for demonstrado de forma inequívoca e idônea.

Art. 53. A autoridade nacional definirá, por meio de regulamento próprio sobre sanções administrativas a infrações a esta Lei, que deverá ser objeto de consulta pública, as metodologias que orientarão o cálculo do valor-base das sanções de multa.

§ 1º As metodologias a que se refere o caput deste artigo devem ser previamente publicadas, para ciência dos agentes de tratamento, e devem apresentar objetivamente as formas e dosimetrias para o cálculo do valor-base das sanções de multa, que deverão conter fundamentação detalhada de todos os seus elementos, demonstrando a observância dos critérios previstos nesta Lei.

§ 2º O regulamento de sanções e metodologias correspondentes deve estabelecer as circunstâncias e as condições para a adoção de multa simples ou diária.

Art. 54. O valor da sanção de multa diária aplicável às infrações a esta Lei deve observar a gravidade da falta e a extensão do dano ou prejuízo causado e ser fundamentado pela autoridade nacional.

Parágrafo único. A intimação da sanção de multa diária deverá conter, no mínimo, a descrição da obrigação imposta, o prazo razoável e estipulado pelo órgão para o seu cumprimento e o valor da multa diária a ser aplicada pelo seu descumprimento.

CAPÍTULO IX
DA AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD) E DO CONSELHO NACIONAL DE PROTEÇÃO DE DADOS PESSOAIS E DA PRIVACIDADE

Seção I
Da Autoridade Nacional de Proteção de Dados (ANPD)

Art. 55. (VETADO).

Art. 56. (VETADO).

Art. 57. (VETADO).

Seção II
Do Conselho Nacional de Proteção de Dados Pessoais e da Privacidade

Art. 58. (VETADO).

Art. 59. (VETADO).

CAPÍTULO X
DISPOSIÇÕES FINAIS E TRANSITÓRIAS

Art. 60. A Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet), passa a vigorar com as seguintes alterações:

"Art. 7º .....

.....

X - exclusão definitiva dos dados pessoais que tiver fornecido a determinada aplicação de internet, a seu requerimento, ao término da relação entre as partes, ressalvadas as hipóteses de guarda obrigatória de registros previstas nesta Lei e na que dispõe sobre a proteção de dados pessoais;

....." (NR)

"Art. 16. .....

.....

II - de dados pessoais que sejam excessivos em relação à finalidade para a qual foi dado consentimento pelo seu titular, exceto nas hipóteses previstas na Lei que dispõe sobre a proteção de dados pessoais." (NR)

Art. 61. A empresa estrangeira será notificada e intimada de todos os atos processuais previstos nesta Lei, independentemente de procuração ou de disposição contratual ou estatutária, na pessoa do agente ou representante ou pessoa responsável por sua filial, agência, sucursal, estabelecimento ou escritório instalado no Brasil.

Art. 62. A autoridade nacional e o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), no âmbito de suas competências, editarão regulamentos específicos para o acesso a dados tratados pela União para o cumprimento do disposto no § 2º do art. 9º da Lei nº 9.394, de 20 de dezembro de 1996 (Lei de Diretrizes e Bases da Educação Nacional), e aos referentes ao Sistema Nacional de Avaliação da Educação Superior (Sinaes), de que trata a Lei nº 10.861, de 14 de abril de 2004.

Art. 63. A autoridade nacional estabelecerá normas sobre a adequação progressiva de bancos de dados constituídos até a data de entrada em vigor desta Lei, consideradas a complexidade das operações de tratamento e a natureza dos dados.

Art. 64. Os direitos e princípios expressos nesta Lei não excluem outros previstos no ordenamento jurídico pátrio relacionados à matéria ou nos tratados internacionais em que a República Federativa do Brasil seja parte.

Art. 65. Esta Lei entra em vigor após decorridos 18 (dezoito) meses de sua publicação oficial.

Brasília, 14 de agosto de 2018; 197º da Independência e 130º da República.

MICHEL TEMER
Torquato Jardim

Aloysio Nunes Ferreira Filho
Eduardo Refinetti Guardia
Esteves Pedro Colnago Junior
Gilberto Magalhães Occhi
Gilberto Kassab
Wagner de Campos Rosário
Gustavo do Vale Rocha
Ilan Goldfajn
Raul Jungmann
Eliseu Padilha

Este texto não substitui o publicado no DOU de 15.8.2018, e republicado parcialmente em 15.8.2018 - Edição extra

*

ANNEX B – WHATSAPP PRIVACY POLICY

The last version of WhatsApp's Privacy Policy, released on 11th of Jannuary, 2021, enabled the encrypted messaging App to share a significant amount of data with parent company Facebook in order to enhance business usage amongst group companies. However, WhatsApp user's data is public for its contacts: any person or businesses with whom it communicates.

Last modified: January 04, 2021 (archived versions)

WhatsApp Privacy Policy

If you live in the European Region, WhatsApp Ireland Limited provides the services to you under this Terms of Service and Privacy Policy.

WhatsApp Legal Info

If you live outside the European Region, WhatsApp LLC ("WhatsApp," "our," "we," or "us") provides our Services to you under this Terms of Service and Privacy Policy.

Our Privacy Policy ("Privacy Policy") helps explain our data practices, including the information we process to provide our Services.

For example, our Privacy Policy talks about what information we collect and how this affects you. It also explains the steps we take to protect your privacy, like building our Services so delivered messages aren't stored by us and giving you control over who you communicate with on our Services.

We are one of the Facebook Companies. You can learn more further below in this Privacy Policy about the ways in which we share information across this family of companies.

This Privacy Policy applies to all of our Services unless specified otherwise.

Please also read WhatsApp's Terms of Service ("Terms"), which describe the terms under which you use and we provide our Services.

Back to top

Information We Collect

WhatsApp must receive or collect some information to operate, provide, improve, understand, customize, support, and market our Services, including when you install, access, or use our Services.

The types of information we receive and collect depend on how you use our Services. We require certain information to deliver our Services and without this we will not be able to provide our Services to you. For example, you must provide your mobile phone number to create an account to use our Services.

Our Services have optional features which, if used by you, require us to collect additional information to provide such features. You will be notified of such collection, as appropriate. If you choose not to provide the information needed to use a feature, you will be unable to use the feature. For example, you cannot share your location with your contacts if you do not permit us to collect your location data from your device. Permissions can be managed through your Settings menu on both Android and iOS devices.

Information You Provide

- **Your Account Information**. You must provide your mobile phone number and basic information (including a profile name of your choice) to create a WhatsApp account. If you don't provide us with this information, you will not be able to create an account to use our Services. You can add other information to your account, such as a profile picture and "about" information.
- **Your Messages**. We do not retain your messages in the ordinary course of providing our Services to you. Instead, your messages are stored on your device and not typically stored on our servers. Once your messages are delivered, they are deleted from our servers. The following scenarios describe circumstances where we may store your messages in the course of delivering them:

- **Undelivered Messages**. If a message cannot be delivered immediately (for example, if the recipient is offline), we keep it in encrypted form on our servers for up to 30 days as we try to deliver it. If a message is still undelivered after 30 days, we delete it.
- **Media Forwarding**. When a user forwards media within a message, we store that media temporarily in encrypted form on our servers to aid in more efficient delivery of additional forwards.

We offer end-to-end encryption for our Services. End-to-end encryption means that your messages are encrypted to protect against us and third parties from reading them. Learn more about end-to-end encryption and how businesses communicate with you on WhatsApp.

- **Your Connections**. You can use the contact upload feature and provide us, if permitted by applicable laws, with the phone numbers in your address book on a regular basis, including those of users of our Services and your other contacts. If any of your contacts aren't yet using our Services, we'll manage this information for you in a way that ensures those contacts cannot be identified by us. Learn more about our contact upload feature here. You can create, join, or get added to groups and broadcast lists, and such groups and lists get associated with your account information. You give your groups a name. You can provide a group profile picture or description.
- **Status Information**. You may provide us your status if you choose to include one on your account. Learn how to use status on Android, iPhone, or KaiOS.
- **Transactions And Payments Data**. If you use our payments services, or use our Services meant for purchases or other financial transactions, we process additional information about you, including payment account and transaction information. Payment account and transaction information includes information needed to complete the transaction (for example, information about your payment method, shipping details and transaction amount). If you use our payments services available in your country or territory, our privacy practices are described in the applicable payments privacy policy.
- **Customer Support And Other Communications**. When you contact us for customer support or otherwise communicate with us, you may provide us with information related to your use of our Services, including copies of your messages, any other information you deem helpful, and how to contact you (e.g., an email address). For example, you may send us an email with information relating to app performance or other issues.

Automatically Collected Information

- **Usage And Log Information**. We collect information about your activity on our Services, like service-related, diagnostic, and performance information. This includes information about your activity (including how you use our Services, your Services settings, how you interact with others using our Services (including when you interact with a business), and the time, frequency, and duration of your activities and interactions), log files, and diagnostic, crash, website, and performance logs and reports. This also includes information about when you registered to use our Services; the features you use like our messaging, calling, Status, groups (including group name, group picture, group description), payments or business features; profile photo, "about" information; whether you are online, when you last used our Services (your "last seen"); and when you last updated your "about" information.
- **Device And Connection Information**. We collect device and connection-specific information when you install, access, or use our Services. This includes information such as hardware model, operating system information, battery level, signal strength, app version, browser information, mobile network, connection information (including phone number, mobile operator or ISP), language and time zone, IP address, device operations information, and identifiers (including identifiers unique to Facebook Company Products associated with the same device or account).
- **Location Information**. We collect and use precise location information from your device with your permission when you choose to use location-related features, like when you decide to share your location with your contacts or view locations nearby or locations others have shared with you. There are certain settings relating to location-related information which you can find in your device settings or the in-app settings, such as location sharing. Even if you do not use our location-related features, we use IP addresses and other information like phone number area codes to estimate your general location (e.g., city and country). We also use your location information for diagnostics and troubleshooting purposes.

- **Cookies**. We use cookies to operate and provide our Services, including to provide our Services that are web-based, improve your experiences, understand how our Services are being used, and customize them. For example, we use cookies to provide our Services for web and desktop and other web-based services. We may also use cookies to understand which of our Help Center articles are most popular and to show you relevant content related to our Services. Additionally, we may use cookies to remember your choices, like your language preferences, to provide a safer experience, and otherwise to customize our Services for you. Learn more about how we use cookies to provide you our Services.

Back to top

Third-Party Information

- **Information Others Provide About You**. We receive information about you from other users. For example, when other users you know use our Services, they may provide your phone number, name, and other information (like information from their mobile address book) just as you may provide theirs. They may also send you messages, send messages to groups to which you belong, or call you. We require each of these users to have lawful rights to collect, use, and share your information before providing any information to us.

  You should keep in mind that in general any user can capture screenshots of your chats or messages or make recordings of your calls with them and send them to WhatsApp or anyone else, or post them on another platform.

- **User Reports**. Just as you can report other users, other users or third parties may also choose to report to us your interactions and your messages with them or others on our Services; for example, to report possible violations of our Terms or policies. When a report is made, we collect information on both the reporting user and reported user. To find out more about what happens when a user report is made, please see Advanced Safety and Security Features.

- **Businesses On WhatsApp**. Businesses you interact with using our Services may provide us with information about their interactions with you. We require each of these businesses to act in accordance with applicable law when providing any information to us.

  When you message with a business on WhatsApp, keep in mind that the content you share may be visible to several people in that business. In addition, some businesses might be working with third-party service providers (which may include Facebook) to help manage their communications with their customers. For example, a business may give such third-party service provider access to its communications to send, store, read, manage, or otherwise process them for the business. To understand how a business processes your information, including how it might share your information with third parties or Facebook, you should review that business' privacy policy or contact the business directly.

- **Third-Party Service Providers**. We work with third-party service providers and other Facebook Companies to help us operate, provide, improve, understand, customize, support, and market our Services. For example, we work with them to distribute our apps; provide our technical and physical infrastructure, delivery, and other systems; provide engineering support, cybersecurity support, and operational support; supply location, map, and places information; process payments; help us understand how people use our Services; market our Services; help you connect with businesses using our Services; conduct surveys and research for us; ensure safety, security and integrity; and help with customer service. These companies may provide us with information about you in certain circumstances; for example, app stores may provide us with reports to help us diagnose and fix service issues.

  The "How We Work With Other Facebook Companies" section below provides more information about how WhatsApp collects and shares information with the other Facebook Companies.

- **Third-Party Services**. We allow you to use our Services in connection with third-party services and Facebook Company Products. If you use our Services with such third-party services or Facebook Company Products, we may receive information about you from them; for example, if you use the WhatsApp share button on a news service to share a news article with your WhatsApp contacts, groups, or broadcast lists on our Services, or if you choose to access our Services through a mobile carrier's or device provider's promotion of our Services. Please note that when you use third-party services or Facebook Company Products, their own terms and privacy policies will govern your use of those services and products.

Back to top

How We Use Information

We use information we have (subject to choices you make and applicable law) to operate, provide, improve, understand, customize, support, and market our Services. Here's how:

- **Our Services**. We use information we have to operate and provide our Services, including providing customer support; completing purchases or transactions; improving, fixing, and customizing our Services; and connecting our Services with Facebook Company Products that you may use. We also use information we have to understand how people use our Services; evaluate and improve our Services; research, develop, and test new services and features; and conduct troubleshooting activities. We also use your information to respond to you when you contact us.
- **Safety, Security, And Integrity**. Safety, security and integrity are an integral part of our Services. We use information we have to verify accounts and activity; combat harmful conduct; protect users against bad experiences and spam; and promote safety, security and integrity on and off our Services, such as by investigating suspicious activity or violations of our Terms and policies, and to ensure our Services are being used legally. Please see the Law, Our Rights and Protection section below for more information.
- **Communications About Our Services And The Facebook Companies**. We use information we have to communicate with you about our Services and let you know about our terms, policies, and other important updates. We may provide you marketing for our Services and those of the Facebook Companies.
- **No Third-Party Banner Ads**. We still do not allow third-party banner ads on our Services. We have no intention to introduce them, but if we ever do, we will update this Privacy Policy.

- **Business Interactions.** We enable you and third parties, like businesses, to communicate and interact with each other using our services, such as Catalogs for businesses on WhatsApp through which you can browse products and services and place orders. Businesses may send you transaction, appointment, and shipping notifications; product and service updates; and marketing. For example, you may receive flight status information for upcoming travel, a receipt for something you purchased, or a notification when a delivery will be made. Messages you receive from a business could include an offer for something that might interest you. We do not want you to have a spammy experience; as with all of your messages, you can manage these communications, and we will honor the choices you make.

Back to top

Information You And We Share

You share your information as you use and communicate through our Services, and we share your information to help us operate, provide, improve, understand, customize, support, and market our Services.

- **Send Your Information To Those With Whom You Choose To Communicate.** You share your information (including messages) as you use and communicate through our Services.
- **Information Associated With Your Account.** Your phone number, profile name and photo, "about" information, last seen information, and message receipts are available to anyone who uses our Services, although you can configure your Services settings to manage certain information available to other users, including businesses, with whom you communicate.
- **Your Contacts And Others.** Users, including businesses, with whom you communicate can store or reshare your information (including your phone number or messages) with others on and off our Services. You can use your Services settings and the “block” feature in our Services to manage who you communicate with on our Services and certain information you share.
- **Businesses On WhatsApp.** We offer specific services to businesses such as providing them with metrics regarding their use of our services.

- **Third-Party Service Providers**. We work with third-party service providers and other Facebook Companies to help us operate, provide, improve, understand, customize, support, and market our Services. We work with these companies to support our Services, such as to provide technical infrastructure, delivery and other systems; market our Services; conduct surveys and research for us; protect the safety, security and integrity of users and others; and assist with customer service. When we share information with third-party service providers and other Facebook Companies in this capacity, we require them to use your information on our behalf in accordance with our instructions and terms.
- **Third-Party Services**. When you or others use third-party services or other Facebook Company Products that are integrated with our Services, those third-party services may receive information about what you or others share with them. For example, if you use a data backup service integrated with our Services (like iCloud or Google Drive), they will receive information you share with them, such as your WhatsApp messages. If you interact with a third-party service or another Facebook Company Product linked through our Services, such as when you use the in-app player to play content from a third-party platform, information about you, like your IP address and the fact that you are a WhatsApp user, may be provided to such third party or Facebook Company Product. Please note that when you use third-party services or other Facebook Company Products, their own terms and privacy policies will govern your use of those services and products.

Back to top

How We Work With Other Facebook Companies

As part of the Facebook Companies, WhatsApp receives information from, and shares information (see here) with, the other Facebook Companies. We may use the information we receive from them, and they may use the information we share with them, to help operate, provide, improve, understand, customize, support, and market our Services and their offerings, including the Facebook Company Products. This includes:

- helping improve infrastructure and delivery systems;
- understanding how our Services or theirs are used;
- promoting safety, security and integrity across the Facebook Company Products, e.g., securing systems and fighting spam, threats, abuse, or infringement activities;
- improving their services and your experiences using them, such as making suggestions for you (for example, of friends or group connections, or of interesting content), personalizing features and content, helping you complete purchases and transactions, and showing relevant offers and ads across the Facebook Company Products; and
- providing integrations which enable you to connect your WhatsApp experiences with other Facebook Company Products. For example, allowing you to connect your Facebook Pay account to pay for things on WhatsApp or enabling you to chat with your friends on other Facebook Company Products, such as Portal, by connecting your WhatsApp account.

Learn more about the other Facebook Companies and their privacy practices by reviewing their privacy policies.

Back to top

Assignment, Change Of Control, And Transfer

In the event that we are involved in a merger, acquisition, restructuring, bankruptcy, or sale of all or some of our assets, we will share your information with the successor entities or new owners in connection with the transaction in accordance with applicable data protection laws.

Back to top

Managing And Retaining Your Information

You can access or port your information using our in-app Request Account Info feature (available under Settings > Account). For iPhone users, you can learn how to access, manage, and delete your information through our iPhone Help Center articles. For Android users, you can learn how to access, manage, and delete your information through our Android Help Center articles.

We store information for as long as necessary for the purposes identified in this Privacy Policy, including to provide our Services or for other legitimate purposes, such as complying with legal obligations, enforcing and preventing violations of our Terms, or protecting or defending our rights, property and users. The storage periods are determined on a case-by-case basis that depends on factors like the nature of the information, why it is collected and processed, relevant legal or operational retention needs, and legal obligations.

If you would like to further manage, change, limit, or delete your information, you can do that through the following tools:

- **Services Settings.** You can change your Services settings to manage certain information available to other users. You can manage your contacts, groups, and broadcast lists, or use our “block” feature to manage the users with whom you communicate.
- **Changing Your Mobile Phone Number, Profile Name And Picture, And “About” Information.** If you change your mobile phone number, you must update it using our in-app change number feature and transfer your account to your new mobile phone number. You can also change your profile name, profile picture, and "about" information at any time.
- **Deleting Your WhatsApp Account.** You can delete your WhatsApp account at any time (including if you want to revoke your consent to our use of your information pursuant to applicable law) using our in-app delete my account feature. When you delete your WhatsApp account, your undelivered messages are deleted from our servers as well as any of your other information we no longer need to operate and provide our Services. Deleting your account will, for example, delete your account info and profile photo, delete you from all WhatsApp groups, and delete your WhatsApp message history. Be mindful that if you only delete WhatsApp from your device without using our in-app delete my account feature, your information will be stored with us for a longer period. Please remember that when you delete your account, it does not affect your information related to the groups you created or the information other users have relating to you, such as their copy of the messages you sent them.

You can learn more here about our data deletion and retention practices and about how to delete your account.

Back to top

Law, Our Rights, And Protection

We access, preserve, and share your information described in the "Information We Collect"" section of this Privacy Policy above if we have a good-faith belief that it is necessary to: (a) respond pursuant to applicable law or regulations, legal process, or government requests; (b) enforce our Terms and any other applicable terms and policies, including for investigations of potential violations; (c) detect, investigate, prevent, or address fraud and other illegal activity or security and technical issues; or (d) protect the rights, property, and safety of our users, WhatsApp, the other Facebook Companies, or others, including to prevent death or imminent bodily harm.

Back to top

Our Global Operations

WhatsApp shares information globally, both internally within the Facebook Companies and externally with our partners and service providers, and with those with whom you communicate around the world, in accordance with this Privacy Policy. Your information may, for example, be transferred or transmitted to, or stored and processed in, the United States; countries or territories where the Facebook Companies' affiliates and partners, or our service providers are located; or any other country or territory globally where our Services are provided outside of where you live for the purposes as described in this Privacy Policy. WhatsApp uses Facebook's global infrastructure and data centers, including in the United States. These transfers are necessary to provide the global Services set forth in our Terms. Please keep in mind that the countries or territories to which your information is transferred may have different privacy laws and protections than what you have in your home country or territory.

Back to top

Updates To Our Policy

We may amend or update our Privacy Policy. We will provide you notice of amendments to this Privacy Policy, as appropriate, and update the "Last modified" date at the top of this Privacy Policy. Please review our Privacy Policy from time to time.

Back to top

Contact Us

If you have questions or issues about our Privacy Policy, please contact us.

WhatsApp LLC
Privacy Policy
1601 Willow Road
Menlo Park, California 94025
United States of America

Back to top